

The responsible use of Artificial Intelligence in archives through the use of paradata

Pierluigi Feliciati^(a)

Luciana Duranti^(b)

a) University of Macerata, <https://orcid.org/0000-0002-2499-8528>

b) University of British Columbia, Canada, <https://orcid.org/0000-0001-7895-1066>

Contact: Pierluigi Feliciati, pierluigi.feliciati@unimc.it; Luciana Duranti, Luciana.Duranti@ubc.ca

Received: 10 December 2024; **Accepted:** 07 March 2025; **First Published:** 15 May 2025

ABSTRACT

Using Artificial Intelligence in archives responsibly, clearly documenting why, how, by whom and with what results a specific system or tool was used in a given context, requires the conscious adoption of paradata throughout the entire period of use of such system or tool and not only at the start or end of a procedure. Paradata are structured and formalized data documenting the methodologies, processes (including technological ones) and persons (physical and juridical) associated with the production and management of records. This contribution focuses on paradata generated in connection with the use of AI systems or tools in archives to perform archival functions. Its content is primarily a result of the research conducted within the context of the international project InterPARES Trust AI.

KEYWORDS

Digital preservation process; Archival Records; Artificial Intelligence; Paradata.

L'uso responsabile dell'Intelligenza Artificiale negli archivi attraverso l'uso di paradati

ABSTRACT

Usare Intelligenza Artificiale responsabile, che documenti chiaramente perché, come, da chi e con quale risultato un certo sistema o strumento sia stato usato in un determinato contesto, richiede l'adozione consapevole dei paradati, necessari nell'intero ciclo di uso di uno strumento e non solo all'avvio o alla fine di una procedura. I paradati sono dati strutturati e formalizzati che documentano le metodologie, i processi (inclusi quelli tecnologici) e le persone (fisiche e giuridiche) associati alla produzione e alla gestione dei documenti. Il focus di questo contributo riguarda i paradati prodotti in connessione all'uso di sistemi o strumenti di Intelligenza Artificiale negli archivi per esercitare funzioni archivistiche. Il suo contenuto è principalmente il risultato parziale delle ricerche condotte nell'ambito del progetto internazionale InterPARES Trust AI.

PAROLE CHIAVE

Processi di conservazione digitale; Archivi digitali; Intelligenza Artificiale; Paradati.

Introduzione

“A pivotal question linked to data, information, and knowledge is what we need to know about data, information, and knowledge and how it has been managed in order to use it? Sometimes the answer is ‘very little.’ [...] This is where paradata, i.e., information on the processes of data creation, curation, and earlier and intended use, can be pivotal for the usefulness of data, information, and knowledge.” (Andersson et al. 2024, 1-2).

Nel contesto archivistico, mentre i metadati sono i dati strutturati e formalizzati relativi ai documenti archivistici, necessari per cercare, identificare, visualizzare, analizzare, valutare e gestire nel tempo quei documenti, i paradata sono dati strutturati e formalizzati che documentano le metodologie, i processi (inclusi quelli tecnologici) e le persone (fisiche e giuridiche) associati alla produzione e alla gestione dei documenti. Come fenomeno informativo, i paradata differiscono dai metadati anche qualitativamente, perché puntano a testimoniare il contesto di produzione e gestione piuttosto che l'identità e l'integrità dei documenti, e quantitativamente, in quanto comprendono informazioni più ampie e meno dettagliate rispetto ai metadati.

Il recentissimo volume curato da Lisa Andersson, Isto Huvila e Olle Sköld della Uppsala University (Andersson et al. 2024), nell'affrontare il tema dei paradata, ha scelto di coinvolgere studiosi di diverse discipline, anche lontane dall'archivistica, che si confrontano con la gestione delle informazioni. Un punto di partenza per comprendere il ruolo dei paradata può essere lo spinoso problema di come gestire gli aspetti taciti e impliciti della conoscenza e unirli con forme esplicite e formalizzate di informazioni e dati (Polanyi 1967), connettendo nei fatti le due principali prospettive sulla gestione della conoscenza, quella sociale e organizzativa di mediazione e traduzione tra le persone e quella tecnica di formalizzazione e gestione in sistemi, ambedue tendenzialmente soluzioni tecniche (Handzic 2004). I paradata, applicati ad esempio nelle scienze statistiche per presentare i dati dei sondaggi, sono finalizzati alla descrizione del processo di elaborazione e conduzione del sondaggio (Couper 1998; 2000; Frauke Kreuter 2010). In modo simile, nel settore delle applicazioni tecnologiche al patrimonio culturale sono stati utilizzati per raccogliere sistematicamente informazioni sulla creazione e l'elaborazione dei dati, specie per le elaborazioni tridimensionali (Denard 2012). Infatti, la creazione di ricostruzioni 3D, un processo tutt'altro che lineare, richiede la registrazione dei dati utilizzati e la decodifica dei processi adottati per garantirne la documentazione (Papadopoulos 2024).

Va menzionato in particolare che, nel contesto archivistico, altri concetti e metodi non lontani dai paradata sono già utilizzati: tra tutti, quello della *provenienza* è un concetto usato comunemente dalla professione archivistica per descrivere sia la creazione che la gestione di oggetti informativi e in particolare dei documenti archivistici. Rappresentare la provenienza può comprendere la descrizione sia delle persone giuridiche coinvolte, sia delle azioni che queste intraprendono per produrre (generare o ricevere), organizzare, gestire, valutare e scartare i documenti archivistici (Lemieux 2016). Il focus di questo articolo è la creazione di paradata in connessione all'uso di strumenti di Intelligenza Artificiale (IA) negli archivi per esercitare funzioni archivistiche. Il suo contenuto è il risultato soprattutto della ricerca condotta nell'ambito del progetto internazionale InterPARES Trust AI (2021-2026).¹

¹ In breve, *I Trust AI*. Il sito web del progetto di ricerca è www.interparestrustai.org.

Lo studio *Preserving AI Techniques as Paradata*

I Trust AI ha lo scopo di concepire, disegnare, sviluppare e usare sistemi o strumenti di IA per garantire la conservazione e l'accessibilità di documenti archivistici affidabili, accurati, e autentici. A questo fine, il gruppo di ricerca porta avanti sia studi di casi, sia studi di carattere generale. Lo studio sui paradata fa parte di questi ultimi.² Il suo obiettivo è sviluppare un approccio per documentare l'uso di sistemi o strumenti di IA che soddisfi le responsabilità dell'archivio rispetto ai suoi utenti. L'IA responsabile (*accountable*) è cosa diversa rispetto all'IA spiegabile (*Explainable AI*), il cui scopo è solo quello di rendere uno strumento comprensibile. L'IA responsabile deve documentare chiaramente perché, come, da chi, e con quale risultato un certo sistema o strumento sia stato usato in un determinato contesto. Il tipo di spiegazione offerta può comunicare come lo strumento funziona, come sappiamo che funziona, perché funziona nel contesto in cui si intende usarlo, ma anche quali sono i riferimenti istituzionali nel cui ambito viene usato, ovvero le policy, i regolamenti, e le procedure pertinenti al suo uso specifico, insieme ai nomi delle persone responsabili per usarlo. Se in ambito informatico i termini IA responsabile e IA spiegabile sono considerati spesso pressoché intercambiabili³, uno dei risultati dei lavori interdisciplinari di InterPARES è stato chiarirne la differenza. *Accountable* non è solo *Explainable*, perché conoscere l'algoritmo e il modo in cui uno strumento funziona non attribuisce esplicitamente nessuna responsabilità, mentre le scelte di utilizzare un determinato strumento, istruirlo con determinati *data sets* creando un modello unico per uno scopo preciso sono attribuibili a una persona naturale o giuridica che le fa. Le persone, le procedure e il loro contesto sono componenti chiave del ciclo della responsabilità; quindi, devono essere documentate in connessione alle informazioni sulle componenti tecnologiche.

Conseguentemente a tali riflessioni, il gruppo di ricerca ha sviluppato prima di tutto una definizione operativa: "I paradata sono informazioni sulle procedure e sugli strumenti utilizzati per creare ed elaborare risorse informative, insieme a informazioni sul funzionamento degli strumenti e sull'esecuzione delle procedure, nonché sulle persone che eseguono tali procedure e utilizzano tali strumenti".⁴

Gli studi di InterPARES considerano inoltre i paradata necessari nell'intero ciclo di uso di uno strumento. Questo significa che i paradata non sono determinati solo all'inizio di una procedura o aggiunti alla fine della stessa, ma sono espressamente dichiarati dal momento della scelta del sistema o dello strumento per uno scopo specifico fino al raggiungimento di tale scopo, includendo la descrizione del suo risultato.

Quali paradata siano necessari e come raccogliarli e conservarli sono due questioni problematiche. Poiché l'implementazione di un processo di IA è unica per ogni soggetto archivistico e dipende dai suoi scopi, dalla sua tolleranza dei rischi e dalle sue risorse, la risposta alle due questioni sarà

² I ricercatori coinvolti nello studio *Research and Preservation 04, Preserving AI Techniques as Paradata*, sono: Patricia Franks, Babak Hamidzadeh, Scott Cameron, Norman Mooradian, Alex Richmond e Mario Beauchamp. Per approfondire cfr. (Franks et al. 2022; Cameron et al. 2023; Davet et al. 2023; Cameron e Hamidzadeh 2024).

³ Cfr. ad esempio Baker, Xiang 2023.

⁴ Traduzione italiana della definizione presente nell'InterPARES Terminology database: "Information about the procedure(s) and tools used to create and process information resources, along with information about the persons carrying out those procedures", <https://interparestrustai.org/terminology/term/paradata/en>.

necessariamente diversa da un contesto ad un altro. Al tempo stesso, lo scopo dello studio di I Trust AI è da un lato identificare elementi comuni che debbano essere documentati in tutti i contesti di uso archivistico di sistemi o strumenti di IA, e dall'altro identificare una tipologia di elementi unici che possono essere adottati da organizzazioni che intendano sviluppare una propria, specifica strategia.

A questo scopo, il gruppo di ricerca ha condotto uno studio di caso nell'archivio della Banca del Canada.⁵ I paradata identificati sono di due tipi: tecnici e organizzativi. Quelli tecnici includono il *data set* e i parametri usati per istruire il modello, i *log* generati durante l'uso dello strumento e, nel caso si tratti di un sistema o strumento commerciale, non sviluppato dall'archivio, la documentazione fornita dal venditore e le informazioni sulla versione usata. I paradata organizzativi consistono nella *policy* riguardante l'uso dell'IA, nei regolamenti e nelle considerazioni etiche, nella pianificazione progettuale, nella formazione del personale dell'archivio, nella valutazione dell'impatto dell'uso della IA, e nella procedura di implementazione.

Per quanto riguarda la conservazione, le domande che il gruppo di ricerca si è posto sono: in quale parte del sistema di tenuta dei documenti o di conservazione archivistica vanno mantenuti i sistemi o gli strumenti di intelligenza artificiale? Come possiamo conservare le loro componenti? In quali situazioni dobbiamo conservare i sistemi o gli strumenti di IA, o parti di questi? Fino a che punto ciò che conserviamo deve essere riproducibile? Se ci sono esseri umani coinvolti nel processo decisionale, come rappresentiamo e conserviamo le informazioni sui loro ruoli?

Dobbiamo infatti considerare che ci sono tre tipi di sistemi di IA. Il tipo più comune è quello che non prevede la partecipazione di essere umani. I sensori raccolgono dati dall'ambiente, gli strumenti di controllo li analizzano e decidono quali azioni compiere, le componenti esecutive agiscono e il risultato di tali azioni torna ai sensori, quindi il ciclo si ripete. Questo comportamento è quello tipico dei robot. Il secondo tipo di IA, con cui abbiamo sempre più familiarità è quello che raccomanda agli esseri umani, sulla base del loro comportamento online, cosa comprare o quale sito visitare. Questo tipo di sistema, in modo simile a quello precedente, non risulta difficile da documentare e da conservare. Gli esseri umani possono reagire o no, ma le loro attività si alternano, non sovrappongono, a quelle del sistema.

⁵ I professionisti responsabili per lo studio di caso sono Alex Richmond, Mario Beauchamp e John Cameron della Bank of Canada.

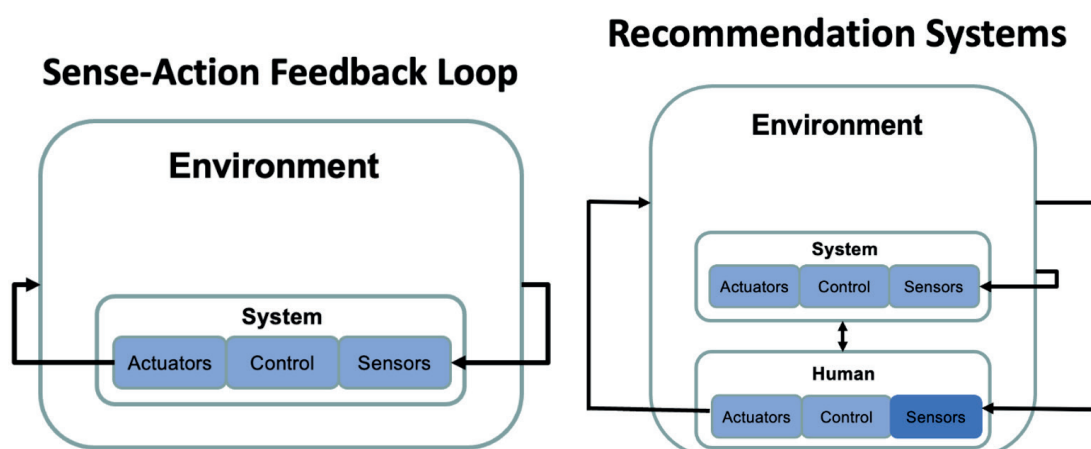


Figura 1. P. Franks, *Paradata - what is it and why do we care?* Presentazione in *Using AI To Interrogate Archives - iTrust AI Public Symposium*, 23 Febbraio 2024, Honolulu.⁶

Il terzo tipo di sistema, più problematico, è quello in cui l'IA e gli esseri umani interagiscono, ma possono farlo in modo indipendente, a volte a turno, altre in sovrapposizione o congiuntamente e altre volte per correggere, aggiungere, o sostituire l'azione l'uno dell'altro (come in caso di emergenze). Questi *action systems* sono dinamici, richiedono azioni in tempo reale e possono coinvolgere un numero variabile di agenti diversi.

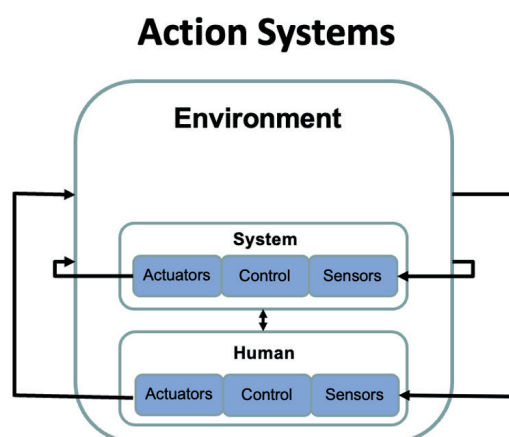


Figura 2. P. Franks, *Paradata - what is it and why do we care?* Presentazione in *Using AI To Interrogate Archives - iTrust AI Public Symposium*, 23 Febbraio 2024, Honolulu.

Cosa risulta preferibile e possibile conservare di questi sistemi o strumenti? Certo, prima di tutto informazioni sul sistema/strumento stesso: l'architettura e il *design* di hardware e software, il codice, il modello, gli algoritmi, la logica e i meccanismi di esecuzione. Inoltre, la documentazione

⁶ La registrazione del Symposium è disponibile in https://fb.watch/qEq_IVLAyu/.

relativa al mantenimento e all'aggiornamento, i dati usati per istruire il modello, testarlo e validarlo e infine i risultati di tutte queste azioni, più gli strumenti per il funzionamento del sistema/strumento. A tutte queste informazioni relative al sistema/strumento e al suo uso bisogna aggiungere i documenti prodotti dall'organizzazione che usa il sistema/strumento relativi al processo di auto-controllo, ai test che precedono l'adozione del sistema/strumento e, nel caso di sistemi/strumenti commerciali, alla storia delle vendite di modelli e configurazioni.

Tutti questi sono da considerarsi paradatai generali, che devono essere accompagnati da paradatai operativi. Questi ultimi includono:

- l'input fornito dai sensori (*log* dei dati e, nel caso di sistemi di *computer vision*, i video e/o le immagini);
- le decisioni prese dal *controller* (*log* delle direzioni e relativi parametri di *setting*, dati sugli intermediari che hanno supportato le decisioni prese, le spiegazioni *post factum* di questi processi, i *log* degli avvertimenti e delle notifiche di passaggio/affidamento di responsabilità);
- le azioni dell'attuatore (componente esecutiva) (*log* delle azioni controllate dagli esseri umani e di quelle controllate dal sistema IA, *log* dei messaggi del sistema per i controllori umani e per i soggetti esterni);
- e i risultati (*log* dei dati percepiti dai sensori dopo che le azioni degli attuatori sono state completate, o video delle conseguenze di tali azioni).

Per quanto riguarda i paradatai, come detto sopra, la documentazione dovrebbe essere continua, ovvero basata su un modello ciclico di *feedback* (sensori, controllori, attuatori) al fine di rispecchiare ogni fase, ma anche a campione e relativa a eventi specifici nel processo. Inoltre, è importante documentare la relazione tra i dati percepiti e registrati dai sensori, la logica usata dal controllo nel determinare l'azione che ne deve conseguire, l'attuazione di tale azione e le sue conseguenze, e infine la dimensione temporale.

Le decisioni prese dall'archivio e le azioni dei sistemi o degli strumenti di IA conseguentemente adottati devono insomma essere documentate. Parte di questa documentazione sarà automatica, come prodotto del sistema o strumento di IA, e un'altra parte sarà generata dall'archivio prima dello sviluppo o dell'adozione dello specifico sistema o strumento (ad esempio in forma di linee guida di natura etica, regolativa o procedurale), oppure durante e successivamente al loro uso, indicando le persone che hanno portato avanti le varie fasi della procedura, come la determinazione del *data set* per istruire il modello o il suo testing. Tutti questi paradatai sono necessari non solo a promuovere la piena trasparenza ma soprattutto a soddisfare i requisiti relativi alle attestazioni di responsabilità. I paradatai e la loro conservazione garantiscono che i processi portati avanti da sistemi o strumenti di IA siano documentati e l'autenticità dei documenti da loro generati sia garantita.

Conclusioni

Ciò detto, va tuttavia sottolineato che i profili applicativi e la necessità dei paradatai rimangono un campo di studio aperto nella maggior parte delle discipline interessate alla creazione e gestione di dati. Allo stesso tempo, le relazioni dei paradatai con i metadati e la prospettiva di incorporare le informazioni relative all'elaborazione dei dati in forma di descrizioni strutturate sono attualmente in fase di sperimentazione da parte di diversi gruppi di ricerca.

Esistono già, infatti, modelli ben definiti che coprono tipi specifici di paradata, come lo standard CIDOC-CRMdig,⁷ che ambisce a documentare le fasi e i metodi di produzione dei prodotti di digitalizzazione e delle rappresentazioni digitali, la specifica PROV del W3C per la modellazione della provenienza dei dati (Groth e Moreau 2013) e il *Common Workflow Language* per la formalizzazione dei flussi di lavoro analitici computazionali (Amstutz et al. 2016).

La inevitabile trasversalità interdisciplinare delle indagini sui paradata deve costituire l'occasione per attivare un effetto positivo, per espandere gli orizzonti della ricerca e delle pratiche di gestione delle informazioni e della conoscenza.

⁷ <https://cidoc-crm.org/crmvig/home-2>.

Riferimenti bibliografici*

Amstutz, Peter, Michael R. Crusoe, Nebojša Tijanić, Brad Chapman, John Chilton, Michael Heuer, Andrey Kartashov, Dan Leehr, Hervé Ménager, Maya Nedeljkovich, Matt Scales, Stian Soiland-Reyes, e Luka Stojanovic. 2016. *Common workflow language, v1.0*. <https://doi.org/10.6084/M9.FIGSHARE.3115156.V2>.

Andersson Lisa, Isto Huvila Isto, e Olle Sköld Olle. 2024. "An Introduction to Paradata." In *Perspectives on Paradata. Research and Practice of Documenting Process Knowledge*, a cura di Lisa Andersson, Isto Huvila, e Olle Sköld. Cham, Switzerland: Springer, <https://doi.org/10.1007/978-3-031-53946-6>.

Baker Stephanie e Xiang Wei. *Explainable AI is Responsible AI: How Explainability Creates Trustworthy and Socially Responsible Artificial Intelligence*. Sottomesso il 4 dicembre 2023. <https://arxiv.org/abs/2312.01555v1>.

Bentkowska-Kafel, Anna, e Hugh Denard, a c. di. 2012. *Paradata and Transparency in Virtual Heritage*. London: Routledge. <https://doi.org/10.4324/9781315599366>.

Cameron, Scott, e Babak Hamidzadeh. 2024. "Preserving Paradata for Accountability of Semi-Autonomous AI Agents in Dynamic Environments: An Archival Perspective." *Telematics and Informatics Reports* 14. <https://doi.org/10.1016/j.teler.2024.100135>.

Cameron, Scott, Patricia Franks, e Babak Hamidzadeh. 2023. "Positioning paradata: A conceptual frame for AI processual documentation in archives and recordkeeping contexts." *ACM. Journal of Computing and Cultural Heritage* 16 (4): 1-19. <https://doi.org/10.1145/3594728>.

Couper, Mick P. 1998. "Measuring survey quality in a CASIC environment." In *Proceedings of the Section on Survey Research Methods of the American Statistical Association* 6: 41-49. http://www.asasrms.org/Proceedings/papers/1998_006.pdf.

Couper, Mick P. 2000. "Usability evaluation of computer-assisted survey instruments". *Social Science Computer Review* 18(4): 384-396. <https://doi.org/10.1177/089443930001800402>.

Davet, Jeremy E., Babak Hamidzadeh, e Patricia Franks. 2023 "Archivist in the machine: paradata for AI-based automation in the archives." *Archival Science* 23: 275-295. <https://doi.org/10.1007/s10502-023-09408-8>.

Franks, Patricia C., Jeremy E. Davet, Babak Jeremy E., Babak Hamidzadeh, e Jenny Bunn. 2022. "Tracking the functions of AI as paradata & pursuing archival accountability." *Archiving 2022: Final Programs and Proceedings, Society for Imaging Science and Technology*, 19: 83–88. <http://dx.doi.org/10.2352/issn.2168-3204.2022.19.1.17>.

Groth, Paul, e Luc Moreau. 2013. *PROV-Overview: An Overview of the PROV Family of Documents*. <http://www.w3.org/TR/2013/NOTE-prov-overview-20130430/>.

Handzic, Meliha. 2004. *Knowledge management: Through the technology glass*. Singapore: World Scientific.

* Data di ultima consultazione dei siti web: 1 marzo 2025.

Kreuter, Frauke, Mick Couper, e Lars Lyberg. 2010. *The use of paradata to monitor and manage survey data collection*. In *Section on Survey Research Methods – JSM 2010*, 282-296. http://www.asasrms.org/Proceedings/y2010/Files/306107_55863.pdf.

Lemieux, Victoria, a c. di. 2016. *Building Trust in Information*. Cham: Springer.

Papadopolous, Costas. (2024). “A Leap of Faith: Revisiting Paradata in 3D Scholarship.” In *Perspectives on Paradata. Research and Practice of Documenting Process Knowledge*, a cura di Lisa Andersson, Isto Huvila, e Olle Sköld, 61- 86. Cham: Springer. <https://doi.org/10.1007/978-3-031-53946-6>.

Polanyi, Michael. 1967. *The tacit dimension*. London: Routledge. Tradotto da Franco Voltaggio come *La conoscenza inespresa* (Roma: Armando. 1979).

What are the technical factors affecting the copyright protection of information resources in Libraries Information Systems? A Fuzzy Delphi Approach*

Zeinab Papi^(a)

a) National Library and Archives of the Islamic Republic of Iran, <https://orcid.org/0000-0002-1495-0723>

Contact: Zeinab Papi, z-papi@nlai.ir

Received: 16 October 2024; **Accepted:** 25 March 2025; **First Published:** 15 May 2025

ABSTRACT

Several factors, such as technical aspects, affect the copyright protection of information resources in information systems (ISs) used in libraries. By identifying the technical factors affecting the copyright protection of information resources in the systems, libraries can take the necessary measures to protect the rights of creators of works and make the resources legally accessible to users. The present study aimed to identify the technical factors affecting the copyright protection of information resources in the ISs of the National Library and Archives of Iran (NLAI).

The qualitative approach and documentary analysis, thematic analysis, and fuzzy Delphi methods were used in this study to collect and validate the data. The research population consisted of seven Iranian specialists and experts selected for interviews, 11 specialists selected for the Delphi panel, and seven ISs of the NLAI selected for examination and study. In this study, 16 indicators were identified as technical factors affecting copyright protection of the systems under study. These indicators were identified as effective technical factors in both bibliographic and digital systems. In general, the systems under study (i.e., digital and bibliographic) had no proper status regarding copyright protection of information resources. Therefore, the managers of such systems must take necessary measures to review the technical checklists to evaluate the systems in terms of protecting the copyright of information resources. Furthermore, they must try to improve the effective technical factors in these systems. Paying attention to the technical factors mentioned in the present research can reduce the risk of copyright infringement in information systems and digital libraries.

KEYWORDS

Copyright; Information resources management; Information systems (ISs); Technical factors; National Library and Archives of Iran (NLAI).

* The present research is taken from the research project approved by National Library and Archives of Iran (NLAI), which was carried out under the title “Review and analyzing the technical requirements copyright protection for information resources in NLAI information systems: Providing solutions” in 2023.

1. Introduction

Intellectual property rights and copyrights pose various challenges for librarians in all stages of the information and source management cycle, from creating digitized resources and providing external resources to access and long-term digital collections management (Muir 2006). Various factors, including technical factors, affect the copyright protection of these resources in Information systems (ISs). Identifying the technical factors affecting the copyright protection of information resources in the information systems allows libraries to take necessary measures to protect the rights of the creators/producers of works and make the resources legally accessible to users. The rights of creators (i.e., copyright holders) of works can be protected in the digital space through technical requirements. However, some technical factors are not implemented in the systems for various reasons. As a result, it affects the protection of the rights of the producers/creators of information resources. The effective technical factors may be few, but libraries cannot neglect them. As we know, the information age has significantly affected all sectors of the society, including libraries. Libraries have experienced many quality improvements in their services, the nature of the information provided, and the beneficiaries, thereby turning into modern libraries. In these libraries, there has been much dependency on technologies in digitalizing information and data for effective access and greater efficiency in information storage and processing (Niqresh 2019). The reason is that technological developments have fundamentally changed many of the common, routine practices in libraries. Library resources are mostly used, delivered, and maintained digitally. Libraries provide collections of music, movies, etc., in digital formats thanks to features such as convenient search, retrieval, access, and transfer (Wahid 2017). They play a vital role in developing the information society by providing the users with information resources in various tools through the Web. Libraries and librarians encounter problems regarding digital information security management policy, data privacy, data backup, data protection against malware and viruses, and technical support. Hence, it is essential to protect library resources (Farid, Warraich, and Iftikhar 2023) from being misused. To prevent unauthorized access to library resources and copyright protection of resources, whether physical or digital, library managers and information experts should consider strategies to provide adequate security for digital information resources in libraries. The reason is that library resources contain valuable information that helps libraries meet users' information needs. In this respect, the development of digital collections, including audio-visual equipment, electronic sources, microforms, etc., will help libraries better meet the information needs of their users (Ferdinand, Ijiekhuamhen, and Nneka 2015).

Following the increased use of ISs, libraries have started their digitization programs to make library resources available to web users. To provide free access to information, libraries have been making their information resources available to users through ISs for years. To this end, they have used various information resources, including books, journal articles, research projects, manuscripts and lithographic books, maps, photos, theses, dissertations, etc. Libraries are usually less interested in providing the full content of resources through ISs as they tend to protect the copyright information resources and decrease the risk of unauthorized access. Although this solution cannot contribute to the survival of libraries in the long run, it may be used more for libraries that cover many diverse resources. National libraries in every country are responsible for providing, processing, maintaining, and disseminating information. In Iran, the National Library and Archives of Iran (NLAI) is

responsible for implementing ISs. Digital systems, including Iran's National Digital Library (Digital Library¹), Iran publications system², Iranian Scientific Publications System³, Iran Publications Database⁴, the National Document Center Network⁵ in Iran, the Catalog of Manuscripts⁶ available in Iran, and Iran Libraries Network⁷ are bibliographic systems that provide users with various resources. These information resources include books, manuscripts, lithographs, research projects, photos, maps, and journal articles. However, offering information resources in these systems has led to many issues related to resource digitization. As was mentioned, one of these issues is copyright protection of resources in the ISs, where the resources are stored and made accessible to users. Since the systems examined in this research were also common among other libraries, the suggestions and results of this research can be used by other libraries as well. Accordingly, considering the stated goal and problem, the current study will look for an answer to the following question to evaluate the systems under study by identifying the effective technical factors from the experts' point of view. The main question to respond to in this study is:

RQ. What technical factors affect the copyright protection of information resources in the systems under study?

2. Review of the related literature

Because of their importance, many studies have focused on security issues in ISs and digital libraries. For instance, Ismail and Zainab (2011) investigated the security of ISs of public libraries in Malaysia, considering network security to be one of the technical factors in the security of library ISs. Researchers proposed using the Library Information Security System Assessment Model (LISSAM) that libraries should also implement good backup policies and recovery procedures to ensure their data and services via information systems can be accessed and shared in a convenient way whenever it was needed and data can be restored quickly during downtime. Therefore, network security was considered an important technical factor.

Khan, Ibrahim, and Hussain (2021) also examined Pakistan university libraries' data and information security status. Their findings showed an unsatisfactory situation of such security measures. Lack of updated antivirus and library management software, firewalls, testing and debugging mechanisms, use of workstations, no backup data, and updated security patches for operating systems are some issues reported in the survey. Similarly, mandatory actions are usually taken by the IT section of any department. The measures such as allowing restricted user access to certain important databases, file sharing mechanism, and maintaining log of access (e.g., wireless connected devices) are some of the critical issues that need to be considered by libraries in this study. Other studies have dealt with the issue of the security of information resources in libraries. Farid,

¹ <https://dl.nlai.ir/ui/forms/index.aspx>

² <https://sana.nlai.ir>

³ <https://iranjournals.nlai.ir/>

⁴ <https://mags.nlai.ir/>

⁵ <https://docs.nlai.ir/>

⁶ <https://scripts.nlai.ir/>

⁷ <https://libs.nlai.ir/>

Warraich, and Iftikhar (2023) investigated the policy of digital information security management in academic libraries using a Systematic Review method. They concluded that digital information security policy and data are secure in some academic libraries, such as those in the United Kingdom, United States of America (USA), China, and the United Arab Emirates (UAE). However, other university libraries, such as those in India, Pakistan, Nigeria, Ghana, and Oman, have no such procedures. In Iran, Kokabi and Kohi Rostami (2015) found that access to information and processes of the organization should be controlled based on organizational and security needs. This rule should be considered in information dissemination and licensing procedure policymaking. The rights access rules of each user or group of users should be explicitly and accurately stated and recorded in the access right granting method document. According to Peykam and Salimifard (2016), organization's information security policy and the documentation of organizational processes are among the factors affecting the security of ISs. In addition, Eidi Ghaleshiri (2015) believes that the settings and management of firewalls and host-based intrusion detection systems (IDSs) at the level of applications and the configuration of firewalls and routers are among the challenges and technical factors in the design of digital libraries, especially the Digital Library of Astan-e Quds-e Razavi in Iran.

Rezvani (2018) considered two categories of solutions and factors to be effective in improving the security of Iranian digital library systems. This study was conducted using the correlation and path analysis. The author concluded that the security of ISs can be improved using encryption principles in the authentication or replication of services to increase the level of access. Also, using an enterprise service bus (ESB) to implement security controls in the architectural design and using security-WS were reported as effective measures for message-level security. In addition, the development of an information security policy, information architecture, organization of information security, development of ISs, security of human resources, and content protection affect information security in digital libraries. Doroudi and Jamshidi (2021) considered physical and environmental security; communication and operations management; access control; preparation, development and maintenance of ISs; human resource security; asset management; information organization; and security policy among the factors affecting information security in the digital libraries of the Qom province in Iran.

Other researchers have investigated the issue of licenses for information resources, technical requirements for protecting the copyright of resources in ISs, digital copyright, and copy detection systems. Fernández-Molina (2004) conducted an analytical study on licensing agreements for information resources and limitations and exceptions to copyright. He believes that the emergence and development of the digital environment have posed major changes in the conditions of access, distribution, and use of intellectual works, thereby affecting the rights holders and users. For online exploitation of intellectual works, contractual agreements and licenses can be used as the most common way of accessing projects. Therefore, there is a need for clear rules for license agreements and their relationship with limitations and exceptions to copyright. The activities performed by legal users without obtaining permission from the rights holders should be limited to not disrupt the adequate use of intellectual works and the proper functioning of the information society.

Gaber (2012) studied the users' rights in the Digital rights management (DRM) system, concluding that DRM prevents illegal access and copying, allowing users to have access only to digital content with permits based on license issuers. Griffin (2013) investigated the opportunities and threats of digital copyright exchange. The results showed that digital copyright can be implement-

ed as an alternative solution for traditional publishers, and the issuance of digital copyright helps present digital works more easily. In other words, it prevents the copyright infringement of works. Although this model may put traditional publishers at a weak point in the competition with digital licensing, it can change the way of creating new creative works and the way of using the copyright of the works. For instance, it can alter digital technology as a phenomenon that has facilitated sharing individuals' thoughts and works worldwide.

Papi et al. (2017) considered the copy detection system among the main infrastructure elements of the Iran's National System for Theses and Dissertation (INSTD). Applying copy detection systems, which use similarity-finding software, prevents unfair use and duplication of theses and dissertations. Thesis and dissertation integrated systems are of importance in setting up the infrastructure for the repository.

Papi (2021) investigated the copyright of printed and digital library resources in Iran using a qualitative approach and documentary analysis method. Fair use and exceptions, access to information resources, material and intellectual rights, license and agreement, for-profit use, type of resources, and supporting and violating the rights of creators are the most important issues to consider developing solutions for material and intellectual rights and the printed and digital source copyright policy in libraries; although there was no copyright policy in Iranian libraries. Finally, this author offered suggestions and operational solutions for the copyright bill of library resources and use in libraries according to the types of resources. Moreover, they suggested creating a field called "Copyright Status" or "Rights and Access" in the bibliographic record in library software, which can contribute considerably to protecting the rights of creators.

From studying various researches, it became clear that a category of studies addressed security components in library information systems. The research of Eidi Qaleh Shiri (2015) falls into this category. His research discussed firewalls in libraries and security issues. Other studies such as Ismail and Zainab (2011); Khan, Ibrahim, and Hussain (2021); Kokabi and Kohi Rostami (2015); Farid, Warraich, and Iftikhar (2023); Peykam and Salimifard (2016); Rezvani (2018); and Doroudi and Jamshidi (2021) also belong to this category.

For example, in the research by Kokabi and Kohi Rostami (2015) and Rezvani (2018), access control indicators in the systems of public libraries in Iran and the digital library systems of Iran were examined. In the first study, security indicators were rated at a moderate level, while in another study, the digital system of the National Library of Iran was rated at a good level. In addition to the above studies, the topic of licensing for information resources was explored in the research by Fernández-Molina (2004).

One of the technical factors for improving the virtual space and the interaction between users and creators is the use of DRM systems and attention to copyright in the digital environment. The application of DRM technology and digital copyright to prevent unauthorized duplication has been highlighted and emphasized in the research of Gaber (2012) and Griffin (2013).

Emphasis on the formulation of copyright policies and obtaining permission from publishers are two points related to the research by Papi (2021), which examined the copyright of printed and digital sources in Iran's library resources. The use of similarity detection can help identify violations of the rights of creators in information systems. In the research by Papi and et al. (2017), the copyright detection system was discussed as a technical factor. In the methodology section, the studies under consideration were conducted using various methods.

Some researchers, such as Ismail and Zainab (2011) and Khan, Ibrahim, and Hussain (2021), have explored data and information security in libraries and their ISs using survey methods and tools such as questionnaires and checklists. The digital information security management policy and copyright policy were analyzed by Peykam and Salimifard (2016), Papi (2021), and Farid, Warraich, and Iftikhar (2023). In the study by Peykam and Salimifard (2016), the researcher designed a questionnaire and used the Fuzzy Analytic Hierarchical Process (FAHP) technique to calculate the weights and ranks of factors. Papi (2021) studied resources using the Documentary analysis method. Farid, Warraich, and Iftikhar (2023) employed a systematic method for analysis.

Kokabi and Kohi Rostami (2015) and Doroudi and Jamshidi (2021) examined technical indicators using the survey research method. The research of Griffin (2013) and Fernández-Molina (2004), was conducted using the review and descriptive method. Also, Rezvani (2018) studied the information security by Correlational research and survey. In the study by Eidi Ghaleshiri (2016), Quantitative Evaluation Methods was used. Gaber (2004) used a qualitative systematic analysis to analyze technical indicators. Papi et al. (2017) research investigated the system by a checklist and the grounded theory approach.

Therefore, according to the examination and analysis of the studies conducted, it can be stated that such a research has not been done in Iran so far. Based on this, several features of the current research can be presented: in this research, the use of technical factors in support of copyright was examined. Additionally, various systems of the NLAI were reviewed in this study. From digital resource systems containing various informational resources (books, theses and dissertations, manuscripts, non-book resources, and the like) with different features and accessibility to other systems were analyzed in the current research. While in other studies, emphasis was placed on the security issues of libraries. Moreover, there were also differences in methodology compared to other studies. In this research, considering the objectives and research questions, methods and tools such as qualitative approaches, document analysis, thematic analysis, fuzzy Delphi methods, as well as tools like interviews and observations were utilized.

3. Methodology

In this applied study, a qualitative approach was employed for data collection. In the first step, different studies and resources were examined and analyzed using the documentary analysis method. For this purpose, Iranian and foreign databases, such as online public access catalog (OPAC) available in NLAI, the INSTED, and the Academic Center for Education, Culture, and Research (ACECR) in Iran, were searched using Persian keywords. Also, foreign databases such as Emerald, Google Scholar, Science Direct, Acalib, and Sage were searched using keywords and phrases such as “copyright in libraries”, “copyright and information resources”, “ISs”, “technical factors in systems”, and “security of ISs”. In the second step, the resources were studied, and the technical factors affecting the copyright protection of information resources were extracted. Until these factors are used alongside the factors extracted from the interviews for the checklist.

In the third step, semi-structured interviews were conducted with 7 Iranian and non-Iranian specialists and experts in person and via e-mail (Table 1). In selecting individuals for the interviews, purposive sampling was used, and then a snowball sampling method was employed.

Table 1. Characteristics of Experts in the Interview

<i>Interviewee Code</i>	<i>Level of Education</i>	<i>Area of Expertise</i>	<i>Work Experience (Year)</i>	<i>Name of Country</i>
1	Dr. at Law, Assistant Professor	Copyright and technology	10	India (Delhi)
2	PhD in Computer Science	Digital library strategist and metadata architect	25	United States (Pennsylvania)
3	Master of Library Science	Library Technology, Digital libraries, Intellectual property	40	United States (California)
4	Ph.D. in Knowledge and Information Science	Information technology	22	Iran (Tehran)
5	Ph.D. in Information Technology	Information systems	20	Iran (Tehran)
6	Ph.D. in Information Technology and Ph.D. in Knowledge and Information science	Information systems, Information technology	21	Iran (Tehran)
7	Master in Information Technology	Digital libraries	17	Iran (Tehran)

To clarify any ambiguities in the interview and to obtain more in-depth information, the interviews continued for three or four rounds. Before conducting the interview, a brief overview of the topic was provided to the interviewee. Numerical codes (1-7) were used to keep the interviewees' names confidential. The interviews continued until new data were obtained and until data saturation was reached. The interviews took place from March 2023 to May 2023.

After the interviews, the Thematic Analysis (TA) method was used to analyze them at two levels of concept and category.

TA is a method for systematically identifying, organizing, and offering insight into patterns of meaning (themes) across a data set. Through focusing on meaning across a data set, TA allows the researcher to see and make sense of collective or shared meanings and experience (Braun and Clarke 2012). An example of coding in the TA method is presented in Table 2.

Table 2. A sample of the Thematic Analysis (TA) method (coding of interviews)

<i>Codes</i>	<i>Sentence examples</i>	<i>Concept</i>	<i>Category</i>
Code 1	"By the type of such works (audio, video, audiobook, and e-book), they would mostly be licensed and so by way of contract ..." (Code 1).	Existence of a license and contract	license and contract
Code 2	"Clear information on copyrights and licenses and/or permissions, maintained alongside digital library content to which they apply ..." (Code 2).	Getting licenses/ permissions for the content	license agreement
Code 6	"Laws and statutes should actually be made to protect digital resources. We need to have information security management policies to prevent internal and external attacks".	information security management policies	Formulating information security management policies

In the fourth step, the technical factors were initially extracted. The new categories extracted from the interviews, which were related to the technical factors of the systems, were added to the checklist. For the sake of reliability, the list of indicators were presented to two experts in digital libraries and ISs, who confirmed all the indicators. Then, Cronbach's alpha was used to measure the reliability of the indicators. The obtained score was 0.99, indicating the checklist's high reliability. SPSS 24 was used for Cronbach's alpha. In the fifth step, after confirming the indicators, the fuzzy Delphi method was used for their validation and group consensus. A 5-point Likert scale was used in the above checklist to examine the effectiveness of technical factors, with 1 indicating very low and 5 indicating very high effectiveness. The fuzzy Delphi method was conducted in two rounds. A panel of 11 experts and specialists with good work experience in ISs and digital libraries and accessibility were selected (Table 3).

Table 3. Characteristics of experts selected as members of the Delphi panel

Row	Level of Education	Area of Expertise	Work Experience (Year)	First Round	Second Round
1	Ph.D. in Communication	Digital libraries	25	✓	✓
2	Ph.D. in Information Technology	Information systems	21	✓	✓
3	Ph.D. in Knowledge and Information Science	Digital libraries	20	✓	✓
4	Ph.D. in Knowledge and Information Science	Digital libraries	20	✓	✓
5	BS in Hardware Engineering	Information systems	30	✓	✓
6	MS in Information Science and Epistemology	Information systems	10	✓	✓
7	Ph.D. in Knowledge and Information Science	Digital libraries	15	✓	✓
8	Ph.D. in Knowledge and Information Science	Digital libraries	10	✓	✓
9	MS in Information Technology	Digital libraries	17	✓	✓
10	MS in Computer and Networks	Information systems	17	✓	✓
11	MS in Knowledge and Information Science	Information systems	17	✓	✓
Total				11	11

The checklist was presented to the experts to receive their opinions. The members of the Delphi panel expressed their opinions through verbal variables (Table 4). The numbers were fuzzified by converting the triangular fuzzy numbers equivalent to the 5-point Likert scale into fuzzy numbers.

Table 4. Triangular fuzzy numbers equivalent to the 5-point Likert scale

<i>Triangular Fuzzy Numbers</i>	<i>Linguistic Terms</i>
(0, 0, 0.25)	Very Low
(0, 0.25, 0.5)	Low
(0.25, 0.5, 0.75)	Average
(0.5, 0.75, 1)	High
(0.75, 1, 1)	Very High

Then, the fuzzy average was calculated from the scores using the equation and converted into a definite number using the andequations. The results of all fuzzification calculations are given in Tables 5 and 6. The threshold value was considered 0.5 in this study.

In the fuzzy Delphi method, which has a screening mode, variables with a minimum agreement of 50% are confirmed (Habibi, Firouzi Jahantigh, and Sarafrazic 2015).

Fuzzy Delphi analysis was performed for all accepted indicators in two rounds, where 16 indicators were confirmed based on the opinions of 11 experts (Tables 5, 6 and 7). If the difference between the mean scores of the first and second-round questions is less than 0.2, the survey process will be stopped (Habibi and Afaridi 2022). None of the indicators scored less than 0.2, suggesting that all 16 indicators were confirmed. The above indicators were placed in 5 components.

In the first round, the indicator “Using different firewalls and remote access to protect (secure) the server” with an average of 0.841 related to network security ranked first among experts. Then, the indicator “Formulating information security management policies to protect digital resources” with an average of 0.795 in the standards and frameworks component ranked next. Additionally, the indicator “Using proxy servers to receive resources and apply access restrictions” with an average of 0.765 related to network security ranked third among experts.

In the second round of the Delphi panel, the indicator “Using different firewalls and remote access to protect (secure) the server” again ranked first with an average of 0.826, and the indicator “Undocumented (unlogged) access” with an average of 0.811 related to standards and frameworks ranked second. Furthermore, the indicator “Formulating information security management policies to protect digital resources” also again ranked third among experts in this round with an average of 0.841 (Table 5, 6 and 7).

Table 5. Results for the first round of fuzzy Delphi

<i>Dimensions</i>	<i>Components</i>	<i>Indicators</i>	<i>Triangular fuzzy average</i>			<i>De-fuzzified average</i>	<i>Status</i>
			<i>l</i>	<i>m</i>	<i>u</i>		
Technical factors	Network Security (4)	Not using a firewall (failure to use a firewall)	0.455	0.705	0.864	0.674	Confirmed
		Using different firewalls and remote access to protect (secure) the server	0.636	0.886	1.000	0.841	Confirmed
		Blocking access to data transfer from the server for system employees	0.409	0.659	0.864	0.644	Confirmed
		Using proxy servers to receive resources and apply access restrictions	0.545	0.795	0.955	0.765	Confirmed
	Technical Infrastructure (3)	Outdated Digital Rights Management (DRM) software	0.409	0.659	0.864	0.644	Confirmed
		Limited computational space (servers, software, etc.) provided by the systems	0.455	0.705	0.909	0.689	Confirmed
		Limited temporary storage space	0.455	0.705	0.932	0.697	Confirmed
	Copy control (4)	Costly audio/video transcription algorithms	0.318	0.568	0.795	0.561	Confirmed
		Not using “look inside” to access pages and different parts of digital content	0.409	0.659	0.864	0.644	Confirmed
		Using “look inside” for text, audio, and video formats	0.455	0.705	0.886	0.682	Confirmed
		Copy (plagiarism) detection systems	0.477	0.705	0.864	0.682	Confirmed
	Frameworks and standards (3)	Undocumented (unlogged) access	0.500	0.750	0.932	0.727	Confirmed
		Formulating information security management policies to protect digital resources	0.591	0.841	0.955	0.795	Confirmed
		Applying Metadata Encoding and Transmission Standard (METS)	0.455	0.705	0.886	0.682	Confirmed
	License (2)	Lacking publisher’s license agreement to publish audio-books on the systems	0.545	0.795	0.932	0.758	Confirmed
		Smart contracts	0.523	0.750	0.886	0.720	Confirmed

Table 6. Results for the second round of fuzzy Delphi

<i>Dimensions</i>	<i>Components</i>	<i>Indicators</i>	<i>Triangular fuzzy average</i>			<i>De-fuzzified average</i>	<i>Status</i>
			<i>l</i>	<i>m</i>	<i>u</i>		
Technical factors (5)	Network Security (4)	Not using a firewall (failure to use a firewall)	0.500	0.750	0.955	0.735	Confirmed
		Using different firewalls and remote access to protect (secure) the server	0.614	0.864	1.000	0.826	Confirmed
		Blocking access to data transfer from the server for system employees	0.455	0.705	0.932	0.697	Confirmed
		Using proxy servers to receive resources and apply access restrictions	0.477	0.727	0.909	0.705	Confirmed
	Technical Infrastructure (3)	Outdated Digital Rights Management (DRM) software	0.477	0.727	0.909	0.705	Confirmed
		Limited computational space (servers, software, etc.) provided by the systems	0.500	0.750	0.955	0.735	Confirmed
		Limited temporary storage space	0.477	0.727	0.932	0.712	Confirmed
	Copy control (4)	Costly audio/video transcription algorithms	0.364	0.614	0.818	0.598	Confirmed
		Not using “look inside” to access pages and different parts of digital content	0.432	0.682	0.909	0.674	Confirmed
		Using “look inside” for text, audio, and video formats	0.432	0.682	0.909	0.674	Confirmed
		Copy (plagiarism) detection systems	0.568	0.818	1.000	0.795	Confirmed
	Frameworks and standards (3)	Undocumented (unlogged) access	0.591	0.841	1.000	0.811	Confirmed
		Formulating information security management policies to protect digital resources	0.636	0.886	1.000	0.841	Confirmed
		Applying Metadata Encoding and Transmission Standard (METS)	0.364	0.614	0.818	0.598	Confirmed
	License (2)	Lacking publisher’s license agreement to publish audiobooks on the systems	0.568	0.818	0.955	0.780	Confirmed
		Smart contracts	0.455	0.705	0.932	0.697	Confirmed

Table 7. Comparison of the first and second rounds

<i>Indicators</i>	<i>De-fuzzified average (first round)</i>	<i>De-fuzzified average (second round)</i>	<i>Difference</i>	<i>Status</i>
Not using a firewall (Failure to use a firewall)	0.674	0.735	0.061	Confirmed
Using different firewalls and remote access to protect (secure) the server	0.841	0.826	-0.015	Confirmed
Blocking access to data transfer from the server for system employees	0.644	0.697	0.053	Confirmed
Using proxy servers to receive resources and apply access restrictions	0.765	0.705	-0.061	Confirmed
Outdated Digital Rights Management (DRM) software	0.644	0.705	0.061	Confirmed
Limited computational space (servers, software, etc.) provided by the systems	0.689	0.735	0.045	Confirmed
Limited temporary storage space	0.697	0.712	0.015	Confirmed
Costly audio/video transcription algorithms	0.561	0.598	0.038	Confirmed
Not using “look inside” to access pages and different parts of digital content	0.644	0.674	0.030	Confirmed
Using “look inside” for text, audio, and video formats	0.682	0.674	-0.008	Confirmed
Copy (plagiarism) detection systems	0.682	0.795	0.114	Confirmed
Undocumented (unlogged) access	0.727	0.811	0.083	Confirmed
Applying Metadata Encoding and Transmission Standard (METS)	0.682	0.598	-0.083	Confirmed
Lacking publisher’s license agreement to publish audiobooks on the systems	0.758	0.780	0.023	Confirmed
Smart contracts	0.720	0.697	-0.023	Confirmed
Formulating information security management policies to protect digital resources	0.795	0.841	0.045	Confirmed

After validating and confirming the indicators, the seven information systems were evaluated by the researcher and the managers of the systems. Two options “Yes” and “No” were used in the checklist. The research population comprised the National Digital (Digital Library), Iran publications systems, Iranian Scientific Publications System, Iran Publications Database, National Document Center Network, Catalog of Manuscripts, and Iran Libraries Network. The first three are considered digital systems, while the rest are placed in the bibliographic category because they provide information on information resources and are of the type of database. The authority and credibility of the Iranian National Library among other libraries, its legal duty to collect all library resources including digital materials, and the diversity of its ISs were among the reasons for selecting it and the NLAI as the research population.

In analysis, all fuzzy Delphi calculations were performed in Excel 2016. Moreover, the methodological triangulation method was used to validate the qualitative findings. This method was employed because using different methods and tools for data collection, such as documentary analysis, thematic analysis, fuzzy Delphi and interview, observation, and checklist, gives further validity to qualitative research.

The most common way of triangulating with method is using both qualitative and quantitative data collection instruments and procedures and data analysis in a single study. Another way of triangulating with the method is to use in-method triangulation. As the name suggests, multiple approaches within the same research approach (qualitative or quantitative) were employed in a study. For example, in one study, the researcher collected data from documentary analysis, observation, and focus group interviews (Donkoh and Mensah 2023).

Before answering the main research question and presenting the results, we need to explain the studied systems briefly.

3.1. What are ISs of NLAI?

The existing systems in the NLAI provide users with information in the form of bibliographic and full-text ISs. Accordingly, of the 7 ISs, only the digital resources system (National Digital) provides the full text of some resources in addition to bibliographic and metadata information. Other systems currently provide only bibliographic and metadata information on the resources. One of the most practical ISs in libraries is the digital library, which most Iranian libraries have designed and implemented.

In general, the web-based ISs of the NLAI are the National Digital (Digital Library), Iran Publications systems, Iranian Scientific Publications System, Iran Publications Database, the National Document Center Network, the Catalog of Manuscripts, and Iran Libraries Network.

Iran Libraries Network, the National Document Center Network, Catalog of Manuscripts, and the Iran Publications Database are provided through the National Iranian Memory (NIM). These resources were set up to facilitate fair access to all library resources, manuscripts, documents, and journal articles in Iran. This platform makes it possible to search in each of the systems separately, e.g., searching book resources through the Iran Libraries Network, searching documents through the National Document Center Network, searching manuscripts through the Catalog of Manuscripts, and searching journal articles through the Iran Publications Database (Search guide in the National Document Center Network n.d.). So far, as many as 3,000 active academics, public, and specialized libraries in the Library Network and 64,000 articles in the Iran Publications Database are available on this platform. Also, it includes the list of documents available in 7 archival centers in the National Document Center Network, a list of 1 million Persian and non-Persian manuscripts in Iran, and the manuscripts of Iraq and the Persian manuscripts of countries in 1868 centers in the Catalog of Manuscripts. In fact, the NIM guides the researchers and the audience to documentary and library collections, enabling them easier access to information resources (National Iranian Memory n.d.).

The Digital Resources System (National Digital Library) has millions of documents, manuscripts, lithographic and lead-printed books, periodicals, theses and dissertations, reports, research plans, photos, maps, etc. Therefore, it enables users to access these resources through the library and digital documents. Users are provided with services after the on-site and online collection, selection, conversion, and processing of the resources in the library and digital documents. Users can browse and search for the resources, see the in-line delivery of the resources, and print their content. The NLAI has adopted this measure for the long-term protection and greater accessibility of the written heritage of Iranians.

Iran publications systems includes publications relate to the historical periods of Qajar, Pahlavi, and the Islamic Revolution. The periods of these publications are: 1) publications of the Qajar period (1193-1304 Lunar Calendar / 1779-1926 Gregorian Calendar); 2) publications of the first Pahlavi period (1304-1320 SH/ 1925-1941 Gregorian Calendar); 3) publications of the second Pahlavi period (1320-1357 SH/ 1941-1978 Gregorian Calendar); 4) publications after the Islamic Revolution (1357 SH/ 1957 Gregorian Calendar); and 5 Latin publications (Iran publications systems 2023). In the first stage, member users have access to the full text of some publications, and to receive the full text, they must send their request to the website administrator. There is no difference in terms of access type (including viewing or downloading) on this system. The publications of the Qajar period are related to the last 100 years. Considering the copyright protection period (up to 50 years after the author's death)⁸, some publications can be made available to the public, and the download and viewing rights can also be given to member users.

Iran Publications Database as many as 61,064 articles are available on this website (Iran Publications Database 2023). It gives no information about the variables, such as articles and historical periods; instead, it merely gives a search guide. It lists the Iranian newspapers and magazines from the beginning until now. It also makes it possible to comment, complete, and participate in data correction (National Iranian Memory n.d.).

As mentioned above, the NIM consists of 4 bibliographic databases. It uses the Oracle database and its indexing with Lucene. Its profile is similar to the Rasa software of the National Library of Iran⁹. Oracle is a relational database that uses the Structured Query Language (SQL).

Iranian Scientific Publications System like those of other systems, have been compiled based on the main approach of the organization. Its publications include those written in English, Russian, Arabic, Persian, and French. The information of a publication is provided, and the system is searched based on the publication's date, author's name, title, and the subject of the publication (Iranian Scientific Publications System 2023). Accessing the full text and viewing the articles is possible through a link given to the publication in the system. Therefore, in terms of the type of access, it is possible to view and download some resources only in the Digital Library (including manuscripts and lithographs, theses and research projects), Iranian Scientific Publications System, and Iran publications systems.

4. Results

What are the technical factors affecting the copyright protection of information resources in the systems under study?

With the explanations given about the systems under study, the main question of the research, i.e., the technical factors affecting the copyright protection of information resources in the systems under study, is answered as follows:

⁸ To get more information about the protection period of the works in the Law on the Protection of the Rights of Authors, Writers, and Artists, see the following two links: <https://rc.majlis.ir/fa/law/show/788903> and <https://rc.majlis.ir/fa/law/show/96427>.

⁹ Messages from Hassan Bagheri (system administrator) to author, 31 August 2022.

As shown in Tables 5, 6 and 7. and based on the fuzzy Delphi method, 16 indicators were identified as the technical factors affecting the systems. Here, the greater the de-fuzzified mean, the more agreement there is on the indicator.

The obtained indicators were identified as effective technical factors in both bibliographic and digital systems. Not using a proper firewall (failure to use a firewall), using different firewalls and remote access to protect (secure) the server, blocking access to data transfer from the server for system employees, using proxy servers to receive resources and apply access restrictions, outdated DRM system, limited computational space (servers, software, etc.) provided by the systems, limited temporary storage space, costly audio/video transcription algorithms, not using “look inside” to access pages and different parts of digital content, using “look inside” for text, audio, and video formats, copy (plagiarism) detection systems, undocumented (unlogged) access, applying the METS Metadata, lacking publisher’s license agreement to publish audiobooks on the systems, and smart contracts related to the license component are important for the ISs of libraries. Evaluating the studied systems also showed that only the two technical factors of “not using a firewall” and “using different firewalls and remote access to protect (secure) the server” were used in the bibliographic systems to prevent intrusion into the systems and maintain network security.

In the digital systems, the indicators “Blocking access to data transfer from the server for system employees” and “using different firewalls and remote access to protect (secure) the server” were used. Besides, according to the manager of digital systems, the METS Metadata was largely used. Other technical indicators, effective as technical factors in protecting the copyright of information resources, were not used in the studied systems. Therefore, based on the fuzzy Delphi outputs, the de-fuzzified average of the indicators shows a consensus among the experts over all the indicators. However, the two indicators of “using different firewalls and remote access to protect (secure) the server” and “formulating information security management policies to protect digital resources” were more effective than the other technical factors considering their de-fuzzified average. Fig. 1 depicts the technical factors affecting copyright protection in the studied ISs. As mentioned above, the indicators were: network security, licensing, frameworks and standards, technical infrastructure, and copy control.

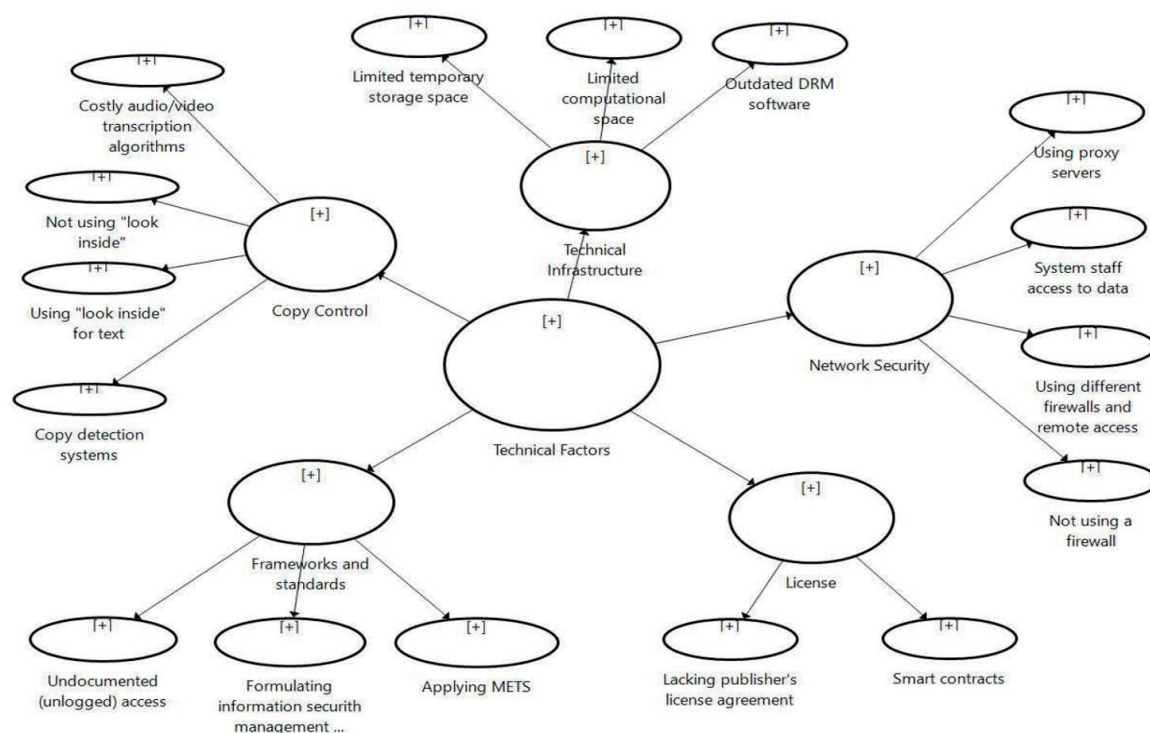


Fig. 1. Technical factors affecting copyright protection in the studied systems (bibliographic and digital)

5. Discussion

The current study aimed to identify the technical factors affecting the copyright protection of information resources in the ISs of the NLAI. Documentary and thematic analysis methods were used to review and analyze the resources and studies. In addition, the fuzzy Delphi method was used to validate the indicators and thus evaluate the studied systems. According to experts and specialists, some indicators are effective as technical factors in the studied systems. Overall, 16 indicators were identified as technical factors affecting the copyright protection of information resources in the studied ISs.

The results analysis revealed that the indicators “not using a firewall”, “using different firewalls and remote access to protect (secure) the server”, “Blocking access to data transfer from the server for system employees”, and “using proxy servers to receive resources and apply access restrictions” were effective in maintaining the security of the network and protection of copyright in the studied systems. This issue was emphasized in an interview with experts:

“What factors can be effective in the copyright protection of the systems? We should see where the servers are located; they must be located in the protected zone of the organization. Also, they must have firewalls. There are different types of firewalls. There must be different remote access” (Code 6).

“Penetration testing must be done for the system to identify the security holes that may exist. We checked many systems, and they were not really safe. I think the system itself must be secure, and we must have an

ISMS in the organization. Developing ISMS in the departments with content electronic services should be a security investment” (Code 6).

The use of indicators such as “network security” and “the use of firewalls and routers” was already emphasized by Ismail and Zainab (2011), Khan, Ibrahim, and Hussain (2021), and Eidi Ghaleshiri (2016). In the current research, network security was also discussed as an important technical factor in information systems. This factor can play a role in preventing attacks from hackers and malware.

Obtaining a license and an agreement with the publishers can considerably help overcome the challenges related to the copyright protection of the resources. In the current study, the lack of license was identified as a technical factor affecting the performance of the systems under study. Considering the nature of the resources, the need for a license is a requirement obtained from the owner of the material and intellectual rights of the work based on the contract. Obtaining license must occur by the libraries before the digitization of resources and uploading of resources to information systems. In such a way that the type of license and access is clearly specified in the information systems. License and access can be provided in the form of an information content package for each of the systems.

Another non-Iranian expert mentioned the licenses of Creative Commons and other licenses:

“If a Creative Commons or other license has been declared for particular material on the Archive, to the extent you trust the declaration and declarer (which is rarely the Internet Archive), you may use the content according to the terms and conditions of the applicable license” (Code 3).

Using Creative Commons licenses and access to articles have become common among scientific journals in Iran for several years. Issuance of digital licenses, giving licenses, and licensed agreements mentioned in this study are consistent with the views raised in the studies of Kokabi and Kohi Rostami (2015), Fernández-Molina (2004), and Griffin (2013) since obtaining permission from publishers reduces the risk of copyright infringement of information resources. While considering the need to get written licenses, Gaber (2012) criticizes that the DRM system has not been used in the systems investigated in the current study. Another technical factor affecting the copyright protection of information resources in the studied systems is the use of “look inside”, which has been ignored in the mentioned systems. This important point was expressed in an interview with one of the experts in this field:

“For some theses, we said, for example, the readers can view pages 4 to 14, which becomes ten pages. Our “Look Inside” required that only the first ten pages or the second ten pages be displayed, which was one of the technical factors that had to be developed. In the “Look Inside” that the item is part of the development plan, we said that we would specify the item access level, i.e., the access level of the item itself, which again required a higher development level processing, for example, to say that all the items can be accessed separately. Thus, we can create some restrictions for all items” (Code 7).

The use of “Look Inside” and the segmentation of images related to digital content in the mentioned digital systems can prevent illegal downloads and uses to a large extent. Not using “Look Inside” for accessing different pages and sections of digital content that were controlled by copy-

right. The use of “Look Inside” in digital libraries were employed for previewing book content. It also enables searching within the content of books.

Other technical factors such as smart contracts, METS metadata, and access documentation are among other technical factors affecting the copyright protection of information resources. The use of smart contracts and METS Metadata was mentioned in an interview with digital library specialists:

“We wanted to assign a part of a content to its copyright notice. We considered composite content for our implementation. But let us mention this in the more developed world. The issue of METS should be considered as well. Yet, it’s much more developed form is the smart contract, which enters the digital asset world. Although it is a digital asset for which there should be a smart contract, we did not use it in our analysis. These are the technical factors of it” (Code 7).

“A part that makes a great contribution is the documentation issue; every file or database should have access to a document. Such documents should be transferred, and we need to have knowledge management to transfer them. In this way, the security issue may not be challenged. Usually, files that are stored for a long time and are forgotten by a technician like me, and there is no writing anywhere. Consequently, we do not know how to access it, which becomes a problem” (Code 5).

Another technical factor mentioned in this study is the copy detection system, an instance of which is a concordance system. This point was outlined in an interview with a specialist in Iran’s ISs:

“I think one of the most effective factors is the next steps. For example, in scientific works, they say that we should use a concordancer to see how much plagiarism has occurred in a work. They use applications and have huge databases to check a work for plagiarism. In my opinion, these are technical requirements that must exist and be observed” (Code 4).

Papi et al. (2017) also emphasized the need for copy-detection systems, which is consistent with the current study’s findings.

Another indicator and effective technical factor mentioned in the current study is “formulating information security management policies to protect digital resources of information”, which is neglected in many digital libraries and ISs. According to Farid, Warraich, and Iftikhar (2023), Peykam and Salimifard (2016), and Rezvani (2018), there is no digital information security management policy or security policy in some university libraries. In Iran, like the current study, and it is in line with the findings of Papi’s study (2021), confirming the lack of a copyright policy in Iranian libraries. Khan, Ibrahim, and Hussain (2021) evaluated the university library system in Pakistan regarding data and information security. These researchers indicated an unfavorable status, which is consistent with the current study’s findings. In the above studies, issues such as the lack of copyright policy have been highlighted.

The subject of content protection and the role of human resources in it has been highlighted in various studies, such as Rezvani (2018) and Doroudi and Jamshidi (2021), since protecting content and human resources are among the technical factors that affect the copyright in the digital libraries.

6. Conclusion

In general, technical factors are effective in developing and protecting authors' copyright affairs in information systems in libraries. The studied systems (digital and bibliographic) are not favorable regarding copyright protection of information resources. Therefore, the managers of such systems must take necessary measures to review the technical checklists to evaluate the systems in terms of respecting and protecting the copyright of information resources. In addition, they are required to improve the effective technical factors in these systems, thereby increasing the trust of the creators of the works in the libraries.

The technical factors that were discussed in the present study were: "not using a firewall", "using different firewalls and remote access to protect (secure) the server", "Blocking access to data transfer from the server for system employees", "using proxy servers to receive resources and apply access restrictions", "Obtaining a license and an agreement", "look inside", "smart contracts", "METS metadata", "access documentation", "copy detection system" and so on.

It seems that focusing on network security measures and developing standards and technical frameworks for better content protection can have a greater impact on copyright protection in systems compared to other indicators. Because the absence of any of these indicators can affect the violation of the rights of the creator of the work. Therefore, the attention to the technical factors mentioned in the present research can reduce the risk of copyright infringement in information systems and digital libraries. Undoubtedly, it is not possible to support the authors' copyright in the information systems of libraries without using technical measures.

6.1 Practical implications

According to the discussion regarding the findings of this study, the following are some of the practical implications that this study may have for libraries, information centers, and policymakers to ensure that libraries have the capacity to help maintain copyright of information resources while securing ISs. Therefore, it was essential for libraries to utilize technical standards to ensure the security of systems before implementing ISs and digitization processes. Some suggestions were made in this study for this purpose. Libraries should strive to enhance the relationship between copyright laws and technical issues to better implement technical infrastructures in ISs. The current research can offer a new approach for interaction between legal and technical discussions in libraries.

6.2 recommendations

After answering the research question, some suggestions for respecting copyright in ISs, both in current digital systems and for the future and development of bibliographic systems, were proposed. The some of following suggestions were the result of data analysis and can be utilized by libraries and information centers.

- Use of "look inside" for digital, audio, and video resources;
- Development of a copyright policy for all digital resources/use of legal metadata for digital resources;

- Documenting all accesses; documenting the type of access based on the type of resource and its publication date;
- Use of web services;
- Use of firewalls and various remote access methods for server protection;
- Use of proxy servers to obtain resources and impose access restrictions.

References*

- Braun, Virginia, and Victoria Clarke. 2012. "Thematic analysis." In *APA handbook of research methods in psychology*, edited by Harris Cooper, Paul M. Camic, Debra L. Long, Abigail T. Panter, David Rindskopf, and Kenneth J. Sher, vol. 2, *Research designs: Quantitative, qualitative, neuropsychological, and biological*, 57–71. Washington DC: American Psychological Association. <https://doi.org/10.1037/13620-004>.
- Donkoh, Sylvester, and John Mensah. 2023. "Application of triangulation in qualitative research." *Journal of Applied Biotechnology and Bioengineering* 10 (1). [10.15406/jabb.2023.10.00319](https://doi.org/10.15406/jabb.2023.10.00319).
- Doroudi, Fariborz, and Zeinab Jamshidi. 2021. "Assessing the Components of Information Security in Accessing & Use of Digital Libraries." *Iranian Journal of Information Processing & Management* 37 (1). [10.52547/JIPM.37.1.117](https://doi.org/10.52547/JIPM.37.1.117).
- Eidi Ghaleshiri, Davoud. "Evaluation of the security challenges of designing digital libraries (case study: Astan Qods Razavi Digital Library)." Master's thesis, Gilan University, 2016.
- Farid, Ghulam., Nosheen Fatima Warraich, and Sadaf Iftikhar. 2023. "Digital information security management policy in academic libraries: A systematic review (2010–2022)." *Journal of Information Science* 1-15. [10.1177/01655515231160026](https://doi.org/10.1177/01655515231160026).
- Ferdinand, Omosekejimi Ademola, Osaze Patrick Ijiekhuamhen, and Ojeme Thelma Nneka. 2015. "Library and information resources' security: traditional and electronic security measures." *International Journal of Academic Research and Reflection* 3 (3). <https://www.idpublications.org/wp-content/uploads/2015/02/LIBRARY-AND-INFORMATION-RESOURCES'-SECURITY-TRADITIONAL-AND-ELECTRONIC-SECURITY-MEASURES.pdf>.
- Fernández-Molina, Juan-Carlos. 2004. "Licensing Agreements for Information Resources and Copyright Limitations and Exceptions." *Journal of Information Science* 30 (4). doi.org/10.1177/0165551504045853.
- Gaber, Tarek. "Support Consumers' Rights in DRM: A Secure and Fair Solution to Digital License Reselling over the Internet." Ph.D. thesis, Department of Computer Science, The University of Manchester, 2012.
- Griffin, James. 2013. "The Digital Copyright Exchange: Threats and Opportunities." *International Review of Law, Computers & Technology* 27 (1-2). <https://doi.org/10.1080/13600869.2013.764136>.
- Habibi, Arash, and Sanam Afaridi. 2022. *Multiple attribute decision making*. Tehran: Narvan Publications.
- Habibi, Arash, Farzad Firouzi Jahantigh, and Azam Sarafrazi. 2015. "Fuzzy Delphi Technique for Forecasting and Screening Items." *Asian Journal of Research in Business Economics and Management* 5 (2). [10.5958/2249-7307.2015.00036.5](https://doi.org/10.5958/2249-7307.2015.00036.5).
- Ismail, Roesnita, and A.N. Zainab. 2011. "Information systems security in special and public libraries: an assessment of status." *Malaysian Journal of Library & Information Science* 16 (2). <https://ejournal.um.edu.my/index.php/MJLIS/article/view/6697/4379>.

* Websites were last accessed on 5 November 2024.

- Khan, Arif, Muhammad Ibrahim, and Abid Hussain. 2021. "An exploratory prioritization of factors affecting current state of information security in Pakistani university libraries." *International Journal of Information Management Data Insights* 1 (2). <https://doi.org/10.1016/j.jjime.2021.100015>.
- Kokabi, Morteza, and Mansor Koi Rostami. 2015. "Information security of Web-based systems in Iran Institution of public libraries." *Research on Information Science and Public Libraries* 21 (1). <http://publii.ir/article-1-1077-en.html>.
- Muir, Adrienne . 2006. "Preservation, Access and Intellectual Property Rights Challenges for Libraries in the Digital Environment." <https://hdl.handle.net/2134/2181>.
- National Iranian Memory. n.d. *Pamphlet of the National Iranian Memory (NIM)*. Tehran: National Library and Archives of Iran.
- Niqresh, Mohammad. 2019. "Digital Library and Intellectual Issues – Issues in Copyright and Intellectual Property." *International Education Studies* 12 (1). <https://doi.org/10.5539/ies.v12n1p114>.
- Papi, Zeinab. 2021. "Manual on Moral and Economic Rights of Print and Digital Library Resources in Iran." Research Project, Ministry of Justice: National Library and Archives of Iran, Tehran.
- Papi, Zeinab, Saeed Rezaei Sharifabadi, Sedigheh Mohammadesmaeil, and Nadjla Hariri. 2017. "Technical requirements for copyright protection of electronic theses and dissertations in INSTED: a grounded theory study." *The Electronic Library* 35 (1): 21-35. <http://dx.doi.org/10.1108/EL-11-2015-0226>.
- Peykam, Alireza, and Khodakaram Salimifard. 2016. "A Framework for Analysis of Inter-Organizational Factors Affecting on the Security of Information Systems using FAHP Approach." *Business Intelligence Management Studies* 4 (16): 147-176. [10.22054/ims.2016.6987](https://doi.org/10.22054/ims.2016.6987).
- Rezvani, Shahla. 2018. "Designing an Information Security Management Model in Digital Libraries." *Library and Information Science Research Journal* 8 (1): 337-356. [10.22067/riis.v0i0.61486](https://doi.org/10.22067/riis.v0i0.61486).
- Wahid, Ratnaria. 2017. "Libraries and technology: Canadian and Malaysian copyright exceptions." *Library Management* 38 (8-9): 415-425. <https://doi.org/10.1108/LM-02-2017-0022>.

Indexing clinical documents for long-term preservation through an international coding system*

Erika Pasceri^(a)
Maria Teresa Chiaravalloti^(b)

a) University of Calabria, <https://orcid.org/0000-0003-4440-2266>

b) IIT-CNR, <https://orcid.org/0000-0003-4695-2026>

Contact: Erika Pasceri, erika.pasceri@unical.it; Maria Teresa Chiaravalloti, maria.chiaravalloti@iit.cnr.it

Received: 16 October 2024; **Accepted:** 03 March 2025; **First Published:** 15 May 2025

ABSTRACT

This paper describes the experience of mapping the document types defined by Medas company, a Digital Preservation Organization for healthcare institutions, to the Logical Observation Identifiers Names and Codes (LOINC) international standard, and consequently assesses its adequacy in representing specificities of the Italian context. Mapping operations were manually performed by LOINC Italia experts. The LOINC database was searched using the LOINC Search web browser and REgenstrief LOINC Mapping Assistant (RELMA) software, employing local names and synonyms of Medas document types. Out of 483 Medas document types, 144 fully match with a LOINC code; 211 were associated with a LOINC code which only partially covers the meaning; 128 are not covered by LOINC. Although the axes of LOINC Document Ontology are quite adequate for document types representation, an extension of Kind of document and Subject Matter Domain axes is desirable. Local Medas codes reflect the variety of clinical document types produced in the Italian healthcare domain and the results of this case study can be used as new terms submissions to enrich the LOINC standard, especially in relation to the specificities that characterize the Italian national context.

KEYWORDS

LOINC; Clinical documents; Long-term preservation; Mapping; LOINC Document Ontology.

* Authors all contributed in the realization of this paper, although Erika Pasceri specifically dealt with subsection “Introduction”, “Material and Method” and Maria Teresa Chiaravalloti dealt with “Background and significance”, “Results” and “Discussion” and together worked on “Conclusion”.

© 2025, The Author(s). This is an open access article, free of all copyright, that anyone can freely read, download, copy, distribute, print, search, or link to the full texts or use them for any other lawful purpose. This article is made available under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. JLIS.it is a journal of the SAGAS Department, University of Florence, Italy, published by EUM, Edizioni Università di Macerata, Italy, and FUP, Firenze University Press, Italy.

Introduction

Long-term preservation is the activity aimed at protecting and safeguarding archives of documents and computer data over time. It is a mandatory task in Public Administration in Italy since the introduction of the Law Decree n. 82/2005 named *Codice dell'Amministrazione Digitale* (CAD). Long-term preservation systems have to guarantee authenticity, integrity, reliability, readability and availability of electronic documents over time (CAD, art. 44). Public Administration can choose to implement an inhouse or outsourcing long-term preservation service.

Actually, it introduced a gradual digitization of the information flows in public administrations by re-engineering processes and consequently promoting a new document approach from a structural point of view and about the semantic management of their content. In this context, health documents are of particular interest, both for the nature of the data they contain and for the requirements introduced by the advent of the Italian Electronic Health Record (EHR) system, namely Fascicolo Sanitario Elettronico (FSE).

FSE was established by the Law Decree n. 179/2012 and subsequently detailed by the Prime Minister Decree n. 178/2015, also in compliance with the European Directive 2011/24/EU concerning the application of patients' rights to cross-border healthcare. The FSE is a federated infrastructure, which enables interoperability services among regional EHR systems, such as, primarily, clinical documents search, retrieve and exchange.

FSE allows to overcome technological barriers of the regional and proprietary EHR systems to trace the entire medical history of a patient, so ensuring healthcare professionals the chance to access clinical documents produced by different stakeholders in order to provide efficient clinical care and pathways. Its architecture is based on a system of registries and repositories. Each Region has a registry, which indexes metadata of the clinical documents of the patients assisted in the Region itself (even if produced by facilities located in other Regions) and can choose to have one centralized repository or more distributed repositories to store documents produced by the healthcare facilities of the Region itself (even if indexed in the registry of another Region) (Frazier et al. 2001; Shapiro et al. 2005; De Pietro et al. 2011). Clinical data of the documents indexed in the FSE have to be represented by using coding and classification systems to ensure, eventually by transcoding, semantic interoperability among the various regional, national and European e-health solutions. Among minimum metadata required for indexing documents in the FSE there is the description of the type of clinical document, which has to be uniquely identified through codes belonging to the standard LOINC (Logical Observation Identifiers Names and Codes). The subset of LOINC codes currently used for this purpose is listed into the Italian Affinity Domain (AD) document.¹ Nonetheless, private organizations, which provide outsourcing services of document long-term preservation, find it insufficient compared to the need to have uniquely identified a great number of different clinical document types produced by healthcare facilities. Furthermore, the names of the clinical documents that they receive for long-term preservation are often assigned by the authors themselves or at most determined by institutional workflows, so they are most of the time not consistent when pooled.

The LOINC codes identifying clinical documents' types are part of the Document Ontology

¹ More info at https://www.fascicolosanitario.gov.it/sites/default/files/public/media/AffinityDomainItalia_v2-3.pdf.

(DO) of the standard, which was born from the initial effort of the Document Ontology Task Force (DOTF) (Dugas and Thun 2009), composed by members of both the US Veterans Health Administration and Health Level Seven (HL7), aiming at creating a system for clinical document naming. They considered a similar previous work of the European Committee for Standardization (CEN) resulting in the publication of ENV 13606-2, which is a domain term list for the electronic healthcare record communication. However, CEN document categories were still too generic and needed to be used in combination with local codes. On the contrary, DOTF aimed to produce a fully specified name. They started from the empiric analysis of over 2,000 documents and defined a structure based on six axes to have the chance of non-ambiguously combining concept names so that they are unique for document retrieval. The intent was to create a vocabulary to represent and communicate knowledge about clinical documents and define a set of relationships among its terms (Frazier et al. 2001).

Moving from these considerations, the objective of this work is mapping the document types managed by one of the major Digital Preservation Organization of clinical documents² towards the LOINC standard to assess the semantic coverage of the standard and consequently identify candidates for LOINC new terms submissions. Results of this work could be also used to enlarge the list of LOINC document type codes used to index documents into the FSE according to the AD.

Background and Significance

The LOINC DO is currently the domain reference standard, but it is worth mentioning similar effort of Brown et al. (2001) in creating a Document-naming Nomenclature (DNN) for the Computerized Patient Record System (CPRS) of the Veterans Affairs (VA) medical centers. They specified three document identifier categories to be populated with values from a nomenclature. Nonetheless their system was not precise enough in defining subspecialties.

Several studies investigated the LOINC DO coverage and usability to correctly identify clinical document types. Hyun et al. (2006) focused on the DO adequacy for representing nursing documents. Ninety-four document types were collected from the Eclipsys Clinical Information System (CIS) and the Medical Entities Dictionary (MED) of the New York Presbyterian Hospital. They were mapped considering primarily the specific subset of nurse terms in the DO but extending to others if the concept was present in LOINC but not specifically designed for nursing. As results 51.1% of terms were mapped to an exactly equivalent concept, 28.7% were classified as *not specified*, and 20.2% as *other* because representative values were not present in the Subject Matter Domain (SMD) axis of the DO. The study also showed the need for an extension and clarification of the Type of Service (TOS) axis to completely represent different subtypes of nursing documents. A limit of this study was the size of the sample documents coming from only two institutions, which could affect the generalizability of the results. Similarly, in Hyun et al. (2009), document collections from the Columbia University Medical Center campus (CUMC)

² The work presented in this paper was realized in collaboration with Medas Srl, an Italian company that offers innovative services and IT solutions in the healthcare domain. Among these, it carries out the long-term preservation of digital health documents.

of New York Presbyterian Hospital (NYPH) were used to test the DO. The study evidenced the need for revisions of both the DO and the local clinical documents classification, as documents classified as *Not Specified* were too granular in their names and those classified as *Other* were not covered by the value list of the SMD axis. Chen et al. (2010) conducted a similar study on document names of two institutions (Fairview Health Services and Fletcher Allen Health Care). They used automated methods to identify exact matches, which resulted in 55% of the total amount of local document names (1,240). Authors opened a fundamental question: when mapping to an international standard, to what extent is it tolerable to lose granularity with respect to local names descriptions? And consequently, when is it appropriate to request the addition of new codes to the standard? Their position is that it depends on the use case (documents exchange, retrieval, etc.) and therefore they propose to maintain both local and standard names to «minimize information loss, preserve semantics, and enable flexible use. Moreover, they advanced the idea of a DO with a post-coordinated approach rather than the current pre-coordinated one. As DO is constantly updated, the results of studies from some years ago could have been different if conducted today, in any case they highlight some fundamental issues that are still detectable today, but which basically constitute the driving force for the improvement of DO itself. Such as in Li et al. (2011) where non-mapped notes are due to i) specificity of local items; ii) abbreviations and acronyms, which do not give clear evidence of what they represent; iii) missing DO values. While Wang et al. (2014) focused on values of Setting and Role axes, recommending specific additions to them, Rajamani et al. (2014a) addressed the Setting axis, with the aim of increasing DO representativeness of care settings other than hospital and office visits. The same authors analyzed also the Role axis finding lack of comprehensiveness, as DO values were either too broad or too specific (Rajamani et al. 2014b).

The limitation most complained by the analyzed previous studies is that the sample of document types used to evaluate the semantic coverage of the DO comes from one or at most two institutions, and therefore lacks a sufficiently broad representativeness. This study manages to overcome this limit as the documentary corpus made available by Medas comes from different types of healthcare facilities located throughout Italy.

More recently, a study used a deep learning model to automate the mapping of values to the DO axes (Zuo et al. 2021). The results showed improved performance compared to the manual mapping of the same corpus, which was used to annotate the framework. In the future it would be interesting to test the document corpus of this work through automatic mapping, in order to study methodologies for its refinement so that it could become a common practice to use the LOINC standard for the unambiguous identification of clinical document types.

Materials

The LOINC DO was created in the first 2000s from the joint effort of the LOINC committee and the HL7 document ontology task force, recognizing the need to uniformly name documents with the same content and distinguish those that could appear identical in the name but actually differ in the substance (Frazier et al. 2001). This is especially important to: i) exchange documents between different clinical information systems; ii) classify and retrieve documents coming from

different sources (as in the Italian EHR registries); iii) standardize naming of clinical documents templates. As for the LOINC codes for laboratory tests, these DO codes are not intended to substitute local names and codes but to come alongside them to ensure effective semantic interoperability.

The five axes of the DO codes are:

- *Kind of document*, which describes the basic structure of a document (e.g. note, certificate, etc.);
- *Type of service*, which identifies the type of service or activity delivered to the patient (e.g. consultation, admission evaluation, etc.);
- *Setting*, which is either the place or the situation where the document is produced (e.g. hospital, outpatient, etc.);
- *Subject Matter Domain (SMD)*, which represents the medical specialty or domain related to the document (e.g. cardiology, family medicine, etc.);
- *Role*, which is the medical level responsible for the document content (e.g. consultant, nurse, etc.).

Through the combination of values of those axes it is possible to model LOINC terms for clinical document types, whether they are more specifically medical or administrative. The LOINC DO class currently includes 3,620 records. The DO Committee constantly works to update and refine values of the axes in order to be representative of the expressive needs of the clinical practice.

The LOINC DO is also distributed as an OWL (Ontology Web Language) file³ to give the users the chance to have a computable format for the hierarchy.

In Italy, LOINC DO codes are primarily used to identify clinical documents indexed into the EHR registries of the Regions. They represent the values of the metadata *document type* in the Italian AD. Because of this and considering their reference role as a standard in this domain, Medas, an Italian private company, which provides services of long-term preservation of clinical documents, decided to map its local document type code system to LOINC in order to evaluate the semantic coverage of the standard and the chance to adopt it to be compliant with the national EHR and speed up the processes of correct document type recognition, effective classification and efficient retrieve.

Medas created its own coding system to identify clinical documents received by healthcare providers for long-term preservation. It was necessary to avoid inconsistencies in data item entries. It is based on two hierarchical levels: Genus and Species. The Genus level is the general form in which the document is presented, it always has the same meaning regardless of the associated Species. The Genus' identifier is indicated in the first part of the code of each document type, e.g. the Genus code 090 is for *Report*. This is a speaking code, as the same number always corresponds to the same document type. The Species level indicates the specific type of the document and, differently from the Genus, it is not a speaking code, as the same number changes its meaning depending on the Genus to which it is associated. For example, 090.060 is the code for *Cardiology Report*. In this case, 060 is the Species code which indicates the medical specialty *Cardiology*. But the same Species code 060, if associated with another Genus, has a different meaning. For exam-

³ For more details see: <https://loinc.org/document-ontology/owl/>.

ple, 100.060 is the code for *Vaccination register*, where 100 is the Genus code for *Register* and 060 is the Species code for *Vaccination*. The total number of clinical document types encoded in the Medas local coding system is 483.

Methods

Mapping aims to establish correspondences from concepts of a starting coding system, which is usually local and called *source* system, to concepts of a destination coding system, which is usually standardized at some level and called *target* system. This is the main way to implement the use of a clinical standardized coding system into a healthcare setting. The mapping is a non-trivial task because it is necessary to establish semantic correspondences, which refer to the content actually represented, beyond the names given to the clinical documents (Chiaravalloti 2022). This is not always simple if the standard is to be used by users who are different from the document producers, who actually knows its content, which may also change depending on the clinical flow in which it is integrated.

For the purpose of this work, mapping operations were manually performed by LOINC Italia experts. The work was approached by considering Medas document types according to their Genus in order to group together documents that are similar in their general meaning. This facilitated the maintenance of mapping coherence for documents that conceptually belong to the same semantic group and, instead, differ in the Species, mainly related to medical specialty. The search into the LOINC database was carried out through the LOINC Search web browser and the software RELMA (REgenstrief LOINC Mapping Assistant) using both local names and possible synonyms of the names of the Medas document types.

The LOINC codes candidates for mapping were primarily searched into the DO class. If they were not found there, the chance of identifying the target codes in other classes of the standard was then considered.

Similarly to previous abovementioned studies, we classified the type of matching between the source and the target codes as:

- *Adequate or Exact 1-1* to indicate terms with a full match between source and target systems;
- *Too broad* to indicate mappings to a LOINC concept, which is more generic than the source term;
- *Many-to-1*: we used this subcategory tag when multiple local codes were mapped to the same LOINC code. In this case, the specificity expressed in the source coding system doesn't find semantic correspondence into the target coding system, which groups it into a single generic code;
- *Too specific* to indicate mappings to a LOINC concept, which is more specific than the source term;
- *1-to-many*: we used this subcategory tag when one local code was mapped to multiple LOINC codes. In this case, the source coding system groups together concepts that are instead distinct in different codes of the target coding system.

Not covered to indicate no matching terms between source and target systems.

In Table 1 some examples for each category are reported.

Type of matching	Medas	LOINC
Adequate (Exact 1-1)	090.100 Referto di chirurgia pediatrica (<i>Pediatric surgery Diagnostic study note</i>)	68794-7 Pediatric surgery Diagnostic study note
Too broad	090.140 Referto di dermatologia e vene-reologia (<i>Dermatology and venereology Diagnostic study note</i>)	78260-7 Dermatology Diagnostic study note
Many-to-1	090.090 Referto di chirurgia maxillo-fac-ciale (<i>Maxillofacial Surgery Note</i>) AND 090.095 Referto di chirurgia orale (<i>Oral Surgery Note</i>)	34813-6 Oral and Maxillofacial Sur-gery Note
Too specific	090.070 Referto di chirurgia dell'apparato digerente (<i>Digestive system surgery Note</i>)	84073-6 Colon and rectal surgery Note
1-to-many	090.030 Referto di anestesia e rianima-zione (<i>Anesthesiology and Critical care medicine Note</i>)	34750-0 Anesthesiology Note AND 34754-2 Critical care medicine Note
Not covered	090.040 Referto di biochimica clinica (<i>Cli-nical biochemistry Diagnostic study note</i>)	-

Table 1. Examples of mappings from Medas to LOINC codes for each defined type of matching.

The identified matchings were then validated by a domain expert from LOINC Italia, different from the one who performed the mapping.

After the validation phase, further meetings with the Medas archivists helped to understand the semantic area covered by Medas codes matched as *too broad*, *too specific* or *not covered* and therefore to determine whether they could be classified differently or they need to be required as a new term submission to the LOINC DO Committee.

Results

Out of the total of 483 clinical document types encoded in the Medas local system, their mappings to LOINC codes resulted in:

- 144 adequate mappings, with an exact correspondence of 1-1 between the source concept in Medas and the target concept in LOINC;
- 209 too broad mappings, where the target LOINC code only partially covers the meaning of the source concept in Medas. In these cases, there is a loss of information moving from the source to the target coding system;
- 2 too specific mappings, where the target LOINC code only partially covers the meaning of the source concept in Medas. In these cases, there is an excessive specialization of the information moving from the source to the target coding system;
- 128 Medas clinical document types were not mapped as no semantic correspondence of concepts could be found in LOINC.

In percentages it means that 30% of the Medas local codes have a full match with one and only one LOINC code, 44% of them have a partial match, because in LOINC only broader or narrower concepts can be found, and finally 26% of the Medas codes do not have any match in LOINC as it was not possible to find a semantic representation of their meaning.

If we look in more detail at these percentages, it is interesting to note that among the Medas Genus codes that have the highest full match percentages there are *Letters*, *Prescriptions* and *Reports*. On the contrary, among the document types with less full match there are *Minutes* and *Declarations*, while the Medas Genus codes related to *Registers* and *Diaries* are not represented at all in LOINC (Table 2).

	Full match	Partial match	No match
<i>Report</i>	76%	5%	19%
<i>Declaration</i>	22%	22%	56%
<i>Letter</i>	80%	20%	0%
<i>Prescription</i>	73%	0%	27%
<i>Minutes</i>	53%	6%	41%
<i>Registry</i>	0%	0%	100%
<i>Diary</i>	0%	0%	100%

Table 2. Percentages of full, partial and no match for some of the Medas Genus.

In searching for correspondences between the Medas Genus and the LOINC codes, differences inevitably emerged deriving from the contexts in which the two coding systems were born. In fact, legislation has a great impact on the definition of types of clinical documents. Codes from the Medas system therefore do not always find semantic correspondence in LOINC, because they represent peculiar aspects of the Italian context. Conversely, many of the LOINC codes were created to reflect the clinical document types typical of the North American context. For these specific cases coding can therefore only take place with the creation of LOINC codes customized for the purpose. An example of this is the Genus *Certificate*. The Medas document types belonging to it are 21, and 7 of them full match with a LOINC code, while the remainder have been classified as *too broad* matchings as they can be either mapped to the LOINC 80571-3 *Certificate* or related to the Kind of Document *Note*, thus not allowing to specify information about the punctual function of a certificate. A certificate is, in fact, defined as a “written declaration, issued by a public authority or even, in specific cases, by a private individual, which expresses knowledge of the existence or truth of a fact or expresses the result of an evaluation”,⁴ which means that it has a more specific and official value than a simple clinical note. Either way the specificity of the information would be lost, therefore those concepts are candidates for new LOINC codes’ submissions.

Another special case is the Genus *Consent*. In the Medas coding systems it contains 133 clinical document types, but only 6 of them have a full match in LOINC. Although the source codes are of such a large number, the LOINC codes retrieved by typing the word *consent* are only 19. This

⁴ The translation is derived from <https://www.treccani.it/vocabolario/certificato/>.

discrepancy is due to the Italian legislative system (Chiaravalloti et al. 2015), which is greatly focused on patients' rights and privacy protection, requiring consent documents for almost every medical procedure or event. Furthermore, there are document types not only to express consent in a positive form, but also to express the denial of consent or the modification of previously given consent. There is thus an information gap between the source and the target coding systems. In Table 3, some examples of mapping *many-to-1* for the Genus *Certificate* are reported. In them, moving from the Medas coding system to LOINC there is a loss of information because Medas document types are narrower than the concepts conveyed by the existing LOINC codes. Therefore, in this case the choice is whether to require LOINC DO Committee to create new codes to maintain the specificity of the source information or whether to use the existing generic LOINC codes as a sort of higher-level categories, but with the knowledge that they act only as containers of *n* more specific document types.

DocumentTypeLabel	LOINCDocumentCode	LOINCDocumentName
Carotid stenosis consent	64293-4	Procedure consent document
MRI consent		
CT consent		
PET CT consent		
Consent for surgery	61358-8	Surgical operation consent document
Obesity surgery consent		
DBP review consent		
Bifemoral bypass consent		

Table 3. Examples of mappings many-to-1 for the Medas Genus Consent.

Regarding the Medas Genus *Report*, it is intended as “the written report issued by the doctor who subjected a patient to a clinical or instrumental examination”.⁵ To map this type of clinical documents, it was initially necessary to disambiguate the meaning of the LOINC document types *Note*, *Diagnostic study note* and *Report*. After also a discussion with members of the LOINC DO Committee, it was established that *Note* is the most generic term to indicate a report resulting from a medical visit, while *Diagnostic Study Note* is used when the document is the result of a diagnostic study, therefore more specifically created with the support of medical devices. *Report*, instead, has a more limited use in English because it concerns what is reported because it has been originally expressed in another way such as the Laboratory report. Out of the 87 *Report* types in the Medas coding system, 60 have been mapped to a LOINC code. Nonetheless, it was not always possible to attribute a 1:1 correspondence between the source and target concepts. Some of the most particular cases detected are presented below. There are cases of 2:1 correspondence, in which two different types of clinical documents of the source coding system are mapped towards the same type of clinical document in the target coding system, which is the LOINC standard. Some examples are:

⁵ The translation is derived from https://www.treccani.it/enciclopedia/referto_%28Dizionario-di-Medicina%29/.

- *Cardiology report* and *Cardiovascular system disease report* have been both mapped to the LOINC code 75425-9 *Cardiology Diagnostic study note*;
- *Oral surgery report* and *Maxillofacial surgery report* have both been mapped to the LOINC code 34813-6 *Oral and Maxillofacial Surgery Note*.

In other cases, it was necessary to carry out mappings towards broader LOINC terms compared to the concepts conveyed by the Medas document types. This type of association inevitably leads to a loss of information in the transition from the local to the standard coding system. Some examples are:

- *Dermatology and Venereology report* was mapped to the LOINC code 78260-7 *Dermatology Diagnostic study note*, thus losing the information relating to the missing SMD Venereology;
- *Anesthesia and resuscitation report* can be mapped either to the LOINC code 34754-2 *Critical care medicine Note* or to the LOINC code 34750-0 *Anesthesiology note*. In both cases the information relating to one of the two mentioned medical branches is missing.

Some types of Medas Report have not been mapped because the corresponding specialist branch is missing in the LOINC DO. It should be represented in the SMD axis; therefore, these could be addition proposals for the standard in order to have greater semantic coverage of the reference domain. Some examples are: *Clinical Biochemistry*, *Medical Physics*, *Thermal Medicine*, *Tropical medicine*, *Health statistics and biometrics*, *Senology*, *Orthodontics*, *Neurophysiopathology*, *Orthopedics and Traumatology*, *Clinical Psychology*, *Neuropsychiatry*, and *Child Neuropsychiatry*.

It is to be noticed that the case of *Forensic medicine report* as *forensic medicine* is present as a possible value for the SMD axis, but the only document type to be associated is *Referral note* therefore the representation of the *report* concept associated with this specialist branch is missing.

Discussion

The abovementioned mapping results show that LOINC is the most suitable standard for uniquely identifying clinical document types, as it was possible to find a semantically representative LOINC code for most of the concepts expressed in the Medas coding system. Out of a total of 483 document types analyzed, it was possible to map 144 of them with a full match of the concepts in the source and target coding systems, and 211 of them with a partial match.

However, although the DO is continuously updated to be representative of emerging clinical needs, its semantic coverage can be improved to adapt to national contexts other than the US master one. In this work we identified some additions to the DO axes that can be submitted to the LOINC DO Committee. Specifically, possible extensions for the Kind of Document axis are *Register*, *Diary*, *Index*. The SMD axis could be extended to include *Venereology*, *Clinical Biochemistry*, *Medical Physics*, *Thermal Medicine*, *Tropical Medicine*, *Health Statistics and Biometrics*, *Orthodontics*, *Neurophysiopathology*, *Orthopedics and Traumatology*, *Clinical Psychology*, *Senology*, *Neuropsychiatry*, and *Child Neuropsychiatry*. Finally, for the Type of Service axis a possible concept to submit is *Dispensation*. The process of creating a new LOINC term after submission is quite fast, which makes the standard particularly flexible in adapting to new scientific discoveries and the needs of the users' community (Cardillo et al. 2024). Moreover, this moment is particularly

timely, as announced during the LOINC 2024 conference, the LOINC DO will undergo a comprehensive review process.

Beyond the lack of a specific LOINC code for each document type, the problem relating to the concepts matching can be considered broader and it regards, for example, the organizational logic used in the different contexts. Therefore, this type of work cannot be separated from a preliminary analysis of the starting context nor to be uniform for all organizations dealing with the long-term preservation of clinical documents. The strategy that proved to be successful in our case for solving some mapping issues was the direct comparison with those who developed the source coding system. This allowed us to identify the most suitable solution, which in some cases was to identify a suitable LOINC code using synonyms, in others to propose the concept for a new submission to the LOINC DO Committee as it was not represented in the standard.

The study also offered the possibility of analyzing synonymy in LOINC and above all the richness of synonyms currently present in the standard database. They are important because they are further, often valuable, search keys to arrive at the target concept. It should be noted that for the documents section of LOINC there is not a wealth of synonyms comparable to that present for the laboratory and clinical sections. On the one hand, this could be an indication of a certain uniformity of the context from which the DO values were created, but on the other hand it could constitute an obstacle in finding the codes in contexts, such as the Italian one, in which there is a high linguistic variability in naming the types of clinical documents.

Furthermore, the study also made it possible to verify the names of the clinical documents used in the Medas coding system, reconciling those that are different in form but equal in substance. It happens, in fact, that in a post-coordinated coding system, such as the Medas one, control over the expansion of all the branches of the hierarchy can be lost. Therefore, repetitions of concepts can be created, expressed through synonyms, due to the different names of the clinical documents used in the local systems of the healthcare facilities, which use the long-term preservation services offered by Medas. However, this aspect can be converted into an advantage, considering that the results of this mapping towards an international standard could lay the foundations for the development of guidelines to regulate naming practices within single healthcare facilities or at a regional and national level. It could mean reorganizing existing naming conventions or primarily regulating them at the time of the creation of the document management system.

Conclusion

This paper deals with the process of aligning the names of the types of clinical documents placed in long-term preservation with the international reference standard for the domain, namely LOINC. This activity is carried out through mapping operations that connect the semantically equivalent concepts of the source coding system to those of the target or destination coding system. The case study described was created in collaboration with Medas, a leading Italian company in providing long-term preservation services for clinical documents. The results obtained show how the Medas local coding system allows a certain flexibility through its structuring into Genus and Species and is well represented through the LOINC codes. The concepts that could not be mapped to the international standard, however, mainly concern specific areas of the Italian context and missing

values in LOINC. It was thus possible to evaluate the semantic coverage of the standard and consequently identify concepts to candidate for LOINC new terms submissions.

Regarding research perspectives, the database of the validated mappings of the Medas coding system towards LOINC could represent a knowledge base for supporting mapping operations of clinical document types defined by other companies in the healthcare domain. The common reference to a standard could, in fact, create uniformity where a high degree of linguistic variability remains in determining the names of clinical documents. It should move from a more general reflection on a rationalization of the document types in the document management system, considering that the transition from analog to digital does not mean “translating” existing documents’ types into the corresponding digital ones, but a “rethinking” of the entire process is necessary in order to streamline not only the single document types but above all the documents’ flows. Furthermore, results of this work could be used to enrich the list of LOINC document type codes, which are the metadata values to index clinic document types into the Italian FSE according to the AD. From an international perspective, the use of a standard like LOINC represents an advantage in terms of cross-border clinical interoperability within the exchange of clinical documents among EU countries and beyond.

Acknowledgements

We would like to thank Medas Company (particularly Umberto Ferri and Carlo Alzati) for supporting and sharing their data used in this study. Details about Medas coding system are not shown in this study because of property rights, therefore only aggregated analysis and significative examples are used to highlight the case study.

References*

- Brown, Steven H., Micheal Lincoln, Shawn Hardenbrook, Olga N. Petukhova, S. Trent Rosenbloom, Paul Carpenter, and Peter Elkin. 2001. "Derivation and Evaluation of a Document-Naming Nomenclature." *Journal of the American Medical Informatics Association* 8 (4): 379–390. <https://doi.org/10.1136/jamia.2001.0080379>.
- Cardillo, Elena, Maria Chiaravalloti, and Erika Pasceri. 2024. "Covid-19 Impact on Standard Coding Systems Update." In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: HEALTHINF*, 490–497. <https://doi.org/10.5220/0012387600003657>.
- Chen, Elizabeth S., Genevieve B. Melton, Mark E. Engelstad, and Indra Neil Sarkar. 2010. "Standardizing Clinical Document Names Using the HL7/LOINC Document Ontology and LOINC Codes." In *AMIA Annual Symposium Proceedings*, November 13, 2010, 101-105.
- Chiaravalloti, Maria Teresa, Mario Ciampi, Erika Pasceri, Mario Sicuranza, Giuseppe De Pietro, and Roberto Guarasci. 2015. "A Model for Realizing Interoperable EHR Systems in Italy." In *Proceedings of the 15th International HL7 Interoperability Conference*, Prague, Czech Republic, February 9-11, 2015, 13-22.
- Chiaravalloti, Maria Teresa. 2022. "A proposal for a LOINC automatic mapping support tool." *AIDAinformazioni* 3-4:47-56.
- Ciampi, Mario, Giuseppe De Pietro, Christian Esposito, Mario Sicuranza, and Paolo Donzelli. 2012. "On Federating Health Information Systems." In *Proceedings of the International Conference on Green and Ubiquitous Technology*, Bandung, Indonesia, June, 30–July 1, 2012, 139-143.
- De Pietro, Giuseppe, and Mario Ciampi. 2011. "L'interoperabilità del FSE: il progetto Open-inFSE." *E Health Care Innovazione e Tecnologia Sanità Elettronica* 3: 8-14.
- Dugas, Martin, Sylvia Thun, Thomas Frankewitsch, and Kai U. Heitmann. 2009. "LOINC Codes for Hospital Information Systems Documents: A Case Study." *Journal of the American Medical Informatics Association* 16 (3): 400-403. <https://doi.org/10.1197/jamia.m2882>.
- Frazier, Pavla, Angelo Rossi-Mori, Robert H. Dolin, Liora Alschuler, and Stanley M. Huff. 2001. "The Creation of an Ontology of Clinical Document Names." *Studies in Health Technology and Informatics* 84 (Pt 1): 94-98.
- Hyun, Sookyung, Rosemary Ventura, Stephen B. Johnson, and Suzanne Bakken. 2006. "Is the Health Level 7/LOINC Document Ontology Adequate for Representing Nursing Documents?" *Studies in Health Technology and Informatics* 122: 527-531.
- Hyun, Sookyung, Rosemary Ventura, Stephen Johnson, and Suzanne Bakken. 2009. "Iterative Evaluation of the Health Level 7—Logical Observation Identifiers Names and Codes Clinical Document Ontology for Representing Clinical Document Names: A Case Report." *Journal of the American Medical Informatics Association* 16 (3): 395-399. <https://doi.org/10.1197/jamia.M2821>.

* Websites were last accessed on 23 November 2024.

Italia. Legge 22 dicembre 2017, n. 219, “Norme in materia di consenso informato e di disposizioni anticipate di trattamento.” Gazzetta Ufficiale n. 12, January 16, 2018.

Li, Li, C. Paul Morrey, and David Baorto. 2011. “Cross-Mapping Clinical Notes between Hospitals: An Application of the LOINC Document Ontology.” In *AMIA Annual Symposium Proceedings*, October 22, 2011, 777-783.

Rajamani, Sripriya, Elizabeth S. Chen, Mari E. Akre, Yan Wang, and Genevieve B. Melton. 2015. “Assessing the Adequacy of the HL7/LOINC Document Ontology Role Axis.” *Journal of the American Medical Informatics Association* 22 (3): 615-620. <https://doi.org/10.1136/amiajnl-2014-003100>.

Rajamani, Sripriya, Elizabeth S. Chen, Yan Wang, and Genevieve B. Melton. 2014. “Extending the HL7/LOINC Document Ontology Settings of Care.” In *AMIA Annual Symposium Proceedings*, November 14, 2014, 994-1001.

Shapiro, Jason S., Suzanne Bakken, Sookyoung Hyun, Genevieve B. Melton, Cara Schlegel, and Stephen B. Johnson. 2005. “Document Ontology: Supporting Narrative Documents in Electronic Health Records.” In *AMIA Annual Symposium Proceedings*, 684-688.

Wang, Yan, Serguei Pakhomov, Justin L. Dale, Elizabeth S. Chen, and Genevieve B. Melton. 2014. “Application of HL7/LOINC Document Ontology to a University-Affiliated Integrated Health System Research Clinical Data Repository.” In *AMIA Joint Summits on Translational Science Proceedings*, April 7, 2014, 230-234.

Zuo, Xu, Jianfu Li, Bo Zhao, Yujia Zhou, Xiao Dong, Jon Duke, Karthik Natarajan, George Hripcsak, Nigam Shah, Juan M. Banda, Ruth Reeves, Timothy Miller, and Hua Xu. 2021. “Normalizing Clinical Document Titles to LOINC Document Ontology: An Initial Study.” In *AMIA Annual Symposium Proceedings*, January 25, 2021, 1441-1450.

Libraries and health literacy

Artemis Chaleplioglou^(a)

a) University of West Attica, <https://orcid.org/0000-0002-6519-7428>

Contact: Artemis Chaleplioglou, artemischal@uniwa.gr

Received: 13 December 2024; **Accepted:** 18 March 2025; **First Published:** 15 May 2025

ABSTRACT

Disparities in health care are a universal problem. Different subpopulations of our society are at different predictive risks of morbidity and have different potentials in seeking and receiving appropriate medical care. The parameters involved in health disparities include living standards, nutritional conditions, preventive medicine and diagnostics, medical treatment in a suitable nursing environment, and the degree of the people's health literacy. Indeed, the latter is a pivotal and urgent priority in promoting public health. This literature review focuses on libraries' crucial role in reducing health disparities by providing medical information and librarian guidance and enhancing health literacy. The local library, a crucial source of public medical information, serves any community or minority, offering activation, inclusion, and relief. The accessibility of health information in libraries provides a sense of reassurance and confidence to the public and empowers them with knowledge and understanding. The public libraries represent an information provider network with dense coverage of the whole country in urban and rural areas. They can deliver health literacy services everywhere. In addition, with its safety and inclusivity, the public library environment is where everyone can seek medical information and training in practical and theoretical terms of health conditions, seasonable medical issues, and personal and family care needs in a secure and welcomed environment.

KEYWORDS

Health literacy; Libraries; Health equity; Digital literacy.

Introduction

National health systems are not available to all citizens or residents of countries on a global scale. The living and working environment, social, economic, and cultural conditions, and education emphasizing health literacy significantly affect individual health, even more so than the genetic predisposition factors for disease. In addition, the health literacy of individuals and society indicates the capacities for understanding, trust, response, compliance, and cooperation during health or environmental crises. Inequalities in health and health literacy are realities that, regardless of their sources, pose severe risks on a local and global scale. They must be addressed with solutions both within and outside the narrow framework of health delivery and treatment (NAS 2024).

In contrast to the situation in the field of health, libraries are spaces for the promotion of equality, inclusion, and democratization of knowledge worldwide. They are the cornerstone of valid information and develop a relationship of trust with their public by offering workspace, material, and environment for information seeking, recreation, and interaction (Audunson et al. 2020). Nowadays, the library's traditional role goes beyond the provision of printed material and has been significantly enriched by providing multimedia material and access to documented knowledge sources on the Internet. In addition, municipal and public libraries form in most countries, and in Greece, an information network of dense territorial coverage serves both urban and remote rural or island areas. Can libraries be included and used in national health policymaking?

This study examines the possibilities of libraries operating as external factors of the health system, as carriers of health information, and improving the health literacy of the population to bridge the gap of health inequalities. First, inequalities in the health field and the factors that cause them globally are examined. Emphasis is placed on those individuals who, due to social, economic, cultural, psychological, or emotional concerns or physical barriers, are unable or reluctant to seek health services or medical information on health issues. Secondly, the library is examined as a provider of valid medical information, a contribution to health literacy, and a promotion of public health based on international experience, proposals, practical applications, and how these are evaluated. Finally, the challenges and possibilities of the practical application of this approach in our country are examined in the context of defending the right to documented knowledge and health in the information age.

Inequalities in Health

According to the World Health Organization, health equity is “the absence of unjust and avoidable or remediable differences in health between population groups based on social, economic, demographic or geographic criteria”. Advances in science and technology have increased life expectancy and quality of life, however unevenly. Existing differences widened between those with the best and those with the worst health and well-being. The difference in life expectancy between the richest and poorest countries reaches two decades. Most premature deaths from non-communicable diseases occur in poor and middle-income countries, while infant mortality is eight times higher in Africa compared to Europe (WHO 2019). However, these differences in health are also found within each country, with economic, geographical, social, and cultural criteria, so almost half of the people in the world do not enjoy the necessary health services (Rid et al. 2016).

According to the non-profit Association of American Medical Colleges (AAMC), in the era before the COVID-19 pandemic, residents living in rural areas represented 15-20% of the population of the United States' Americans face inequities and had worse access to health services than residents of urban and suburban areas. It is found that rural America's health disparities are deeply rooted and attributed to economic, social, racial, ethnic, and geographic factors related to the choices of health workers, forming a two-sided problem of access to medical care (Warshaw 2017). The problem is growing in intensity as evidenced by the increase in the number of rural hospitals closing or ceasing to operate as hospitals and converting to short-stay health centers in the United States, with an average of 8.6 hospitals closing per year in five years 2005-2009, 8.8 in the period 2010-2014, and 13.6 in the period 2015-2019, a phenomenon partially interrupted by the COVID-19 pandemic (UNC 2020). Typically, communities that lack hospital infrastructure or suffer economic hardships have unique characteristics (Thomas, Holmes, and Pink 2016). The children in these areas are a sensitive population group exposed to poverty and health and behavioral problems that affect the development and future of their health (Kelleher and Gardner 2017). According to the Centers for Disease Control and Prevention (CDC), focused interventions on the problem of health disparities in the rural population in the United States of America, including health literacy, would result in an overall increase in life expectancy by reducing the five leading causes of death related to heart diseases, malignant neoplasms, unintentional injuries, chronic lower respiratory diseases, and strokes, the spectrum of which statistically burdens rural populations more than urban populations (Garcia 2017).

There is a significantly uneven distribution of hospital units in capital cities and large urban centers rather than in Greece's countryside and islands. On top of this, peripheral health institutions suffer from a lack of personnel or resources. Based on the data of the Hellenic Statistical Authority (ELSTAT), in a total of 267 units, 34% of public and private hospitals are located in Athens, 16% in Thessaloniki, and 12% in Thessaly, cumulatively 62% of the units and the rest are shared in the rest of the country, 6% in Eastern Macedonia and Thrace, 3.4% in Western Macedonia, 2.6% in Epirus, 2.2% in the Ionian Islands, 5.2% in Western Greece, 3.7% in Central Greece, 3.7% in the Peloponnese, 2.6% in the North Aegean, 3.4% in the South Aegean and 5.6% in Crete. Hospital coverage follows the demographic distribution of the population and less the territorial area. Thus, most peripheral hospitals in the region's capitals cover larger territorial areas with a lower population density than the urban ones. These figures are more evident when examining the distribution of beds available for inpatient care, with 41% of the 45,101 beds located in Attica, 18% in Thessaloniki and its wider region, and 9% in Thessaly, followed by 5.6% in Eastern Macedonia and Thrace, 2.4% in Western Macedonia, 3.5% in Epirus, 1.5% in the Ionian Islands, 4.5% in Western Greece, 2.2% in Central Greece, 2.7% in the Peloponnese, 1.4% in the North Aegean, 2.6% in the South Aegean and 5.6% in Crete (ELSTAT 2021). The allocation of hospitals and closed hospitalization beds makes it mandatory for rural and island hospitals to provide mobile units and ambulances for the timely treatment of emergencies. However, regular medical follow-up and laboratory examination become particularly difficult and expensive in these areas for the chronically ill, the elderly, pregnant women, and infants under five years of age since they will have to travel to and from the nearest urban center to be examined. In addition, between 2011 and 2020, the number of hospitals in Greece decreased by 15%, while the number of available beds decreased by 9%. This reduction seems to be a result of the ongoing and prolonged eco-

conomic crisis in the country as it concerns almost exclusively private hospitals and clinics that were mainly in Attica (20% reduction) with a relatively small capacity to treat patients, as shown by the number of beds that were lost (ELSTAT 2021). The Greek data show that in the outskirts of Athens and Thessaloniki, the supply of health services with hospitalization is based almost exclusively on public structures. Thus, in the countryside and islands of Greece, no significant reduction in hospital infrastructure was observed in the last decade. It is worth noting that the Greek state clearly recognizes and legally enshrines Health as a fundamental human right. Whereas the Greek Legislation stipulates that regardless of nationality, color, place of residence, legal status, religious or philosophical beliefs, sexual orientation, economic status, or insurance coverage, everyone has the right to receive health services in the public health system, respecting religious and ideological beliefs of citizens, recognize the right to consent or refuse any diagnostic or therapeutic procedure for oneself or a protected member (MoH 2024). Therefore, health inequalities in Greece are a mixture of geographical population distributions, economic conditions, and social, cultural, or religious beliefs. In the context of individual beliefs and perceptions, health literacy is of value so that choices in health, lifestyle, diagnosis, or treatment are made with knowledge of the health damage that may be caused due to specific choices and the real diagnostic and therapeutic possibilities of different methods. The conceptual framework through which social, economic, political, cultural, educational, biological, and material factors operate, interact, and influence public health to produce disparities is thoroughly analyzed in Solar and Irwin's 2010 World Health Organization study (Solar and Irwin 2010). The factors of social health inequalities can be divided into structural and intermediate. Structural factors are of two kinds, state and individual, with the former shaping the social environment and the latter concerning social stratification. State factors are defined as the organization and operation of the state and, in particular, the state structures that are part of the triptych of Social, Economic, and Political organization, specifically in governance, the macroeconomic policies that are applied, in the social policies of the state and the welfare state, the Health System, and social care for the population, in education, as well as in the legal and regulatory provisions that regulate market, work or housing issues. In addition, the cultural and social values of the society and individual structural factors within this geopolitical environment are also major players. Social and economic stratification define the consumption possibilities of the individual. On top of that, sex, age, race, or ethnic origin may affect social and economic status. Individual structural social factors include the individual's education, occupation, and income. Consumption possibilities are associated with the quality of healthcare access. Social inequalities and their extent determine the degree of social cohesion and social capital. In the intermediate social factors that affect inequalities in health, we have the health conditions that prevail and affect the local populations in the context of the geographical and climatic zones. These factors define the material conditions, life, work, third places, nutrition, hygiene, availability of food or medicine, attitudes, and lifestyle. In addition, the factors in this unit include biological factors such as pathogens, age, and psychological factors that operate independently or are intertwined with material conditions, behaviors, and the state of the state, individual liberties, freedom of speech, freedom of expression or its suppression, the state of peace or war, any natural disasters or situations of crisis and insecurity. The public healthcare system and access to its services should be culturally compatible. This compatibility is based on the values and beliefs of the population being served, general, local, or minority.

These determinants are essential for equality in health and the population's quality of life. Inequalities in health, while primarily affecting and with greater intensity, sensitive social groups, are a matter for the entire society and acquire a global dimension as they are the source of health problems that can affect everyone, especially concerning infectious diseases (Cooper 2021). The US National Academies of Sciences, Engineering, and Medicine's recent report on Diversity, Equity, Inclusion, and State Commitments and Actions for Health Equity by Executive Decision-Making Centers notes that "over the past four decades, a wide variety of special commissions and research groups from international organizations have published numerous reports demonstrating that national health systems are not available to all citizens or residents of states and that systematic restrictions make equal access difficult for all so that they enjoy appropriate health services compatible with their cultural values. Research from the same period shows that where someone lives, works, and plays, i.e., its relevant social factors, affects health more strongly than genetic background. The COVID-19 pandemic has emphatically demonstrated the systemic chronic problems of health systems, especially those related to unequal access to medical care, erosive inequities, and systemic racism that make health systems complicit in erosive health inequities" (NAS 2024). Therefore, solutions to the problem of inequalities in the field of Health should be sought extra-systemically with scientific actions of a developmental and organizational nature (Galaviz et al. 2020).

While the COVID-19 pandemic firstly caused morbidity and increased mortality worldwide, on a second level, it affected lifestyles, individual behaviors, direct interpersonal communication, production, and the global economy; on a third level because it occurred in the Information Age was followed by a phenomenon of cataclysmic proportions referred to as an information pandemic (infodemic). While the term infodemic was introduced by David J Rothkopt in criticism of the role of media news agencies in extrapolating SARS regional health crisis into a debacle that stroke global economy in 2003, SARS-CoV-2 virus caused the pandemic and the information pandemic was caused by unintentional or intentional misinformation affecting and confusing societies in all countries (Scales, Gorman, and Jamieson 2021). At a time when governments were making difficult decisions to manage the health crisis, decisions that required society's compliance and cooperation to achieve and deliver results, conflicting news, scientific rebuttals, pseudoscientific theories, fake news, and theories conspiracy theories flooded the media, and the internet forming an explosive mix. This situation was reinforced by the exchange of messages in social networks that functioned as isolation bubbles (filter bubbles) in which only the echo chamber of the same news or fake news reached them (echo chamber) (Geschke, Lorenz, and Holtz 2019). Evaluating information with reliable evidence and scientific knowledge was essential for rationally shaping public opinion and developing trust in experts. In this information pandemic, the importance of society's health literacy in dealing with health crises has become more evident. Health literacy at the individual and societal levels determines individual health behavior choices. It affects family relationships, organizational efforts, policy decision-making, quality of life, morbidity and mortality indicators, and national economies (Sentell, Vamos, and Okan 2020). Disparities in health literacy substantially affect the outcome of health events.

Furthermore, the conditions demonstrated the close relationship between health literacy and digital literacy, with digital inclusion emerging as a social demand of crucial importance for protecting individual and public health. This request concerns not only access to broadband networking and the use of mobile computing technology but also the use of technology in telemedicine by

utilizing “smart” portable biomedical health monitoring devices (Sieck 2021). However, achieving these methods requires quality education and digital literacy, combined with health literacy so that the public can gain confidence and be encouraged to participate and collaborate in such programs, enhancing understanding of scientific methods and methods of safeguarding personal data against any reservations and hesitations due to the adoption of these technologies.

Timely and valid information on health and technology issues, access to information, and lifelong learning emerge at the forefront of the fight against social inequalities, geographical limitations, and prejudices that shape health inequalities. These actions cannot proceed only through the public health system but need the active participation of external educational agents, with local libraries appearing as the ideal intervention organizations.

Library, Health Information, Health Literacy

In decision-making centers, state actors and health professionals rarely consider libraries when it comes to issues of treatment or prevention in public health or the elimination of health disparities. However, all the evidence points to libraries as the best organizations to solve these problems effectively. Libraries function as regional centers of organized and documented public information. They are the pre-eminent education and literacy providers, providing up-to-date and valid information in physical or electronic form, and medical information is already on the list (Morgan 2016). Guided by this presumption, librarians and information scientists can contribute substantially to the work of health professionals to promote it. Medical information as a library service has added value for vulnerable populations due to economic, social, and physical limitations or exclusions. Having gained the public’s trust and being a living part of the local community (Horrigan 2015), libraries contribute significantly to community engagement in social actions. With targeted health actions, libraries can substantially contribute to cultivating good health practices, healthy lifestyles, basic medical knowledge, and protection of the natural environment, as well as activating and promoting society’s participation in these directions (Aguilar-Gaxiola 2022). Thus, they become shapers of public opinion by promoting public policies to address health issues holistically beyond the self-evident physical to the intellectual and lifestyle level. For the most effective dissemination of community involvement messages, the contribution of specialist health professionals is often necessary, who, with their expertise, can enhance participation and provide feedback on issues that concern the public. In these coordinated actions, the public can receive information on telemedicine issues to better serve geographically isolated communities by utilizing modern “smart” portable health monitoring technologies (wearables) (DeGuzman, Jain, and Loureiro 2022).

A statistical study of public libraries in the United States of America before the COVID-19 pandemic by the Pew Research Center showed that 17,000 public libraries cover 95% of the population. In contrast, 50% visit a public library at least once a year, and they serve four million visitors daily. The reach of public libraries in the US is impressive and demonstrates their importance in the design of health policies (Horrigan 2016). Of the millions of searches, more than 30% are related to personal, family, or friend health issues. In the broader context of health research issues, issues related to the health and care of animals, pets, or production animals, environmental issues, and food and water safety issues are included. Also, issues related to diseases such as infectious

diseases, transmitted by human contact, epidemics, and pandemics or zoonotic diseases, but also non-communicable diseases or health conditions such as cancer, cardiovascular and neurological diseases, psychological issues, aging, lifestyle, occupational hazards, the care of children, the sick and the elderly. In the most specialized topics, we have diagnostic issues that may include modern genomics, proteomic methods, cell biology or stem cell topics, and practical clinical issues such as drugs, medication, their correct dosage, their potential side effects, personalized medicine topics, new therapeutic approaches, surgical approaches, but also bioethical issues or issues of personal data and their protection (Rubenstein 2012).

Recognizing these unique features of the extra-systemic operation of libraries for the benefit of public health, the British National Health Service (NHS) puts forward in its latest strategic planning program the library space as a field of targeted action and funding. The Knowledge for Healthcare 2021-26 program aims to ensure equal access to medical information. It aims as a priority to mobilize evidence for the dissemination of knowledge and the improvement of therapeutic results (NHS 2021). The purpose of the program is achieved through the tripartite of appropriate sources of information, which is achieved through the guidance of library services, a proper intervention group, and a correct division of roles regarding the interaction with health experts and medical information services for all, built on the principles of equity, diversity, and inclusion.

Great Britain's libraries in the system developed by the NHS act as hubs for the dissemination of evidence-based scientific medical knowledge and as a field of action, education, and funding for programs of health education, vaccination, preventive medicine, and systematic medical information for common chronic conditions such as diabetes or hypertension, but also for infectious diseases, epidemics, and pathogenic microorganisms. Furthermore, the libraries present more general health suggestions based on the experts' recommendations concerning changing the public's lifestyle by adopting a healthier lifestyle with proper diet and exercise.

The statistics from the previous period, 2019-2020, demonstrate the success and justify the continuation of the NHS library funding strategy to strengthen medical information. The first phase of the project ensured (a) an increase in staff to 6 librarians per library, (b) the special retraining of the staff in matters of information search and analysis, (c) the enrichment of the physical collections of the libraries in both specialized and popular medical and health literature, (d) access to electronic resources and health databases subscription content, such as the National Medical Information Discovery System interface through EBSCO, (e) the library outreaches, (f) the creation of suitable library websites for remote public access, and (g) the uninterrupted flow of valid information during the pandemic period from the libraries.

In a low-literacy environment, with 20% of adults in Great Britain having a literacy level equal to or below that of a nine-year-old child and with 43% of adults unable to understand and comply with the written medical information on a prescription, the investment in health literacy is expected to contribute to improving the health of the population. In addition, the availability of popularized medical information helps the general population make better individual health decisions. In addition, patient compliance with medical prescriptions and prescriptions increases. In contrast, patients can now actively participate in planning their medical care to achieve better management of chronic diseases and increase trust in specialists while reducing the feeling of embarrassment toward the doctors and fear of ridiculed by them (Gursul 2022). Beyond the social parameters, the program's performance can also be seen in the financial aspect by applying new clinical methods

and adopting good practices. Libraries in Great Britain providing medical information to the public offer public health doctors and nurses time credit. In general, the patient and the public resorting to the library instead of the infrastructure of hospitalization and medical care reduces the number of daily clinical decisions.

At the same time, librarians assist in collecting medical information and specialists by synthesizing evidence-based knowledge necessary for case management. Libraries' support of the public's individual health choices further reduces the burden on the public health system by reducing the incidence of lifestyle-related chronic diseases (Karki et al. 2024). Economically, it is estimated that for every euro spent by the NHS on Great Britain's libraries, the benefit is 2.4 euros.

Libraries worldwide are the most critical enablers of children's and adult literacies, including digital, scientific, environmental, and medical (Koltay, Spiranec, and Karvalics 2016). The libraries' information activities are aimed at the public, regardless of educational level, degree of education, age, or profession. Regarding digital literacy, libraries educate their public to evaluate information, emphasizing information from the internet, helping to fight misinformation and debunk fake news.

Users can check the reliability of information from the website's accreditation data, its publication, updated calendar data, and its general appearance, including the speech within it. Thus, its authenticity (authority), any prejudices (bias), and how modern and up-to-date the information is (current) are examined, an essential element in medical information where new observations and scientific studies can overturn previous data (Atkinson 2016). In terms of scientific literacy, the purpose is to inform the public about the content of science, its mathematical foundation, the use of technology in measurements, trial and error methods, observations, experiments and theory, revision, and formulation of new theories focusing on the real possibilities and limitations (Kearns and Hybl 2005). A particular case is environmental literacy, which aims to provide individual and collective information for decision-making related to the environment and its protection for preserving biodiversity, natural resources, and drinking water, as well as improving the quality of life of people and societies. Therefore, it also touches on health issues (Finn and O'Fallon 2018).

Health literacy refers to how individuals acquire, process, and understand basic information and information needed to make appropriate health decisions. Understandably, a low level of health literacy is related to social inequalities and significantly reduces the adequacy of understanding of health conditions and interpersonal communication between the public and nursing staff. Libraries, safe spaces for inclusion, information, and training, offer medical information commensurate with the needs and capabilities of the public in practical and theoretical health issues, current medical issues, or personal or family needs, improving health literacy quantitatively and qualitatively. The local library should be a vital source of medical information for the public, every community, and every minority, offering meaningful activation, inclusion, and relief (Philbin et al. 2019).

Hypothesis for the Implementation of a Health Information Pole in Greece

In Greece, based on the census data of the ELSTAT, a total of 347 libraries for 2022, 52% of which are public or municipal, while the rest are legal entities. Libraries in our country serve more than three million readers. At the same time, almost seven million users use library websites, four million users use electronic library databases, and more than two million items in physical form were

borrowed by users for external use (ELSTAT 2022). Municipal and public libraries, along with particular academic, research, school, and private libraries, have formed the Greek Libraries Network under the authority of the National Library of Greece (NLG) (2020). This network of 273 libraries has been predominantly built around the public or municipal entities, representing nearly 88% of the total. The Greek Libraries Network aims to deliver better information services to all participants through a standard set of authorities, according to Anglo-American Cataloging Rules, 2nd edition, Revised (AACR2), an aligned legal documentation environment, sharing manuals and guidelines of good librarianship practice, and by facilitating interlibrary loan and resource sharing (NLG 2020). The Network supports coordinated actions to promote reading and creativity. The Greek Libraries Network offers a dense coverage of the regions of Greece, which locally may exceed the hospitals' coverage (Figure 1). It can be postulated that the Greek Libraries Network may deliver health information to regions the healthcare system has not reached, thus covering the information needs of most of the Greek general population.

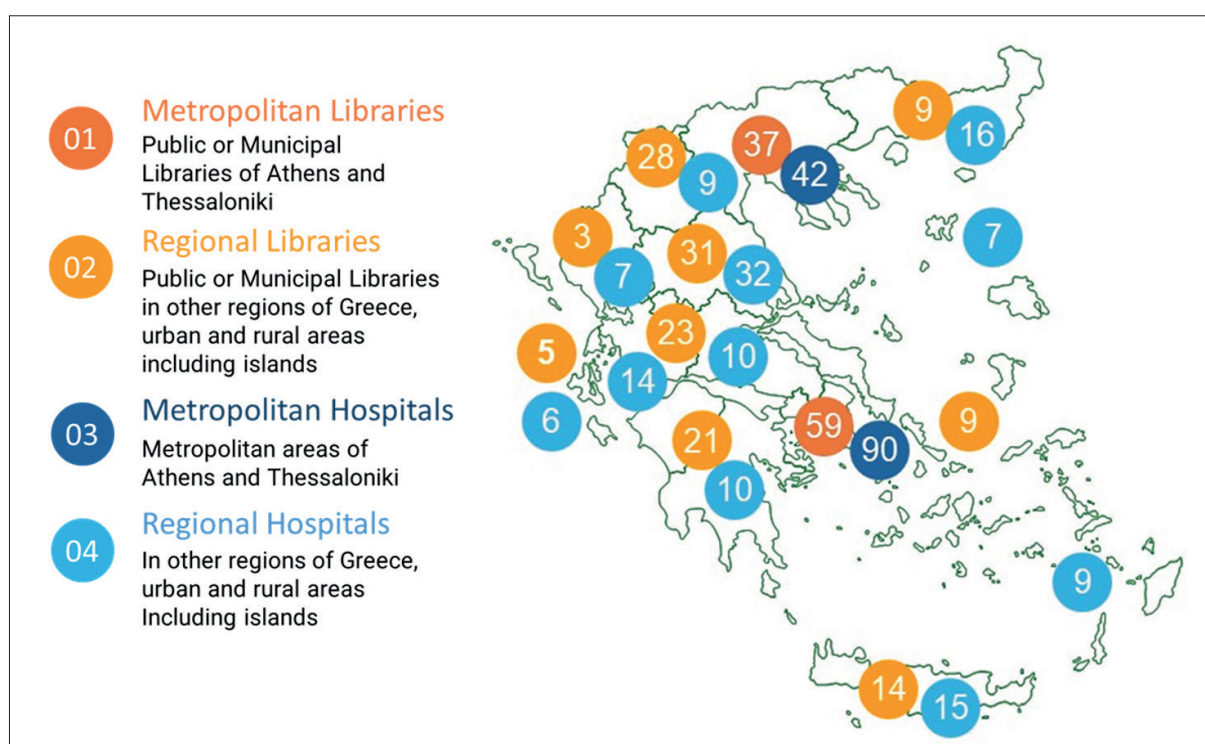


Figure 1: The Greek Libraries Network coverage map with the hospital coverage of regions of Greece.

Taking into account the current situation (state-of-the-art) and international practice, libraries in Greece (a) Form a panhellenic network covering both urban and rural areas, (b) can provide printed and digital material related to health issues (books, magazines, brochures, online databases), (c) they can perform health searches, (d) educate their audience on methods of searching for valid information, while (e) they are a safe, inclusive space for information and education. Therefore, they could be a pole for the development of regional actions to combat inequalities in Health, either due to limited access to therapeutic infrastructure due to geographical coverage,

language restrictions, social inequalities, economic inequalities, fear of stigmatization or ridicule, and discharge the pressure of the public health system in matters of individual counseling and medical information. At the same time, by contributing to public health literacy and combating limited public knowledge of health, prevention, safety and care, and misinformation, they can help reduce morbidity and improve public health.

To achieve these goals, it is necessary: (a) the central coordination, (b) targeted financial support, (c) the enrichment of the libraries' collections in medical records, (d) the strengthening of their human potential, (e) training programs for the existing staff, (f) the access to electronic digital health resources, and (g) the development of a popular source of digital health content for the general public. The inclusion of libraries in the "toolbox" of drawing up public health policy is expected to contribute significantly to the reduction of social inequalities and the improvement of the quality of life of citizens.

References

- Aguilar-Gaxiola, Sergio, Ahmed M. Syed, Anise Ayodola, Atum Azzahir, Kellan E. Baker, Anna Cupito, Milton Eder, Tekisha Dwan Everette, Kim Erwin, Maret Felzien, Elmer Freeman, David Gibbs, Ella Greene-Moton, Sinsi Hernández-Cancio, Ann Hwang, Felica Jones, Grant Jones, Marita Jones, Dmitry Khodyakov, J. Lloyd Michener, Bobby Milstein, Debra S. Oto-Kent, Michael Orban, Burt Pusch, Mona Shah, Monique Shaw, Julie Tarrant, Nina Wallerstein, John M. Westfall, Asia Williams, and Richard Zaldivar. 2022. *Assessing meaningful community engagement: a conceptual model to advance health equity through transformed systems for health: organizing committee for assessing meaningful community engagement in health and health care programs and policies*. NAM Perspectives. Washington, DC: National Academy of Medicine. <https://doi.org/10.31478/202202c>.
- Atkinson, Jeremy. 2016. *Quality and the academic library: Reviewing, assessing and enhancing service provision*. Cambridge, MA: Chandos Publishing.
- Audunson, Ragnar, Herbjørn Andresen, Cicilie Fagerlid, Erik Henningsen, Hans-Christoph Hobohm, Henrik Jochumsen, Håkon Larsen, and Tonje Vold. 2020. *Libraries, archives and museums as democratic spaces in a digital age*. Munich, Germany: De Gruyter Saur. <https://doi.org/10.1515/9783110636628>.
- Cooper, Lisa. 2021. *Why are health disparities everyone's problem?* Baltimore: Johns Hopkins University Press. <https://doi.org/10.1353/book.85070>.
- DeGuzman, Pamela B., Neha Jain, and Christine G. Loureiro. 2022. 'Public libraries as partners in telemedicine delivery: a review and research agenda.' *Public Library Quarterly* 41(3): 294-304. <https://doi.org/10.1080/01616846.2021.1877080>.
- ELSTAT (Hellenic Statistical Authority). 2022. Ελληνική Στατιστική Αρχή (ΕΛΣΤΑΤ). Research on Greek Libraries (except School Libraries) (2022) (Edition 1.0). https://www.statistics.gr/el/home?p_p_id=com_liferay_portal_search_web_portlet_SearchPortlet_INSTANCE_3&p_p_lifecycle=0&p_p_state=maximized&p_p_mode=view&_com_liferay_portal_search_web_portlet_SearchPortlet_INSTANCE_3_mvcPath=%2Fview_content.jsp&_com_liferay_portal_search_web_portlet_SearchPortlet_INSTANCE_3_assetEntryId=18242052&_com_liferay_portal_search_web_portlet_SearchPortlet_INSTANCE_3_type=document.
- ELSTAT (Hellenic Statistical Authority). 2023. *Hospitals and local health centres (beds, staff, equipment)/2021*. <https://www.statistics.gr/el/statistics/-/publication/SHE06/2021>
- Finn, Symma, and Liam R. O'Fallon, eds. 2018. *Environmental health literacy*. Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-319-94108-0>.
- Galaviz, Karla I., Jessica Y. Breland, Mechelle Sanders, Khadijah Breathett, Alison Cerezo, Oscar Gil, John M. Hollier, Cassandra Marshall, J. Deanna Wilson, and Utibe R. Essien. 2020. 'Implementation science to address health disparities during the coronavirus pandemic.' *Health Equity* 4(1): 463-467. <https://doi.org/10.1089/heq.2020.0044>.
- Geschke, Daniel, Lorenz Jan, and Peter Holtz. 2019. 'The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers.' *British Journal of Social Psychology* 58(1): 129-149. <https://doi.org/10.1111/bjso.12286>.

- Gursul, Deniz. 2022. *Health information: are you getting your message across?* National Institute for Health and Care Evidence. https://doi.org/10.3310/nihrevidence_51109.
- Horrigan, John B. 2015. *Libraries at the Crossroads*. Pew Research Center: Internet, Science & Tech. 2015. <https://www.pewresearch.org/internet/2015/09/15/libraries-at-the-crossroads/>
- Horrigan, John. B. 2016. *Libraries 2016*. Pew Research Center: Internet, Science & Tech. 2016. <https://www.pewresearch.org/internet/2016/09/09/libraries-2016/>
- Karki, Manisha, Marie Line El Asmar, Eva Riboli Sasco, and Austen El-Osta. 2024. 'Public libraries to promote public health and wellbeing: a cross-sectional study of community-dwelling adults.' *BMC Public Health* 24(1): 1226. <https://doi.org/10.1186/s12889-024-18535-5>.
- Kearns, Katherine, and Tracy Thrasher Hybl. 2005. 'A collaboration between faculty and librarians to develop and assess a science literacy laboratory module.' *Science & Technology Libraries* 25(4): 39-56. https://doi.org/10.1300/J122v25n04_04.
- Kelleher, Kelly J., and William Gardner. 2017. 'Out of sight, out of mind—behavioral and developmental care for rural children.' *N Engl J Med* 376 (14): 1301-1303. <https://doi.org/10.1056/NEJMp1700713>.
- Koltay, Tibor, Sonja Spiranec, and Laszlo Z. Karvalics. 2016. *Research 2.0 and the future of information literacy*. Cambridge, MA: Chandos Publishing.
- Macarena, Garcia C., Mark Faul, Greta Massetti, Cheryll C. Thomas, Yuling Hong, Ursula E. Bauer, and Michael F. Iademarco. 2017. 'Reducing potentially excess deaths from the five leading causes of death in the rural United States.' *MMWR. Surveillance Summaries* 66(2): 1-7. <https://doi.org/10.15585/mmwr.ss6602a1>.
- MoH (Ministry of Health). 2024. *Rights of Healthcare Services Customers*. <https://www.moh.gov.gr/articles/citizen/dikaiwmata-lhptwn-yphresiwn-ygeias>
- Morgan, Anna U., Roxanne Dupuis, Bernadette D'Alonzo, Andria Johnson, Amy Graves, Kiahana L. Brooks, Autumn McClintock, Heather Klusaritz, Hillary Bogner, Judith A. Long, David Grande, and Carolyn C. Cannuscio. 2016. 'Beyond books: Public libraries as partners for population health.' *Health Affairs* 35 (11): 2030-2036. <https://doi.org/10.1377/hlthaff.2016.0724>.
- NAS (National Academies of Sciences). 2024. *Exploring diversity, equity, inclusion, and health equity commitments and approaches by health organization C-suites: Proceedings of a workshop*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/27458>.
- NHS (National Health Service). 2021. *Knowledge for Healthcare* 2021-26. <https://www.hee.nhs.uk/our-work/knowledge-for-healthcare>.
- NLG (National Library of Greece). 2020. *Participating libraries by region of Greece*. <https://net-work.nlg.gr/library/>.
- Philbin, Morgan M., Caroline M. Parker, Mary Grace Flaherty, and Jennifer S. Hirsch. 2019. 'Public libraries: A community-level resource to advance population health.' *Journal of community health* 44: 192-199. <https://doi.org/10.1007/s10900-018-0547-4>.

Rid, Annette, Michael A. Johansson, Gabriel Leung, Hannah Valentine, Esteban G. Burchard, Sam S. Oh, and Cathy Zimmerman. 2016. 'Towards equity in health: researchers take stock.' *PLoS medicine* 13(11): e1002186. <https://doi.org/10.1371/journal.pmed.1002186>.

Rubenstein, Ellen. 2012. 'From social hygiene to consumer health: Libraries, health information, and the American public from the late nineteenth century to the 1980s.' *Library & Information History* 28 (3): 202-219. <https://doi.org/10.1179/1758348912Z.00000000016>.

Scales, David, Jack Gorman, and Kathleen H. Jamieson. 2021. 'The Covid-19 infodemic—applying the epidemiologic model to counter misinformation.' *New England Journal of Medicine* 385 (8): 678-681. <https://doi.org/10.1056/NEJMp2103798>.

Sentell, Tetine, Sandra Vamos, and Orkan Okan. 2020. 'Interdisciplinary perspectives on health literacy research around the world: more important than ever in a time of COVID-19.' *International Journal of Environmental Research and Public Health* 17 (9): 3010. <https://doi.org/10.3390/ijerph17093010>.

Sieck, Cynthia J., Amy Sheon, Jessica S. Ancker, Jill Castek, Bill Callahan, and Angela Siefer. 2021. 'Digital inclusion as a social determinant of health.' *NPJ digital medicine* 4 (1): 52. <https://doi.org/10.1038/s41746-021-00413-8>.

Solar, Orielle., and Alec Irwin. 2010. *A conceptual framework for action on the social determinants of health*. Geneva: WHO Document Production Services <https://doi.org/10.13016/17cr-aqb9>.

Thomas, Sharita R., George M. Holmes, and George H. Pink. 2016. 'To what extent do community characteristics explain differences in closure among financially distressed rural hospitals?.' *Journal of Health Care for the Poor and Underserved* 27 (4): 194-203. <https://doi.org/10.1353/hpu.2016.0176>

UNC (Cecil G. Sheps Center for Health Services Research). 2020. *Sheps Center for Health Services Research, University of North Carolina at Chapel Hill. Rural Hospital Closures*. <https://www.shepscenter.unc.edu/programs-projects/rural-health/rural-hospital-closures/>.

Warshaw, Robin. 2017. *Health Disparities Affect Millions in Rural U.S. Communities*. AAMC <https://www.aamc.org/news/health-disparities-affect-millions-rural-us-communities>.

WHO (World Health Organization). 2019. *Social determinants of Health*. https://www.who.int/health-topics/social-determinants-of-health#tab=tab_3.

From Manuscript to Metadata: experiments on Handwritten Text Recognition, Tagging and Importation for the *Memoriali* series (1265-1452)*

Edward Loss^(a)
Fabiana Guernaccini^(b)
Manuel Carassai^(c)

a) University of Genoa, <https://orcid.org/0000-0002-9837-8321>

b) Regesta.exe, <https://orcid.org/0000-0002-2101-882X>

c) Regesta.exe

Contact: Edward Loss, edwarddettmamloss@gmail.com; Fabiana Guernaccini, fguernaccini@regesta.com;
Manuel Carassai, mcarassai@regesta.com

Received: 30 December 2024; **Accepted:** 21 March 2025; **First Published:** 15 May 2025

ABSTRACT

The paper describes the experiment of creating a tailored tagging system for the *Memoriali* – a series of notarial records produced in Bologna almost uninterruptedly between the second half of the 13th and first half of the 15th century – using the Handwritten Text Recognition (HTR) software Transkribus. Product of the conjoint effort of the MemoBo project and Regesta.exe, the aim of this system is to facilitate the export of a specific set of information from these documents directly to a customised database on xDams. This tagging system explores a secondary feature of Transkribus, highlighting its potential and its limits for establishing a coherent workflow, which initiates with the application of *ad hoc* models for creating semi-automatic transcriptions of these late medieval notarial records and concludes with the exportation and proper placement of their data inside our xDams database.

KEYWORDS

HTR; Transkribus; Memoriali; xDams; Bologna.

* This work was developed as part of the European Union initiative – Next Generation EU, mission 4, component 1, ON: *Objects in network. The social life of things in the fifteenth century between notarial sources and semantic web*, PRIN2022 (P. I. Tommaso Duranti), CUP:J53D23000510006; Code MUR: 2022XTSEZ3_001. Edward Loss is the author of sections 1, 2, 4 and 5; Fabiana Guernaccini and Manuel Carassai are the authors of sections 3 and 6. All three authors wrote section 7.

1. Introduction

This paper describes the issues and challenges of creating a tailored tagging system for the *Memoriali* – a series of notarial records produced in Bologna almost uninterruptedly between the second half of the 13th and first half of the 15th century – using Transkribus and aimed at facilitating the export of a specific set of information from these documents directly to a customised database on xDams. Product of the conjoint effort of the MemoBo project and Regesta.exe, this tagging system explores a secondary feature of the Handwritten Text Recognition (HTR) software, highlighting its potential and its limits for establishing a coherent workflow, which initiates with the application of *ad hoc* models for creating semi-automatic transcriptions of the late medieval notarial records just mentioned and concludes with the exportation and proper placement of their data inside our xDams database. The article is thus divided into four main sections: the first one concerns the *Memoriali* project, its objectives and its custom-made database, produced by Regesta.exe. The categories of the database predate the project's entire experimentation with HTR, imposing, therefore, the key-elements of the tagging system; a peculiarity which marked some of its features. For this specific reason, this section includes a detailed and exhaustive description of the database.

The second section focuses on the experimentation of the MemoBo project with HTR, and more specifically, with Transkribus. It briefly describes a tortuous process, which culminated with the creation of Large HTR Models based on *Ground Truth*¹ (Hodel 2022) on the scale of hundreds of thousands of words, capable of not only successfully transcribing the late medieval Bolognese *Memoriali*, but also demonstrating potential for geographical transposition in the same chronological context. The section highlights how the projects' initial goal of producing semi-automatic transcriptions in order to aid the manual population of the database changed considerably once we realised Transkribus' tagging feature left an imprint in the XML version of the transcribed material, which could be potentially used by Regesta.exe to export this material directly into the xDams database.

Sections three and four constitute the core of the paper. They illustrate, on one hand, the different levels and elements of the customised tagging system, justifying how some of the decisions made were due to structural limitations of Transkribus. On the other hand, they specify all the key issues of mapping and converting a Transkribus XML tagged with this custom system into an XML file compatible with a xDams XML file for proper semi-automatic exportation.

The paper concludes with the complete structure of this tagging system, offering some general considerations on how its implementation constitutes a time-saving tool in comparison with manually cataloguing the *Memoriali*.

¹ In the context of Machine Learning and of Handwritten Text Recognition, the term *Ground Truth* means the portion of the material (transcribed text) provided to the software which will serve as the reference for any of its autonomous attempts to transcribe scanned text. Also known as *Golden Rule*, HTR software consider *Ground Truth* material as its source of learning, since these systems are not taking input from the open web. *Ground Truth* can either be generated completely manually by the user or by applying a previously existent model (such as “public models”, those made available by pre-existent research groups) and correcting the resulting transcriptions.

2. The MemoBo project

Created as a partnership between the University of Bologna, the State Archive of Bologna and Regesta.exe in 2020, the MemoBo project constitutes a big data initiative centred on a rich notarial late medieval series – the *Memoriali* – of nearly three million records produced in Bologna between 1265 and 1452 (Bruno et al. 2023). The scope of the *Ufficio dei Memoriali* – the institution from which the series and its surviving records borrow their name – was to safekeep the authenticity of notarial acts created inside the city’s jurisdiction. A statutory disposition established its existence in 1265 declaring that any notarial act valued on at least 20 Bolognese *lire* that was not properly brought to the office’s attention should not be considered valid (Continelli 1988, IX-XL). This legal obligation is at the root of the extraordinary value of the series for the study of the economic and social life of a large late medieval commune, which constitutes an almost unique case of a documentary series of this sort preserved almost uninterruptedly for nearly 200 years. The aim of the MemoBo project is to render available the information contained in the 322 tomes that hold these almost three million acts, transforming it into searchable, shareable and reusable data for scholars and the general public. Only a set of fundamental information is extracted from these records – at this stage the project’s goal is not to provide complete transcriptions of these sources – and the core of MemoBo is the custom-made database produced in xDams to host this data.

3. The database in xDams

As mentioned in the introduction, the experimentation with Transkribus constituted a posterior strategy of the initial objectives of the MemoBo project. Therefore, the entire experimentation and mapping operations were executed in the already implemented database. A parallel and temporary database was created to host the test data, considering a possible subsequent transfer of the material to the main database.

MemoBo’s xDams database adopts a hierarchical structure, allowing for the organization of levels for the description of volumes, registers, deeds, and finally, properties – movable, immovable, and rights - which are the subject of the deeds contained in the *Memoriali*.

The descriptive entry for “volume” and “register”

The databases’ initial macro-level entries are “volume” and “register”. These two allow the inclusion of identifying information such as:

- denomination
- registering notary
- chronology
- notes

The descriptive entry for “deed”

Under those two more general entries, one finds the entry “deed”, which allows the insertion of information regarding the individual notarial record, both from a physical and content perspec-

tive. This entry, thus, contains information such as *folio* identification, the position of the deed inside a specific *folio*, and the language of the document. These details trace the physical perimeter of the object, but also contain some extra information: in addition to the drafting and recording notary, the entry collects data on the topical date, the record's registration date, and indexes all of the people who are mentioned in the deed, with their respective roles. For each person, it is possible to insert details such as gender, name, surname, age, origin or residence, qualification, belonging or reference entity, condition, profession, function inside a specific deed, and any other eventual note. Each deed may have as its object one or more assets or properties, described in a specific sheet to maintain the peculiarities of each type: “movable properties”, “immovable properties” and “rights”.

The descriptive entry for “property”

The entry allows the description of the three types just mentioned, after selecting the type via a drop-down menu.

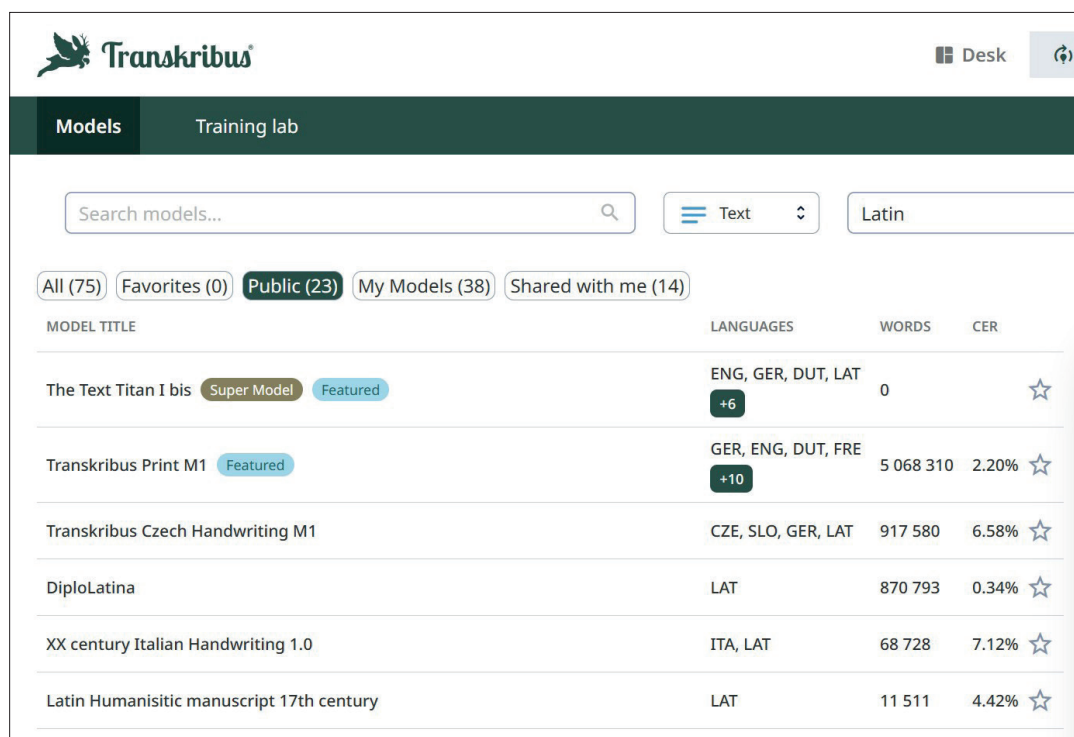
Figure 1. The first part of the descriptive entry for property on xDams.

The image provides an example of the creation of a property record. In the depicted case, the property is categorized as “rights”, but it is equally possible to create records for “movable property” or “immovable property”. The input form is organised into sections, filled out differently depending on the type of object previously selected. Subsequently, it is possible to specify the quantities, measurements, value, rent, and an index of the names of the people and places associated with the individual property. Together all these different categories of information constitute a set of over 200 potential elements that could be extracted from each deed of the *Memoriali* series.

4. MemoBo and HTR

The colossal dimension of the series together with the rather complex and detailed set of information selected for transposition into the database motivated the MemoBo project to look for digital solutions that could complement and overcome the time constraints imposed on manual cataloguers. In 2022, the project started experimenting with Handwritten Text Recognition software, especially with Transkribus, thanks to its popularity in the sector and its relatively user-friendly interface, in particular to scholars who have no background in computer sciences. The initial aim of the experiment was to produce fairly readable transcriptions with the software that could expand the pool of cataloguers recruited for the project – with transcriptions, no longer would we have to rely only on cataloguers highly skilled in Palaeography and late medieval Latin to populate the database.

The experimentation was divided into three phases: the first one focused on the creation of *pure models*, based on a single notary – a “single hand” – and was especially centred on the first volumes of the collection from the late thirteenth-century, which possesses a clearer handwriting from a palaeographical perspective. The idea was to use these models as *base-models* for more daring attempts to create *mixed-models* and *mega-models* able to cover bigger chronological portions of the collection. This initial approach was mostly motivated by the fact that there are no *public models* available at the platform that focuses both on late medieval Latin and on documents produced in the peninsula. As a matter of fact, the timeframe from the second half of the thirteenth- all the way into the first decades of the fourteenth-century is largely uncovered by the research groups that decided to make their models available to others in the open platform.



MODEL TITLE	LANGUAGES	WORDS	CER
The Text Titan I bis Super Model Featured	ENG, GER, DUT, LAT +6	0	☆
Transkribus Print M1 Featured	GER, ENG, DUT, FRE +10	5 068 310	2.20% ☆
Transkribus Czech Handwriting M1	CZE, SLO, GER, LAT	917 580	6.58% ☆
DiploLatina	LAT	870 793	0.34% ☆
XX century Italian Handwriting 1.0	ITA, LAT	68 728	7.12% ☆
Latin Humanisitic manuscript 17th century	LAT	11 511	4.42% ☆

Figure 2. The 17 Latin Text Recognition Public Models available on Transkribus.

In this first phase of the experiment, we decided to measure the time spent in each step of the creation of these *pure models* – from uploading images to *training*, including layout and *Ground Truth* preparation. This data could allow us to consider if Transkribus and HTR-based software were viable options for the project: unless the overall time spent in all phases up to the production of the AI-based transcriptions was significantly lower than the time needed for human-created transcriptions, this tool could not be considered viable for the project, especially taking into account it is a paid software. We produced 26 models in this first phase, opting for a system of *Ground Truth* creation based on the insertion of *direct transcriptions* – i.e. Latin text with expanded abbreviations and free of graphic signs – which proved to be an efficient system for the specific purposes of the project. The other option, an input system of *diplomatic transcriptions* – i.e. abbreviated Latin text containing graphic signs more coherent for the relationship between image and text – not only required much more time for the production of *Ground Truth* comparatively, but also revealed incompatibility issues between Transkribus and Unicode platforms specialized in characters of medieval manuscripts, such as MUFI (*Medieval Unicode Font Initiative*).

One model that is particularly representative of the group is the MemoBo, Me. 69, Model 4, created with 50 pages – 44 in the *Train Set* and 6 in the *Validation set* – in all 16243 words.

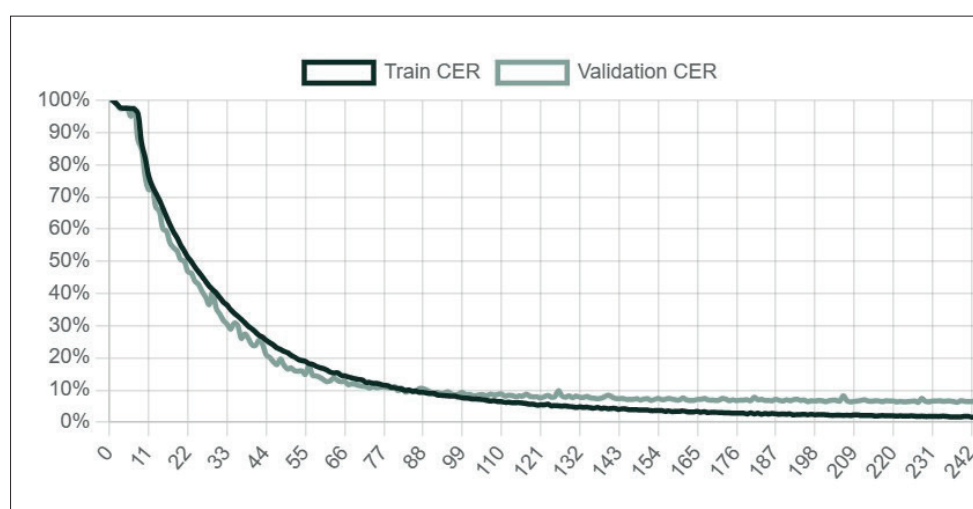


Figure 3. The Learning Curve of MemoBo, Me. 69, Model 4.

We spent roughly 30 hours in the production of this model from scratch and its parameters are the following: 1,60% of CER (Character Error Rate) on train set² and 6,30% of CER on validation,³

² The Character Error Rate (CER) on train set consists of the numbers of errors in terms of characters produced by the software when trying to learn from the material – the *Ground Truth* – provided to the machine. The percentages given represent the portion of the text and of the operations used to created it – for example, expanding abbreviations – that the machine did not understand. Therefore, a CER on train of 1,60% means the machine understood and took into consideration 98,4% percent of all the elements and processes present in the *Ground Truth* given to it.

³ The Character Error Rate (CER) on validation set consists of the number of errors in terms of character produced by the software when trying to apply everything it learned from the *Ground Truth* in producing its own transcription. To establish this rate, the software takes one a pre-selected page of the *Ground Truth*, erases it completely then attempt to transcribe

which allow us to consider this model as already optimal according to the Transkribus guidelines. In practical terms, it means that a transcription using this model with this specific material – a single-hand script – contains one mistake every 15 words.

This was an excellent first outcome, especially considering the CER parameters provided by the guidelines of the software, as becomes clearer in the table below:

	CER	Training Data
<i>Printed text</i>	0,5-2%	~ 5.000 words / 25 pages
<i>Single hand simple writing</i>	2-4%	+10.000 words / 50 pages
<i>Several hands – all seen during training</i>	4-6%	several 10.000 words training data / several 50 pages
<i>Many hands from same period and region – not all seen during training</i>	6-8%	+100.000 words / more than 500 pages

Table 1. Transkribus' guideline parameters for evaluating the CER on Validation

Despite this excellent outcome, these 26 models demonstrated no versatility whatsoever, with attempts to apply them individually to other units of the collection composed by different notaries failing considerably, as expected.

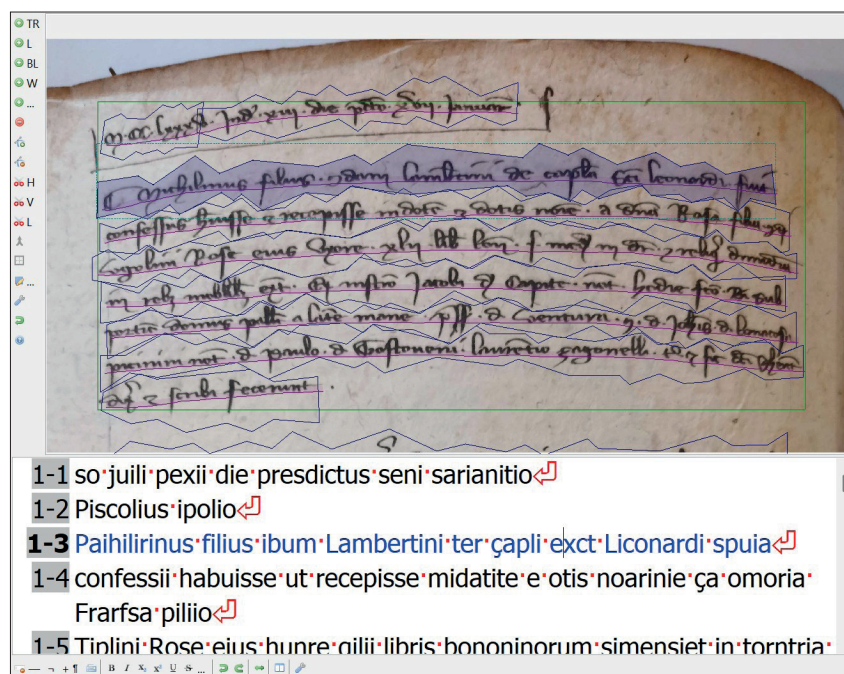


Figure 4. Attempt to apply the MemoBo, Me. 69, Model 4 on a different volume of the *Memoriali*.

it. The resulting transcription is then compared to the original to determine how many mistakes the machine made. This rate is usually considerably higher than the CER on train, since it is an active and generating operation. Therefore, a CER on validation of 6,30% means that the software produced correct transcriptions compared to the selected examples of the *Ground Truth* 93,70 % of the time.

The second stage of the experiment initiated the merging of these 26 models in different combinations to create the first multi-hand *mixed models*. Interestingly, the resulting models started to demonstrate incredible flexibility as early as we introduced a second hand to a previously *pure model*. An example is the MemoBo, Me. 65 and 69, Model. 3, composed of 50 pages of each of these units – both of them from the late thirteenth-century – and presenting a CER on train of 4,90% and a CER on validation of 9,70%.

MemoBo, Me. 65 and 69, Model. 3	
N° of GT pages	100
CER on train	4,90%
CER on validation	9,70%
Hours of work	50

Table 2. Parameters of a MemoBo *mixed model*

These rates are higher than the individual *pure models*, but still optimal, especially considering that this model is composed of mixed material in terms of resolution – half of these 100 images are professional and the other half are amatorial reproductions. An attempt to apply this dual-hand model on material from the first years of the fourteenth century – units not included in the production of its *Ground Truth* – resulted on some perfect lines of transcription even on amatorial reproductions, allowing us to reach an initial conclusion: the software has less difficulties in handling the late medieval abbreviation system than the significant variation of notarial hands inside the collection. This outcome indicated a promising path to the experiment and allowed us to move forward trying on more complex units of the *Memoriali* collection dating from the second half of the fourteenth century.

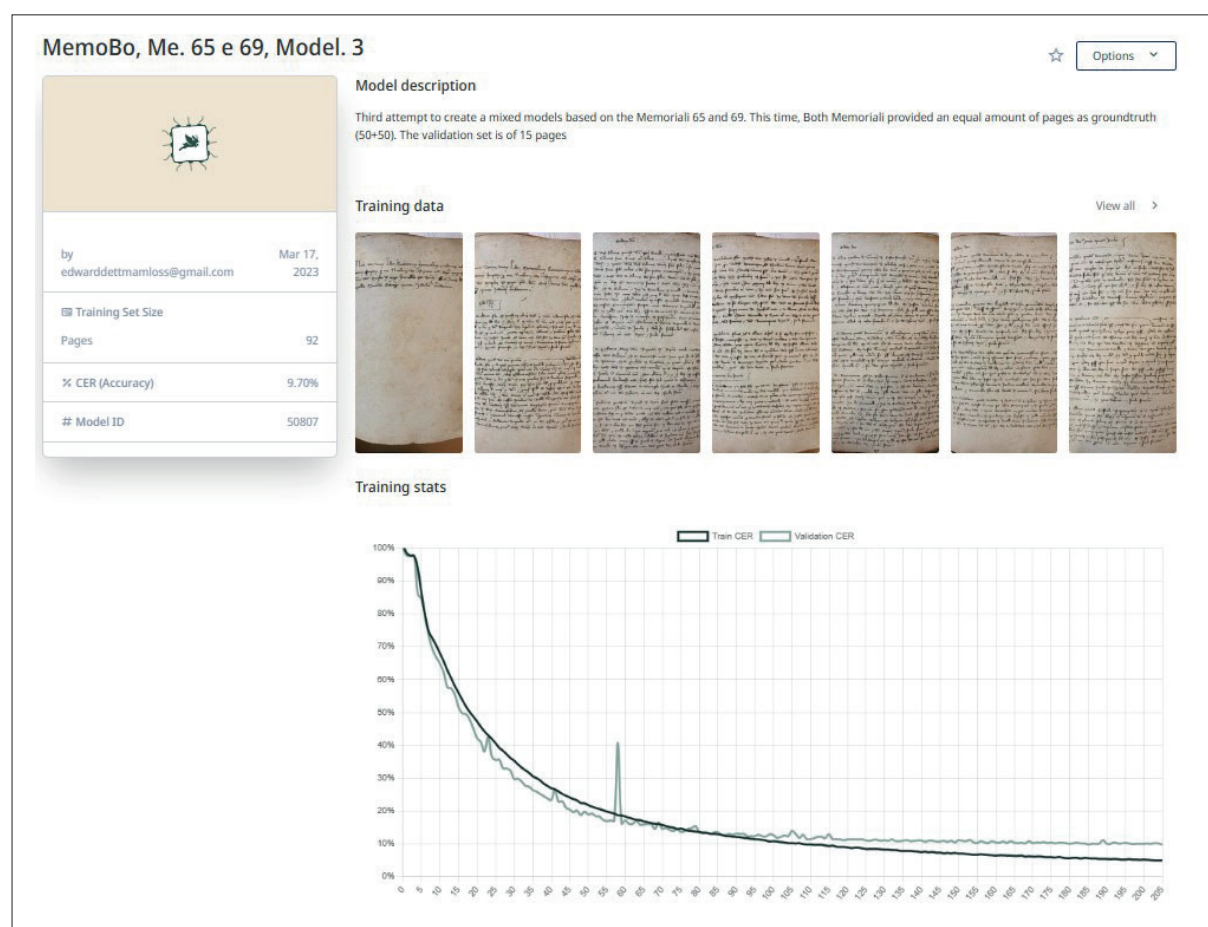


Figure 5. Training set and Learning Curve of MemoBo, Me. 65 and 69, Model. 3.

The third phase of the experiment drew from the previously created *pure models* and these *mixed models* to face the most challenging part of the *Memoriali* collection: the fifteenth-century units, which sometimes contain 40 different hands in a single volume, some of them represented only in fragments of a few *folios* each. The only method capable of treating these units was to invest on *Mega Models*, with large quantities of *Ground Truth*. Apart from substantially augmenting the number of pages and words, the creation of these *mega models* also counted heavily on the previous experiences by the direct inclusion of the *mixed models* above-mentioned on their training. The first one of this category was the MemoBo, Me. 320 and 321, Model 13, created with a *Train Set* of 310 pages (385712 words) and a *Validation Set* of 47 pages (15% of the *Train Set*), proportionally selected from the 40 notarial hands that compose both of the fifteenth-century volumes of the series. We chose to include a combination of *mixed* and *pure models* previously created on thirteenth- and fourteenth-century volumes of the *Memoriali* (MemoBo, Me. 1 and 58, Model. 3 and MemoBo, Me. 298, Model. 5) and the resulting parameters were a CER on *Train* of 4,27% and a CER on *Validation* of 5,78%, with the peculiarity of a *Learning Curve*⁴ that started its learning process already under a scale of 50%.

⁴ A *Learning Curve* represent the learning process of the machine based on the *Ground Truth* provided to it. The X-axis

MemoBo, Me. 320 and 321, Model. 13	
N° of GT pages	310
CER on train	4,27%
CER on validation	5,78%
Base models	MemoBo, Me. 1 and 58, Model. 3 MemoBo, Me. 298, Model 5

Table 3. Parameters of a MemoBo *mega model*

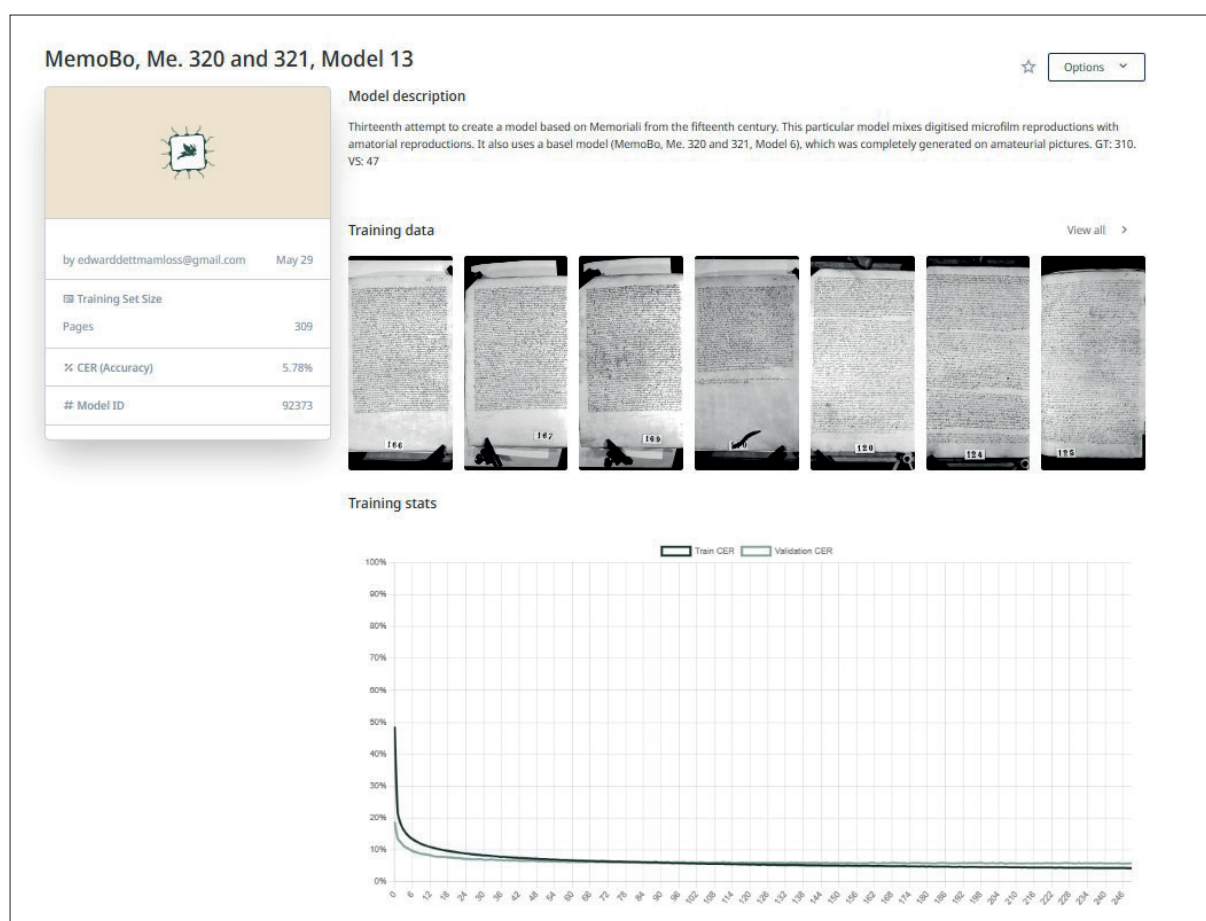


Figure 6. Training set and Learning Curve of *MemoBo, Me. 320 and 321, Model. 13*.

represent the number of times the software read through the *Ground Truth* material, also called cycles. In this example, the software processed the entire material from beginning to end 250 times, doing thus 250 cycles of learning. The Y-axis represents the Character Error Rate in Train and the Character Error Rate in Validation in each of these cycles, allowing the user to follow closely the accuracy of the machine in each of these learning attempts.

In practical terms, the precision of these *mega models* was such as to allow us, for the first time since the beginning of the experiment, to try applying these models produced exclusively with Bolognese documentation on other fifteenth-century notarial records from the Italian peninsula. Bologna and its notaries highly influenced notarial practices in the Late Middle Ages, with notaries, such as Rolandino Passeggeri, who even defined the canons of the notarial practice more generally in the period (Tamba 1998). His *Ars Notarie* was a true best-seller even north of the Alps, and traces of Rolandino's innovative protocols and *formulae* resonate in documentation produced in many cities of modern Southern France and Germany (Tamba 2002). The idea was to verify how much this theoretical influence was translatable in terms of HTR, since *mega models*' main characteristic is their ability to overcome hand variety inside documentary *corpora* of the same chronology. If in terms of format and documental structure other cities followed the Bolognese example, models based in large quantities of *Ground Truth* produced on these documents should, in theory, be able to read successfully records from elsewhere.

We selected Genova and Perugia for this specific part of the third phase of the experiment, since both cities are the ones with the largest collections of surviving notarial records of the peninsula (Macchiavello and Rovere 2018, 43-44; Bartoli Langeli 2006). This ongoing attempt has already demonstrated a lot of potential: for example, attempts to apply the same MemoBo, Me. 320 and 321, Model 13 on the production of notaries, such as the Genoese Antonio de Inghibetis de Castro, without including any additional *Ground Truth* – either from this specific geographical context nor from this specific notary – produced transcriptions with precision verified with Transkribus' tool of *compute accuracy* (still only available on the Expert version) on the scale of 70%, especially applying *compare text versions*. This promising outcome signals larger *mega models* than the MemoBo, Me. 320 and 321, Model 13 could become reference *base models* to explore late medieval notarial documentation on a national or even international level.

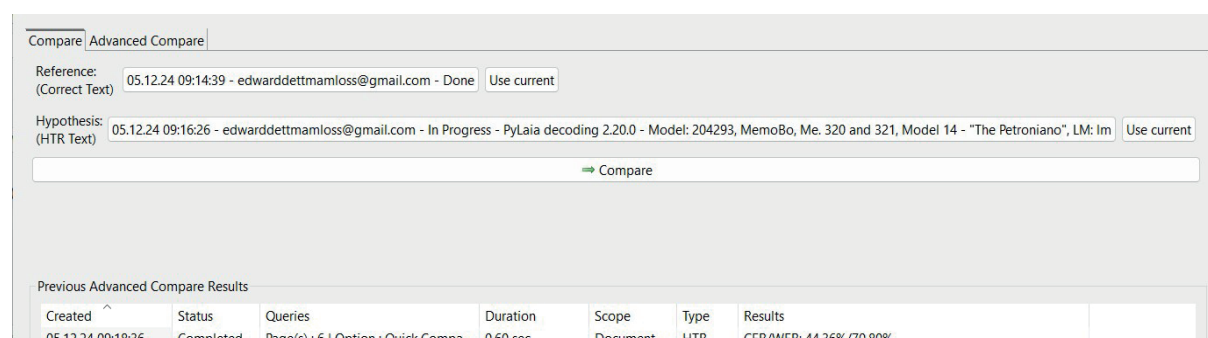


Figure 7. Compare text versions function applied on Antonio de Inghibetis records.

While investing time on the creation of these *mega models*, we also started exploring other functionalities of Transkribus that could have been useful for the project. Since we are interested in extracting specific information from these records, the existence of a customizable tagging tool in Transkribus could offer us the possibility to expedite the input of information inside our database.

5. Transkribus and Tagging

While HTR is the main feature of Transkribus, the software also allows users to mark (*tag*) specific parts of the uploaded images and generated text both during pre-processing and post-processing. The function of this tagging feature is varied and it operates in two levels. First, tags can be *structural*, meaning that they intervene directly on the image section of Transkribus and on the different *Text Regions*⁵ created during layout generation (fig. 8). Second, they can be *textual*, interfering – as its name suggests – directly on the transcriptions, and while *structural tags* cannot be juxtaposed, it is possible to overlay *textual tags* in many different levels (fig. 9).

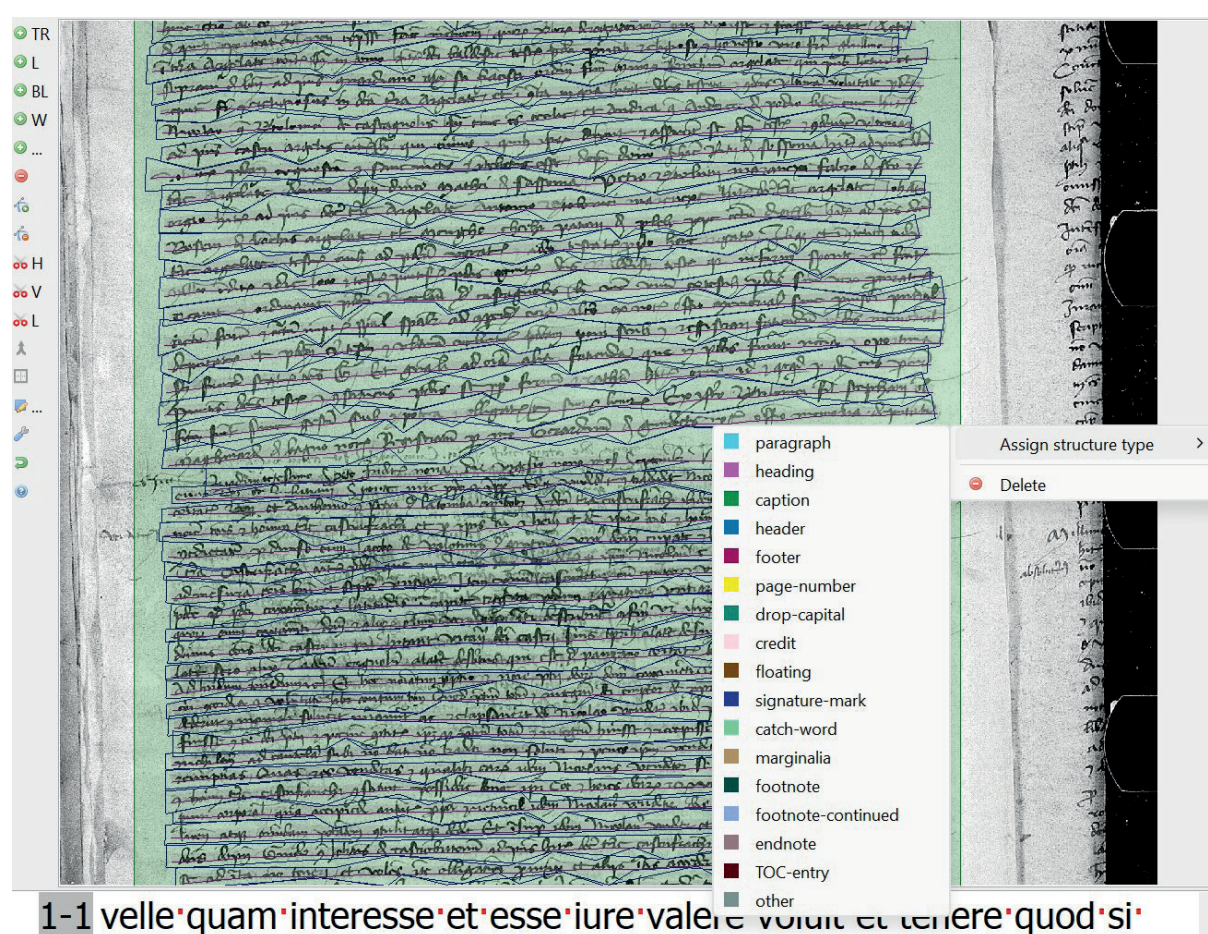


Figure 8. Text Regions and Structural Tags.

⁵ A *Text Region* (TR) is the space within an image that contains the text that user considers to be relevant both for the machine to learn and to transcribe. It does not need to correspond to all text existent inside an image, as it can be seen in Figure 8, where a lateral column of text coming from a different folio was not encased in a text Region on purposes. Elements not included in a Text Region are not even seen by the software.

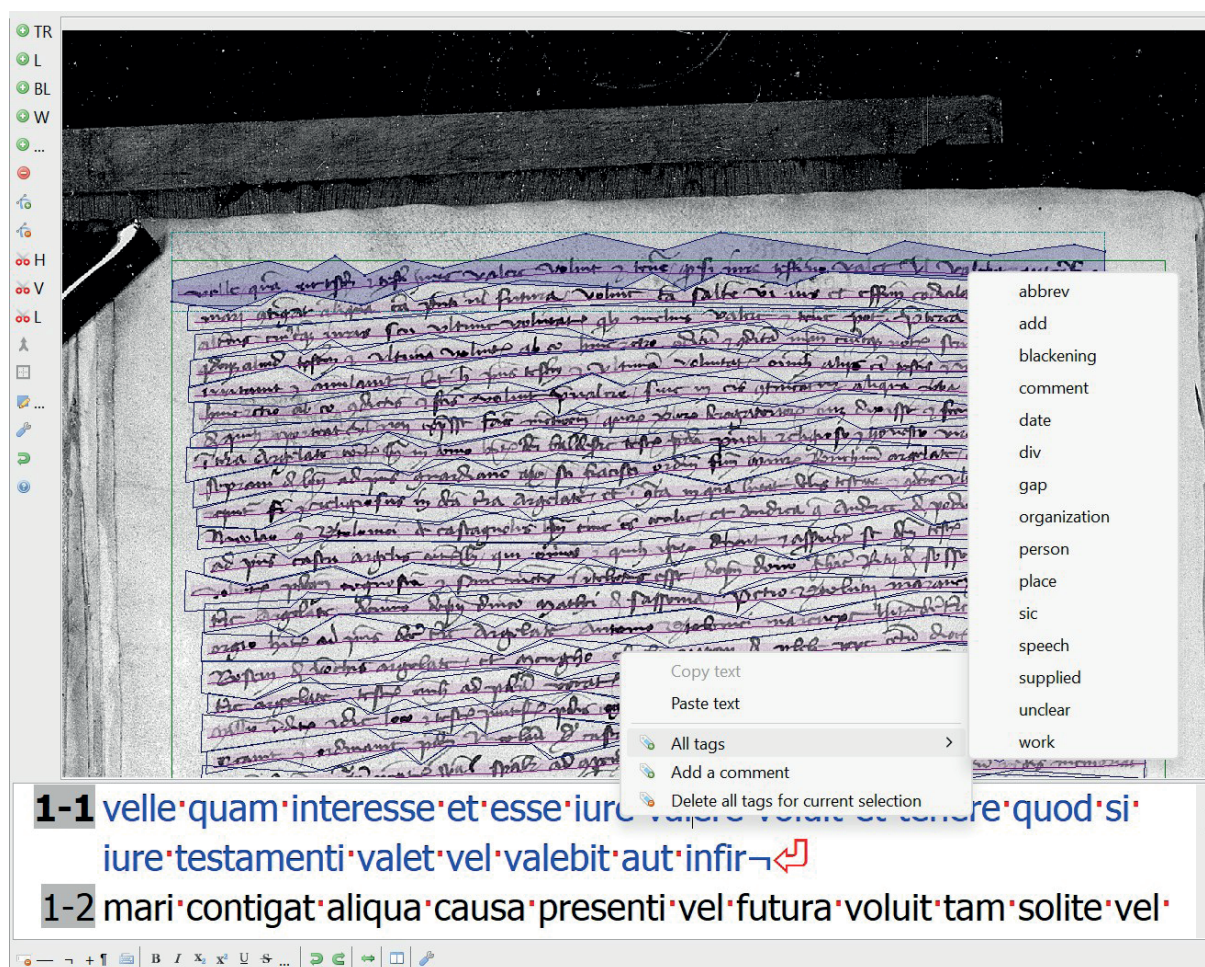


Figure 9. Transcriptions and Textual tags.

Tags leave a substantial imprint on the XML version of exported transcription files, which the user only has access to after properly exporting the material from Transkribus – an important limitation, as highlighted in sections 3 and 4 of the paper. *Structural* tags create and label *Anonymous Blocks* (<ab>) while *textual* tags only operate inside *Line Breaks* (<lb>). While the software offers users with a package of pre-established tags in both categories, it is possible to customize this tool without imposed limitations on the number of tags. While these customised tags must be inserted individually and by hand in the software, once they are created they become an accessible part of the feature in every page, document and collection, turning them into recursive elements (figure 10).

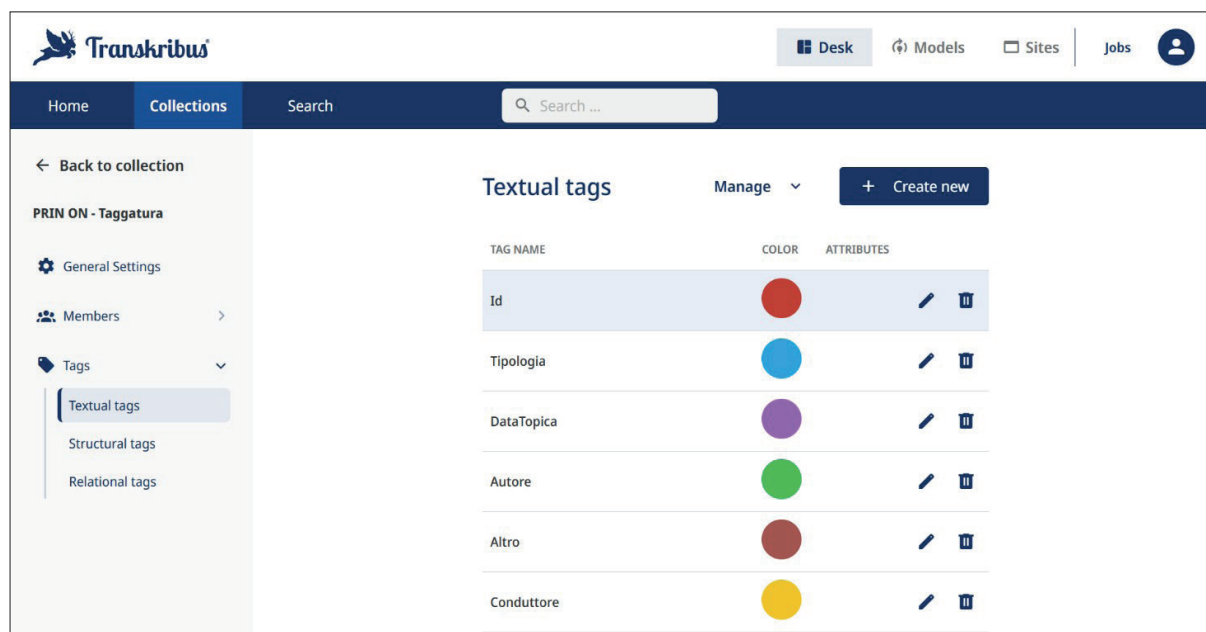


Figure 10. Customised tags as an accessible feature to all collections⁶

Creating a customized tag system

Considering these specificities, the first step to create a customised tagging system on Transkribus that somehow mirrored the categories of our xDams database was to render the software's *Text Regions* circumscribed to each individual act contained in one *folio* of the *Memoriali*. The easiest way to do so was to slice a single macro-*Text Region* automatically generated with a standard Transkribus layout model, posteriorly adding to each of these resulting different *Text Regions* a single *structural tag* purposely created and called <Deed> (record). As mentioned above, the record constitutes the basic unit of our xDams database, therefore, the *structural tag* <Deed> generated the base within which we built our tagging system. Records, however, are not circumscribed to a single *folio*, sometimes continuing on the *verso* side of the manuscript, or – especially in the case of later volumes of the *Memoriali*, produced after 1321 – through several folios. Transkribus structural tags are circumscribed to a single photo, not allowing users to create tags that run from one picture to the other. We came across this limitation with the first attempt of exporting the material into the xDams database – the more technical aspects of this operation in its different phases are better described in the next section – and solutions to overcome this restriction, thus respecting the documentation's basic unit, surfaced only after continuous exchange of ideas and collaboration among the members of the project who are responsible for the Transkribus direct experimentation, in one side, and those more involved with the database and its population in the other. We did not operate in separate compartments; therefore, the creation of each new tagging element was

⁶ The screenshots are taken from the applications used for the project, which are currently set to Italian. To facilitate readability, we have translated the terminology into English; however, it was not possible to translate the applications themselves at this time. In the screenshot in Figure 7, “tipologia” stands for “type”, “data topica” stands for “topical date”, “autore” stands for “author”, “altro” stands for “other” and “conduttore” stands for “conductor”.

followed by an importation attempt, which highlighted interruptions in the basic structure of the act. A strategy to solve them was to create another structural tag <DeedIntegration>. In order to reconduct a certain structural tag <DeedIntegration> to its correspondent <Deed>, we started intervening and populating these resulting *Text Regions* with *textual* tags. The first one we created was a *textual* tag <id>, which, once placed on a user introduced number in both the structural tag <Deed> and <DeedIntegration>, could easily produce an unequivocal element – necessary criteria, again better explained in the next section, for proper importation.

After the *textual* tag <id>, we created other tags on the textual level that mirrored our database and its inherent “record” categories. For example, we invented the tag <TopicalDate> to mark the portion of the automatically transcribed text that corresponded to this field of our database and so on.

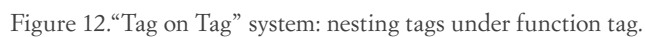
During our first attempt to implement all of these textual tags, we noticed that whenever the information changed from one line to another, Transkribus ruptured our *textual* tags into two different ones of the same category. This peculiarity forced us to operate with the same strategy as the one chosen for the creation of the structural tag <Deed>, which required the creation of a <DeedIntegration> tag. Therefore, all of the *textual* tags developed during this first attempt also included their “integration” version. This case is also better explored in greater detail in the next section.

Another problem that emerged was connected to all information regarding the people involved in the transactions described in the *Memoriali*. The multiplicity of individuals involved in a single deed, together with the intent of the *MemoBo*’s project to classify very specific details from each one of them, imposed a series of potential unwanted juxtapositions of general *textual* tags, such as <name> and <surname>, which could be applied to each and any of the people contained in the acts. A solution to this problem concerning the absence of univocal entries was to classify individuals inside a deed according to their function, and then overlay all these general textual tags to this specific function. We called this operation of overlaying two levels of *textual* tags as the “*Tag on Tag system*”.

The Tag on Tag system

When tagging the various deeds inside Transkribus following this system, we first inserted a tag corresponding to one of the 18 functions with which we classified agents in our xDams database (fig. 11). Once applied, we proceeded nesting inside this first macro-textual tag concerning the *function* all of the more general ones, such as <name> and <surname>, carefully paying attention to not juxtapose their extremities in order to not corrupt this hierarchical scheme (fig. 12).

Figure 11. The list of 18 functions of agents in the xDams platform.



Since many different agents can exercise the same function inside a certain record, we also introduced the previously mentioned tag <id> inside each of these nests organised according to function in order to preserve their univocity (fig. 13 and 14).

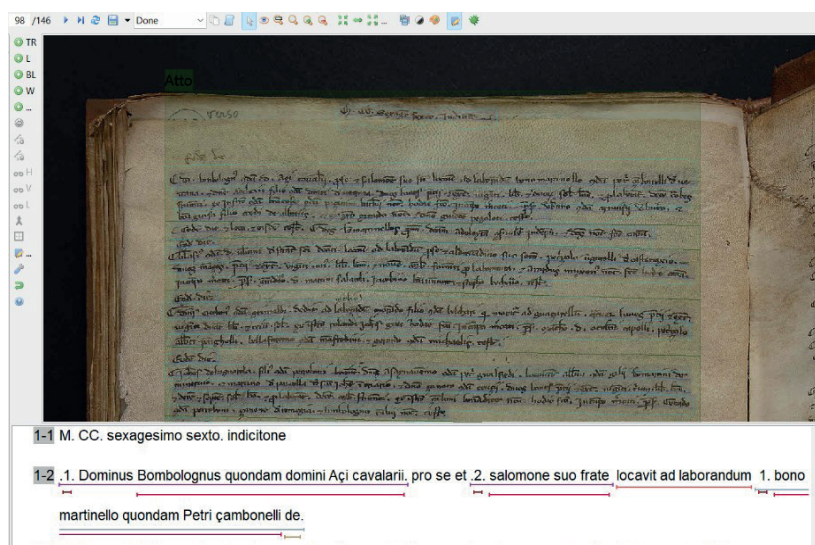


Figure 13. “Tag on Tag” system: textual tag <id>.

```

<Witness>
  <id>1</id>
  . dominus
  <Name>branchaleonus</Name>
  <Surname>de lambertacius</Surname>

</Witness>
et dominus
<Witness>
  <id>2</id>
  <Name>madius</Name>
  <Surname>de lambertacius</Surname>

</Witness>
<lb facs="#facs_105_tr_1_tl_7" n="N006"/>
et dominus
<Witness>
  <id>3</id>
  <Name>guillelmus bocatiis</Name>

</Witness>
et dominus
<Witness>
  <id>4</id>
  <Name>Lambertinus quondam domini mulinaroli</Name>

</Witness>
et
<Witness>
  <id>5</id>
  <Name>marchisinus</Name>
  <Surname>de datis</Surname>

</Witness>
<Witness>
  <id>6</id>
  <Name>guidonus</Name>
  <Craft>sartor</Craft>

</Witness>
et
<Witness>
  <id>7</id>
  <Name>Guilleminus</Name>

</Witness>
<lb facs="#facs_105_tr_1_tl_8" n="N007"/>
<WitnessIntegration>

```

Figure 14. Tag <id> on Transkribus-generated XML file.

Creating and tagging a “level object”

Our xDams database contemplates all properties, assets and rights contained inside a late medieval notarial record as a separate hierarchical level, as described in the previous session. This peculiar configuration imposed a different logical scheme for tagging all portions of the *Memoriali* records that concerned this specific information, since it rendered necessary to detach elements that appear in the original records as united. In order to separate these assets into a portion of the record that could be easily read in the posterior phase of importation – described in detail in the next topic – we added yet another level of tagging in our “Tag on Tag system”: a macro textual tag <ObjectLevel>, which should nest the three textual tags correspondent to each of the asset categories of our database – <ImmovableProperty>, <MovableProperty>, <IntangibleObject>. Each of these asset categories then hosts a general *textual* tag called <Description> – a solution found to accommodate all the differences and varieties of elements that compose the set of information regarding this specific level. The juxtaposition of the three tags hierarchically renders the following structure (fig. 15).

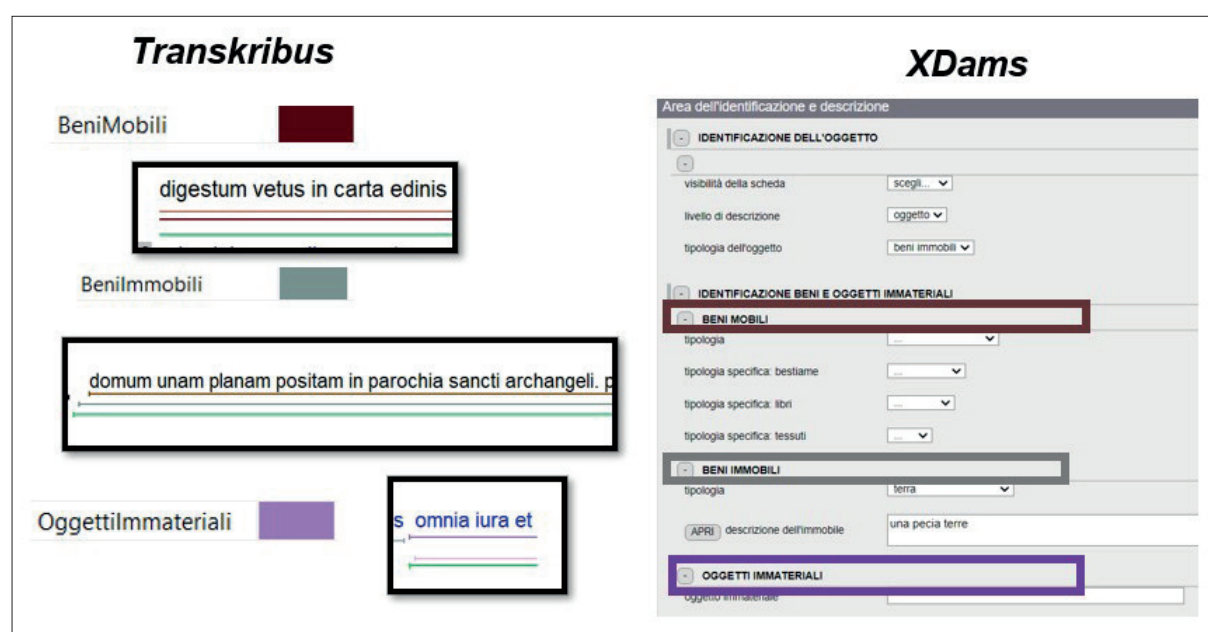


Figure 15. Constructing the level object – comparison between Transkribus and xDams categories.

Once this structure was completed, we then moved on to the first attempts of direct importation from Transkribus to xDams.

6. Importing Transkribus XML into xDams

To obtain a file compatible with the *Memorialis*' layout, it was necessary to carry out analysis and mapping work, which required iterative interventions to achieve an optimal result.

The process was divided into the following phases:

- analysis of Transkribus XML and mapping
- import procedures
- data validation.

The analysis of Transkribus XML and mapping

The analysis was a preparatory and essential phase for the successful completion of the subsequent steps. The XML file generated by Transkribus, which employs a schema provided by the software, has the following structure:

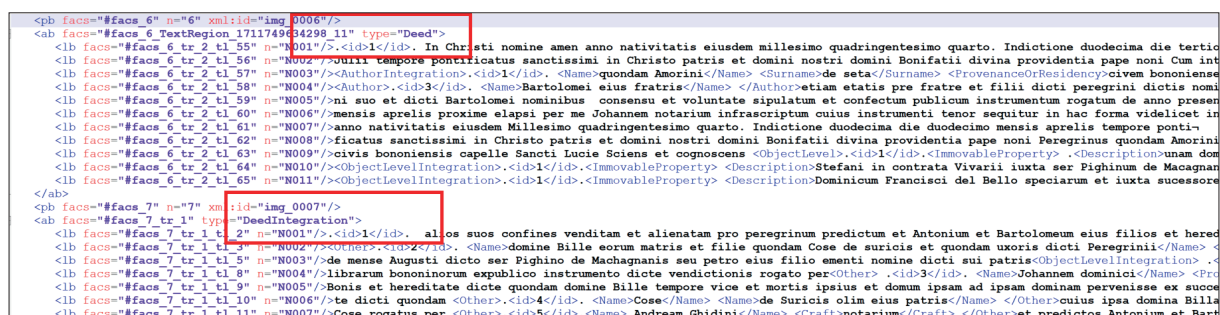
- `/teiHeader/fileDesc/titleStmt/title/text()`: xPath where the registry information is stored;
- `/text/body/div/pb`: each `<pb>` element represents a page contained within a registry;
- `/text/body/div/ab[xx]/@type`: each `<ab>` element represents an act, contained within a page;
- `/text/body/div/ab[xx]/lb[xx]`: each `<lb>` element represents a row, contained within an act.

Given the configurability of Transkribus, tagging was carried out, as far as possible, based on the scheme already used by the *Memoriali* project, as described in the previous section.

During the analysis phase, several critical issues emerged:

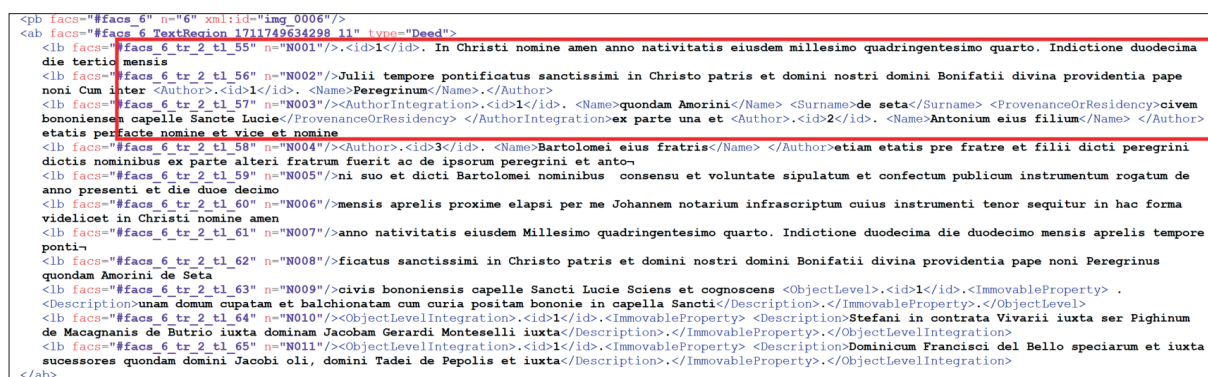
1. Transkribus creates a `</TEI/text/body/div/pb>` tag for each physical page it encounters, and it closes this tag before moving on to the next page. This tag also contains the `"@xml:id"` attribute, which specifies the name of the corresponding image. Simultaneously, one or more `</TEI/text/body/div/ab[xx]>` tags are created, identifying a "deed", which are still closed when moving to a new physical page. Thus, two situations may arise: either pages that contain multiple documents, represented by different `</TEI/text/body/div/ab[xx]>` tags opened and closed before proceeding to the next `</TEI/text/body/div/pb>` tag, or a "deed" that spans two or more pages, exceeding the limit of the physical page. The issue lies in the fragmentation of a single document into multiple parts, thereby losing its unity.
2. Transkribus renders the two nodes `</TEI/text/body/div/pb>` and `</TEI/text/body/div/ab>` as siblings and parallel to each other, without any hierarchical relationship: this causes issues when nesting the physical page information during procedural processing.
3. Each deed contains multiple `<lb>` lines, and each line can include several elements that map the information contained in the deeds, such as people and objects involved in the transactions. It is possible that the description of an element does not end at the physical end of the line, so the data is stored on the next line, creating a new element that would overwrite the previous one during procedural processing if not handled correctly. The issue here is the fragmentation of the information across multiple tags of the same type, leading to transcription errors.

As mentioned in the previous section, it was, therefore, necessary to adjust tagging in Transkribus by introducing an `<integration>` tag, which allows documents spanning different pages and elements whose descriptions exceed the physical line to be linked. This linkage is possible by the use of an `<id>[n]</id>` tag, placed both in the main element and in the supplementary one (fig. 16 and 17).



```
<pb facs="#facs 6" n="6" xlmid="img_0006"/>
<ab facs="#facs 6 TextRegion 1711749634298 11" type="Deed">
  <lb facs="#facs 6 tr 2 tl 55" n="N001"/><id1</id>. In Christi nomine amen anno nativitat...
  <lb facs="#facs 6 tr 2 tl 56" n="N002"/>Julii tempore pontificatus sanctissimi in Christo patris et domini nostri domini Bonifatii divina providentia pape noni Cum inter
  <lb facs="#facs 6 tr 2 tl 57" n="N003"/><AuthorIntegration>.<id1</id>. <Name>quondam Amorini</Name> <Surname>de seta</Surname> <ProvenanceOrResidency>civem bononiense
  <lb facs="#facs 6 tr 2 tl 58" n="N004"/><Author>.<id3</id>. <Name>Bartolomei eius fratris</Name> </Author>etiam etatis pre fratre et filii dicti peregrini dictis nomi
  <lb facs="#facs 6 tr 2 tl 59" n="N005"/>ni suo et dicti Bartolomei nominibus consensu et voluntate stipulatum et confectum publicum instrumentum rogatum de anno presen
  <lb facs="#facs 6 tr 2 tl 60" n="N006"/>mensis aprilis proxime elapsi per me Johannem notarium infrascriptum cuius instrumenti tenor sequitur in hac forma videlicet in
  <lb facs="#facs 6 tr 2 tl 61" n="N007"/>anno nativitat... Millesimo quadringentesimo quarto. Indictione duodecima die duodecimo mensis aprilis tempore ponti-
  <lb facs="#facs 6 tr 2 tl 62" n="N008"/>ficatus sanctissimi in Christo patris et domini nostri domini Bonifatii divina providentia pape noni Peregrinus quondam Amorini
  <lb facs="#facs 6 tr 2 tl 63" n="N009"/>civis bononiensis capelle Sancti Lucie Sciens et cognoscens <ObjectLevel>.<id1</id>. <ImmovableProperty> .<Description>unam dom
  <lb facs="#facs 6 tr 2 tl 64" n="N010"/><ObjectLevelIntegration>.<id1</id>. <ImmovableProperty> <Description>Stefani in contrata Vivarii iuxta ser Pighinum de Macagnan
  <lb facs="#facs 6 tr 2 tl 65" n="N011"/><ObjectLevelIntegration>.<id1</id>. <ImmovableProperty> <Description>Dominicum Francisci del Bello speciarum et iuxta suocore
</ab>
<pb facs="#facs 7" n="7" xlmid="img_0007"/>
<ab facs="#facs 7 tr 1" type="DeedIntegration">
  <lb facs="#facs 7 tr 1 tl 2" n="N001"/><id1</id>. alios suos confines venditam et alienatam pro peregrinum predictum et Antonium et Bartolomeum eius filios et hered
  <lb facs="#facs 7 tr 1 tl 3" n="N002"/><Other>.<id2</id>. <Name>domine Bille eorum matris et filie quondam Cose de suricis et quondam uxoris dicti Peregrini</Name> <
  <lb facs="#facs 7 tr 1 tl 5" n="N003"/>de mense Augusti dicto ser Pighino de Machagnanis seu petro eius filio ementi nomine dicti sui patris<ObjectLevelIntegration>.<id1</id>
  <lb facs="#facs 7 tr 1 tl 8" n="N004"/>librarum bononorum expulso instrumento diete venditionis rogato per<Other>.<id3</id>. <Name>Johannem dominici</Name> <Pri
  <lb facs="#facs 7 tr 1 tl 9" n="N005"/>Bonis et hereditate diete quondam domine Bille tempore vice et mortis ipsius et domum ipsam ad ipsam dominam pervenisse ex suocore
  <lb facs="#facs 7 tr 1 tl 10" n="N006"/>te dicti quondam<Other>.<id4</id>. <Name>Cose</Name> <Name>de Suricis olim eius patris</Name> </Other>cuius ipsa domina Billa
  <lb facs="#facs 7 tr 1 tl 11" n="N007"/>Cose rogatus per<Other>.<id5</id>. <Name>Andream Ghidini</Name> <Craft>notarium</Craft> </Other>et predictos Antonium et Bart
```

Figure 16. Linkage of Deeds spanning different pages in XML.



```
<pb facs="#facs 6" n="6" xlmid="img_0006"/>
<ab facs="#facs 6 TextRegion 1711749634298 11" type="Deed">
  <lb facs="#facs 6 tr 2 tl 55" n="N001"/><id1</id>. In Christi nomine amen anno nativitat...
  <lb facs="#facs 6 tr 2 tl 56" n="N002"/>Julii tempore pontificatus sanctissimi in Christo patris et domini nostri domini Bonifatii divina providentia pape
  <lb facs="#facs 6 tr 2 tl 57" n="N003"/><AuthorIntegration>.<id1</id>. <Name>quondam Amorini</Name> <Surname>de seta</Surname> <ProvenanceOrResidency>civem bononiense
  <lb facs="#facs 6 tr 2 tl 58" n="N004"/><Author>.<id3</id>. <Name>Bartolomei eius fratris</Name> </Author>etiam etatis pre fratre et filii dicti peregrini
  <lb facs="#facs 6 tr 2 tl 59" n="N005"/>ni suo et dicti Bartolomei nominibus consensu et voluntate stipulatum et confectum publicum instrumentum rogatum de
  <lb facs="#facs 6 tr 2 tl 60" n="N006"/>mensis aprilis proxime elapsi per me Johannem notarium infrascriptum cuius instrumenti tenor sequitur in hac forma
  <lb facs="#facs 6 tr 2 tl 61" n="N007"/>anno nativitat... Millesimo quadringentesimo quarto. Indictione duodecima die duodecimo mensis aprilis tempore
  <lb facs="#facs 6 tr 2 tl 62" n="N008"/>ficatus sanctissimi in Christo patris et domini nostri domini Bonifatii divina providentia pape noni Peregrinus
  <lb facs="#facs 6 tr 2 tl 63" n="N009"/>civis bononiensis capelle Sancti Lucie Sciens et cognoscens <ObjectLevel>.<id1</id>. <ImmovableProperty> .
  <Description>unam domum cupatam et balchionatam cum curia positam bononie in capella Sancti</Description>.</ImmovableProperty>.</ObjectLevel>
  <lb facs="#facs 6 tr 2 tl 64" n="N010"/><ObjectLevelIntegration>.<id1</id>. <ImmovableProperty> <Description>Stefani in contrata Vivarii iuxta ser Pighinum
  <lb facs="#facs 6 tr 2 tl 65" n="N011"/><ObjectLevelIntegration>.<id1</id>. <ImmovableProperty> <Description>Dominicum Francisci del Bello speciarum et iuxta
  <lb facs="#facs 6 tr 2 tl 66" n="N012"/>suocore quondam domini Jacobi olii, domini Tadei de Pepolis et iuxta<Description>.</ImmovableProperty>.</ObjectLevelIntegration>
</ab>
```

Figure 17. The linkage with `<id>` element.

This aspect deserves greater attention, as it is closely related to the transformation process, as seen below.

Technologies

For the analysis, processing, and importing of the data produced by Transkribus, we chose to use the Java programming language in conjunction with the XMLBuilder library, a parser developed by Regesta.exe. This library efficiently handles CRUD (create, read, update, delete) operations on XML-formatted resources as well as the migration of data into the designated database. As the final repository, we used xDams, an Open Source document management platform developed in ASP (Application Service Provider) mode, which uses the Web for data access and interaction. It is designed for the management and integrated use of multimedia historical archives.

Phases of the procedure

The first step necessary for the procedure was to process the XML version generated by Transkribus, as described previously, in order to create a new version with some modifications, particularly related to pagination and the structure of “deeds”.

The hierarchical “sibling” relationship between the <pb> and <ab> elements is not an optimal structure for processing, which required intervention by first removing the </TEI/text/body/div/pb> nodes from their original position and reinserting them, hierarchically nested, within the corresponding </TEI/text/body/div/ab[xx]> tags as their “children”. This way, by iterating only through the </TEI/text/body/div/ab[xx]> nodes, the information related to the reference pages is directly accessed.

Next, we addressed the management of the multi-line and multi-page issues: in both cases, Transkribus generated multiple identical nodes where it stores information related to lines affected by line breaks or page changes. This behaviour is also not ideal for procedural processing. We resolved it by introducing two interrelated concepts: the first is “integration”, a system we use to identify nodes referring to the same resource (block of lines, block of pages, etc.); the second is the <id> tag, which is used as a key to recognize and link those nodes. Here is a practical example to better clarify the concept:

```
<lb facs="#facs _ 7 _ tr _ 1 _ tl _ 21" n="N014"/>
<ObjectLevel>
  <id>2</id>
  <ImmovableProperty>
    <Description>una petie terre aratorie</Description>
  </ImmovableProperty>
</ObjectLevel>
<lb facs="#facs _ 7 _ tr _ 1 _ tl _ 22" n="N015"/>
<ObjectLevelIntegration>
  <id>2</id>
  <ImmovableProperty>
    <Description>vidate et arborate sex tornaturarum et cum
eo pluri vel minori quod     esse reperitur posita in guardia terre
funi comitatus bononie in loco gualfundo</Description>
  </ImmovableProperty>
</ObjectLevelIntegration>
```

In this example, we can see how the data contained in the <ObjectLevel/ImmovableProperty/Description/text()> node begins on a specific line, ‘N014,’ but continues on the following one, marked by the <lb> tag with n=‘N015’. To retrieve the values contained in both “Description” nodes, we created the <ObjectLevelIntegration> tag, which acts as an integration of <ObjectLevel>, and linked them through <id> tags, both set to “2” in this case. By doing so, during procedural processing, it becomes easier to loop through the various nodes, retrieve the text in the <Description> nodes, and concatenate them into the string “una petie terre aratorie

vidate et arborate sex tornaturarum et (...)”, which will then be written into the designated *XPath* for that value in the target database schema. The solution just described, regarding the use of ‘integration’ and `<id>`, is applied and repeated across all levels (`<Deed>`, `<MovableProperty>`, `<ImmovableProperty>`, and `<IntangibleObject>`), and involves all the XML nodes, including repeatable ones (`<Author>`, `<Recipient>`, `<Conductor>`, `<Neighboring>`, `<Creditor>`, `<Curator>`, `<Debtor>`, `<Heir>`, `<Executor>`, `<Guarantor>`, `<Legatee>`, `<Present>`, `<Procurator>`, `<TestamentarySubstitute>`, `<Witness>`, `<PreviousOwner>`, `<Tutor>`, `<Other>`, `<ObjectLevel>`) and those that occur only once per “Deed” (`<Type>`, `<TopicalDate>`, `<RecordingNotary>`, and `<ProvenanceOrResidency>`). In this way, we ensure that all portions of resources “split” by Transkribus can be uniquely linked and reconstructed.

This solution, however, introduces another critical issue: Transkribus, and consequently the integrations we added, generate *textual* tags based on the physical order of the manuscript lines, including line breaks. This means that nodes may contain entire words or sentences, but also fragmented sentences or, worse, truncated words, as a result of line breaks. Managing and correctly resolving all possible concatenations automatically is impossible, and human intervention is required to correct eventual introduced errors.

We considered two possible approaches: the first involved processing each retrieved value with a method (`.trim()`) that removes any leading and trailing spaces from each string, then adding a space before each concatenation. The second approach involved inserting a distinctive symbol (in this case, a double pipe, ‘||’), which is easily recognizable and certainly foreign to the rest of the text, before each concatenated string. Although the first approach had a success rate of approximately 70-80%, it made it very difficult to address the remaining 20-30%. For example, consider the word “habitante”, which is interrupted by a line break in the manuscript and concatenated as “habitant”. The second approach, while it “creates errors” in 100% of cases by introducing symbols not present in the original text (e.g., producing ‘hallbitante’ in the previous example), makes it much faster to locate and manually correct these issues. This may involve simply deleting the chosen symbol to join the two parts of the string, as in the case of “habitante”, or replacing it with a space when necessary. Since post-procedural human intervention was required in any case, we mutually agreed to adopt the second approach, aiming for lower complexity and increased speed for those responsible for this final data cleanup in the database.

Once these preparatory steps are completed, the remainder of the procedure—namely, the actual extraction of data and its writing into the database—proceeds smoothly, consistently following the same methodology. The process works in levels, starting from the highest and moving downward. Initially, all records corresponding to the `</TEI/teiHeader>` nodes, labelled according to the value found in `</TEI/teiHeader[xx]/fileDesc/titleStmt/title/text()>`, are recognized as “registry” type records and saved in the xDams database under the root element of the hierarchical tree.

The next step is processing “deed” type records, identified by the `</TEI/text/body/div/ab[xx]>` nodes. For each of these, the values contained in the various repeatable and non-repeatable nodes are extracted, reconstructing strings when necessary (in the presence of integrations), and then migrated into their respective *XPath* in the historical schema. Finally, the “deed” records obtained are saved in the database, each hierarchically linked to its corresponding “registry” record.

For all “deed” records, the value contained in the `</TEI/text/body/div/ab[xx]/@facs>` attribute is also retrieved and saved, as it will later serve as a key to identify them unequivocally. After processing deeds, the procedure examines the integrations for these same deeds, repeating the steps described above and adding the corresponding information to the already saved deed records, which are identified using a query based on an exact match between the `<id>` tag values. The final step is the processing of “goods/object” type records, which are of a lower hierarchical level and identified by the tag `</TEI/text/body/div/ab[xx]/ObjectLevel>`. Here, too, the data is extracted in the same way, with string concatenation in cases of integration. These records are further differentiated based on whether they are `<MovableProperty>`, `<ImmovableProperty>`, or `<IntangibleObject>`, and saved accordingly as “children” of their respective document, identified using a query on the database via the previously saved ‘@facs’ attribute. At the end of the procedure, we will have a database in xDams with the following structure (fig. 18):

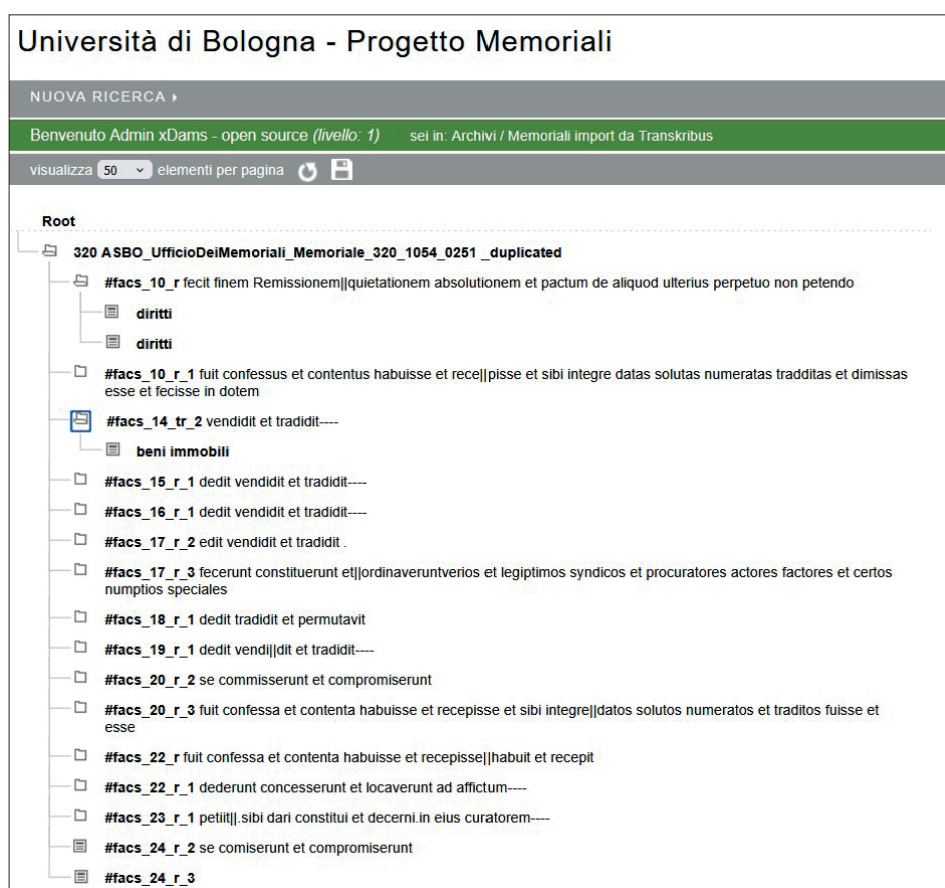


Figure 18. The hierarchical structure in xDams database.

The final step: data validation

At this stage, the results of the procedure are analysed directly within the xDams environment: an initial random check is performed on critical cases (integration cases) to first verify the successful grouping of the ‘integration’ fields.

If issues are identified during the check, the problem must be located, and the mapping and/or transformation procedure corrected.

If the check yields positive results, the data can be handed over to cataloguers, who will complete the descriptions (fig. 19).

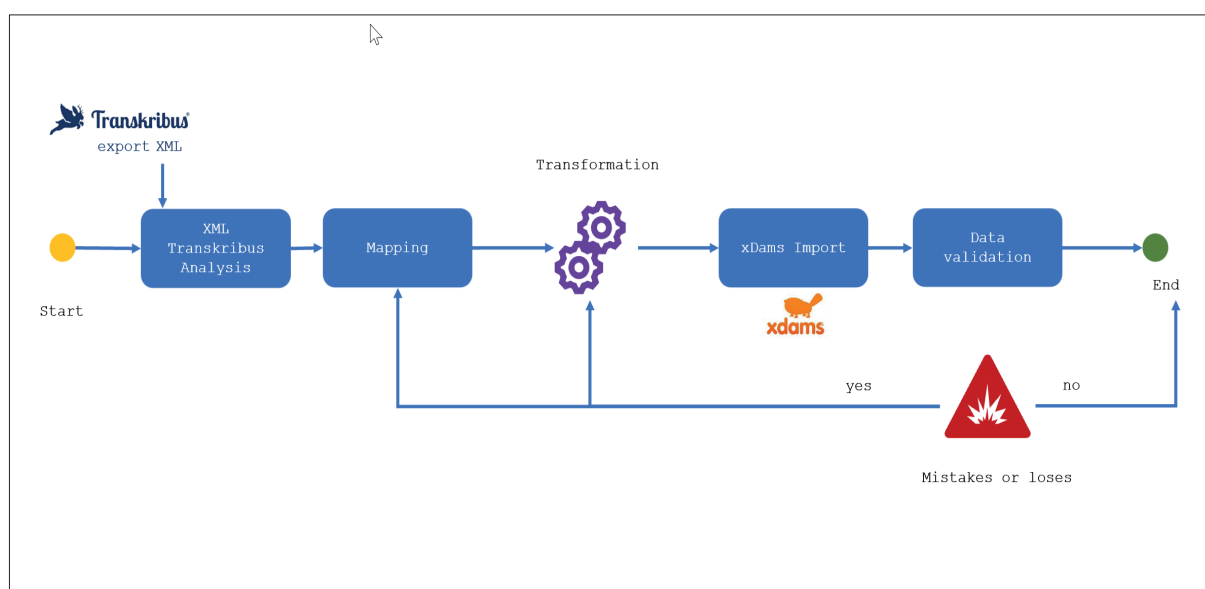


Figure 19. The complete process.

The described process is based on an iterative development cycle: the approach adopted is an iterative and incremental one, with short work cycles and frequent adaptations. The final result has allowed us to obtain a stable schema for reuse and we created a semi-automated loading pipeline.

7. The complete system and final remarks

After this meticulous experimentation, composed, as seen above, by many failed attempts, discussions and retrials, we established a general tagging system capable of producing XML files semiautomatically absorbed by our xDams database with the following structure: 2 *Structural* tags (<Deed> and <DeedIntegration>), 26 *textual* tags relative to the level “Act” (<Author>, <Other>, <Conductor>, <Surname>, <Neighboring>, <Creditor>, <Curator>, <TopicalDate>, <Debtor>, <Recipient>, <Affiliation>, <Heir>, <Executor>, <Guarantor>, <Legatee>, <Occupation>, <RecordingNotary>, <Name>, <Present>, <ProvenanceOrResidency>, <Procurator>, <TestamentarySubstitute>, <Witness>, <Type>, <PreviousOwner>, <Tutor> and <Qualification>) and their 26 integration (<Integration>) versions, plus 5 *textual* tags concerning the level “Object” (<ObjectLevel>, <MovableProperty>, <ImmovableProperty>, <IntangibleObject>, <Description>) and their 5 integration (<Integration>) versions, including, at last, a *textual* tag <id>, useful, as seen before, both to the level “Act” and level “Object”.

The following image gives a visual idea of the application of all these 65 tags in their appropriate cases on Transkribus (fig. 20):

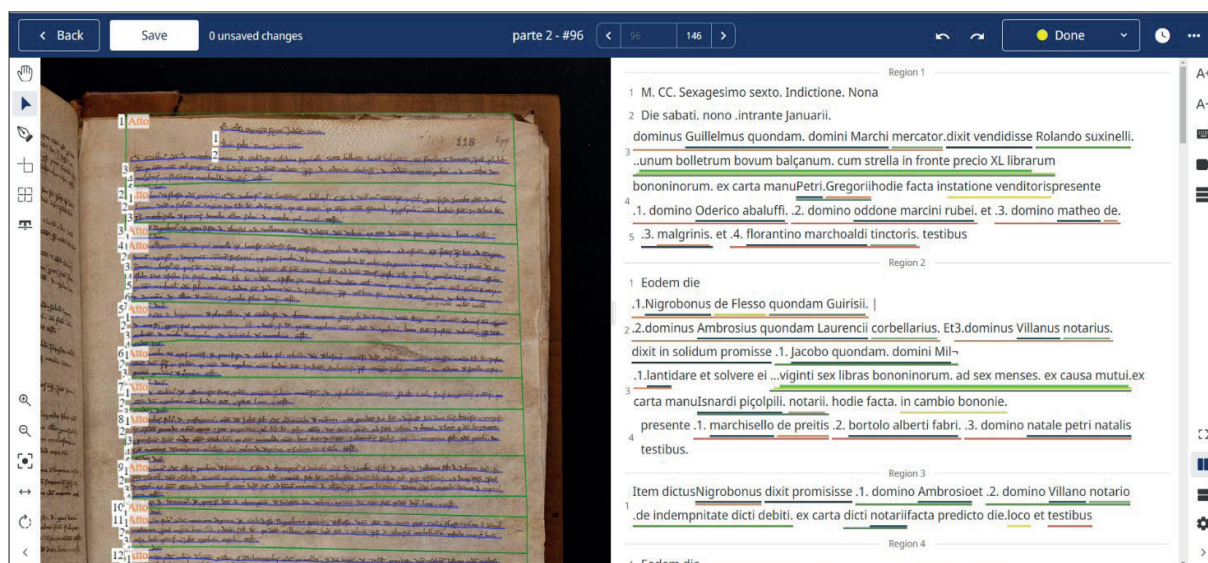


Figure 20. The complete tag system in Transkribus.

As mentioned in the introduction, the goal of this tagging system was not only to test the technological limitations and possibilities of both Transkribus and our xDams database, but also to find solutions that could expedite the general aim of the project to populate the database with complex information extracted from the *Memoriali*. With this objective in mind, all operations described in this paper were meticulously timed to see if and how much faster this system could be in comparison to a human cataloguer that manually inserted all this information in the xDams database. The outcome of this operation is that a regular cataloguer takes approximately 10 minutes to insert manually a complete deed in the xDams database, considering as a parameter the first unit of the *Memoriali* series – the *Memoriali* of Nascimpace Petrizani – marked by the presence of very simple descriptions of all elements relative to the “object” level. Tagging a deed from the same *Memoriali* unit inside Transkribus with all the categories corresponded to the database takes, instead, approximately 3 minutes, but we should also add to this timeframe the time needed to import the resulting XML Transkribus file into the xDams database. To generate a Transkribus XML file that can be acquired to initiate the transformation procedures, the server usually takes up to 1 minute, after which the Regesta.exe can produce a MemoBo XML file that is xDams-compliant and imported into the system. We thus conduct a testing phase, and if the result is satisfactory, we move the output to the working database. The procedure takes approximately 30–40 seconds to process 30 records – equivalent to 1 second per record – including the preliminary stages of correction and hierarchy adjustment. Adding these numbers up, including the time it takes to share a tagged Transkribus XML file among the different members of the project, we could estimate that through this semi-automatic importation system, it would take us approximately 5 minutes to include a deed inside our xDams database. Contrary to the manually inserted records, this deed would require further corrections to bring it to the standards of the database, such as altering names of the deed’s agents to Latin nominative; an operation that we estimate could take up to 2 minutes. The outcome is that in approximately 7 minutes the deed initiated by tagging portions of

a Transkribus HTR generated transcriptions of *Memoriali* has the same standards as the one included by manual cataloguers in as much as 10 minutes. The gain is of nearly 30% of time, which, considering the dimensions of the collection, allow us to consider this strategy, and, consequently, the fruits of our Transkribus/xDams tagging experiment as a viable future for the development of the project.

Apart from the consistent gain in terms of time, it is important to highlight that the adoption of this tagging procedure as a cataloguing method also includes an AI-generated transcription of the material as a final product, with a high level of intervention by human hands applying the tags to the documentation. Though it is not the scope of the MemoBo project to offer full transcriptions of these sources, the resulting texts may be useful for teaching and research, opening new possibilities and applications. There is much that could be said about the cultural implications of rendering available large quantities of information concerning late medieval notarial records; a source, which contrary to normative texts, grant us access to a series of agents and their margins of actions, for example women, not formally included in the channels of political and institutional participation (Piccini, 1996). By applying artificial intelligence tools, thinking on their limitations and possibilities for this specific sort of serial documentation, our project aims to contribute to providing materials to scholars interested on tackling such topics, which could transform our understanding of the dynamics in large communes of the late Middle Ages, such as late medieval Bologna.

References

- Bartoli Langeli, Attilio. 2006. *Notai: scrivere documenti nell'Italia medievale*. Roma: Viella.
- Bruno, Giovanni, Francesca Delneri, Tommaso Duranti, and Fabiana Guernaccini. 2023. "Il Progetto MemoBo: sinergie e nuove sfide a partire dai Memoriali bolognesi." *DigItalia* 18: 150-63.
- Continelli, Luisa. 1988. *L'archivio dell'Ufficio dei memoriali. Inventario. Volume I Memoriali, 1264-1463. Tomo I, 1265-1333*. Bologna: presso l'Istituto per la storia dell'Università.
- Hodel, Tobias. 2022. "Supervised and Unsupervised: Approaches to Machine Learning for Textual Entities." In *Archives, Access and Artificial Intelligence: Working with Born-digital and Digitized Archival Collections*, edited by Lise Jaillant, 157-78. Bielefeld: Bielefeld University Press.
- Macchiavello, Sandra, and Antonella, Rovere. 2018. "The Written Sources." In *A Companion to Medieval Genoa*, edited by Carrie E. Benes, 27-48. Leiden-Boston: Brill.
- Piccinni, Gabriela. 1996. "Le donne nella vita economica, sociale e politica dell'Italia medievale." In *Il lavoro delle donne*, a cura di Angela Groppi, 5-46. Roma-Bari: Laterza.
- Tamba, Giorgio. 1998. *Una corporazione per il potere: il notariato a Bologna in età comunale*. Bologna: Clueb.
- Tamba, Giorgio, ed. 2002. *Rolandino e l'ars notaria da Bologna all'Europa: atti del Convegno Internazionale di Studi Storici sulla Figura e l'Opera di Rolandino. 9 - 10 ottobre 2000*. Milano: Giuffrè editore.

Learning the Renaissance cosmos: The circulation of Sacrobosco's *De sphaera mundi* in sixteenth-century Italy*

Andrea Ottone^(a)
Erika Squassina^(b)

a) University of Oslo, <https://orcid.org/0000-0003-0138-5793>

b) EMoBookTrade project, <https://orcid.org/0000-0002-7966-8737>

Contact: Andrea Ottone, andrea.ottone@iakh.uio.no; Erika Squassina, squassina.erika@gmail.com

Received: 31 December 2024; **Accepted:** 10 March 2025; **First Published:** 15 May 2025

ABSTRACT

This article investigates the late Renaissance dissemination of Joannes de Sacrobosco's *Tractatus de sphaera mundi* in the Italian book market. The *Tractatus* was a widely circulated textbook of geocentric astronomy that served as the foundation of astronomical knowledge and applied mathematics from the thirteenth to the mid-seventeenth centuries. Its influence remained strong throughout the establishment of Copernicanism, and gradually faded as Newtonianism provided the permanent paradigm for the Enlightenment science soon to come. Using quantitative methods, digital tools, and social network analysis, this article traces a social geography of the *Sphaera*'s owners and aims to profile the a cultural morphology of the private collections of graduate and post-graduate individuals at the end of the sixteenth century.

KEYWORDS

Sacrobosco; Cosmology; Book trade; Book market; Church history.

* The present paper takes into account and was largely inspired by the work brought forward by *The Sphere: Knowledge System Evolution and the Shared Scientific Identity in Europe* led by Professor Doctor Matteo Valleriani at Max-Planck-Institut für Wissenschaftsgeschichte, Berlin. Andrea Ottone collaborated on this project and learned much about the topic during his visiting fellowship at the MPIWG (2020-2024). This work is supported by ERC grant (BE4COPY, 101042034). Funded by the European Union. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Although the two authors have discussed the whole paper and jointly designed the database and assembled the dataset, Andrea Ottone is the author of sections 1-2 and 6; Erika Squassina is the author of sections 3-5. Both authors are grateful to Diane Booton for her help in revising the text.

© 2025, The Author(s). This is an open access article, free of all copyright, that anyone can freely read, download, copy, distribute, print, search, or link to the full texts or use them for any other lawful purpose. This article is made available under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. JLIS.it is a journal of the SAGAS Department, University of Florence, Italy, published by EUM, Edizioni Università di Macerata, Italy, and FUP, Firenze University Press, Italy.

Introduction

The Renaissance was an exciting time for astronomy, a time characterized by an experimental attitude and a daring drive. Scholarship of the time had recently been shaken by Copernican heliocentrism and various fields of study, especially philosophy and theology, were attempting to adjust to the ongoing paradigm shift. So much innovative attitude from Renaissance scholarship tends to overshadow how scholars of the time, Copernicus himself, were born and raised in a geocentric cosmos and that a finite space of concentric spheres revolving around the Earth was what they studied in school. Most of these studies were done by reading the *Tractatus de sphaera mundi*, an introductory manual on the use of the armillary sphere, which was necessary to locate the stars, browse the cosmos, and navigate the Earth. Throughout the Renaissance, the *Tractatus* became a school textbook instrumental in acquiring the rudiments of applied mathematics and geometry and constituted the minimum requirement to pursue the quadrivium study program. The book originated in late medieval France and was written by the English Augustinian monk and mathematician John of Holywood (<fl. 1230), Latinized as Johannes de Sacrobosco, a professor of mathematics at the University of Paris. The *Tractatus* soon gained momentum, earning popularity in the manuscript era, maintaining its leadership throughout the incunabula era, and flourishing in the Cinquecento, just as the geocentric paradigm came under attack. Its publishing trajectory began to decline in the Baroque era, to disappear from the scientific horizon by the mid-Seicento, when heliocentrism became dominant in the general scope of the Republic of Letters and geocentrism began to lose its grip also on the curricula of European universities.

As a quasi-standard requirement of entry-level university students, the *Sphaera mundi* transitioned from being a single-authored text to, especially in the Cinquecento, being part of a literary tradition and a corpus of texts, an intellectual arena in which many scholars engaged to contribute with editorial nuances, novel commentaries, numerical tables, expanded visual aids (including bidimensional volvelles and tridimensional build-it-yourself diagrams), *proemia*, dedications, sonnets, analytical indexes and so on.¹

Most of these new and old addenda were shuffled around in Latin or vernacular renditions to propose new editorial formulas to the reading audience with the objective of winning the fierce competition of the first and secondhand markets across the continent and beyond (Valleriani and Ottone 2022; Ulla Lorenzo 2022).

1. Sources

Much of the information about the literary success of the *Sphaera* is usually derived from the publishing history that Sacrobosco's work underwent in its many revisions, editions, and reprints mainly during the printing era. The present paper, instead, aims to provide an overview of the distribution of the editions, based on the point of view of the recipients, collectors, and, presumably, consumers of the text, hence providing a bottom-up view of how Sacrobosco's textbook earned and maintained a solid tradition in academic syllabi and a respectable sales record during the long Renaissance and beyond.

¹ On the volvelle in the *Sphaera* corpus, see Citron (2025).

Limited to the Italian peninsula alone, admittedly a relevant portion of both the European book market and of the Republic of Letters, such a survey is possible for the late Cinquecento, with a retrospective view mainly of the second half of the century. The sources allowing such study are mainly, though not exclusively, limited to the clergy. The available information comes from the complex dynamics of the implementation of the third Roman Index of Forbidden Books in the Italian dioceses. The Index, promulgated by Clement VIII in 1596, generated a number of jurisdictional conflicts, some of which concerned the regular clergy, with special regard to the § 1 of the paragraph *De prohibitionem librorum*, which mandated bishops and inquisitors to apply the Index territorially, allowing them to pursue visits to relevant book collections, which included monasteries and convents within their dioceses. The jurisdictional conflict was particularly aggravated in the dioceses of the Viceroyalty of Naples, where some monasteries and convents maintained feudal control over several surrounding settlements. These were the first to request a dispensation from the Congregation of the Index in order to avoid pastoral visits by nearby bishops in such *nullius diocesis*, which could have set a dangerous precedent for future jurisdictional incursions into the territorial possessions of the regular orders.²

The situation escalated when the Congregation, eager to facilitate the local enforcement of the newly published Index, promptly granted the requested dispensations. This internal dispute among the regular clergy led almost all the orders present in the Italian peninsula to request a similar dispensation, not only for their secular properties, but also for their entire network of religious houses and dependent parish churches present in their subalpine religious provinces, thus requesting that the new Index be applied by the orders independently of the diocesan authorities. This was motivated not only by the concern for their jurisdictional prerogatives but also by the fear that the arbitrary censorship of bishops and inquisitors could have damaged their individual and collective patrimony of books. The Congregation of the Index, in order to avoid harmful conflicts between monasteries and diocesan authorities, with unpredictably long and asymmetrical delays in the pursuit of the new censorship campaign, agreed to grant the heads of each religious order the authority to carry out the implementation of the Clementine Index independently.

The only condition imposed by the Congregation of the Index was that the central authorities of each of the religious orders involved were required to send a copy of the resulting documentation to Rome. The Congregation's initial request was for a partial list of book holdings, identifying only those books that violated the rules of the new censorship catalog (this request was in keeping with the mandate of the new Index). The purpose of this was not only for the Dicastery to monitor the actual progress of the systematic removal of prohibited material from the libraries; and second, to obtain information that the Cardinal of the Index would later use to guide the gradual correction of the offending texts. All this was done in view of the imminent massive project of an expurgatory index that the Congregation had initiated in those years.

However, it soon became clear to the Congregation that the local authorities appointed by the regular orders (mainly the priors of the monasteries and convents involved in the survey) were unable to reliably apply the rules of the new Index while inspecting the individual and communal book

² *Nullius diocesis*, by definition, were territorial spaces carved out of the regular jurisdiction of bishops and assigned to priors endowed with bishop-like prerogatives on secular vassals. The whole set of events here accounted is detailed in Fragnito (2006).

collections. In a change of strategy, the Congregation instructed the individual monasteries to send to Rome groupings of book lists containing a complete inventory of the holdings of monks, friars, and eventually, secular individuals under the jurisdiction of the regular orders. The Vatican functionaries would then have taken care of compiling the comprehensive lists and indicating which were the offending copies that had to be temporarily confiscated by the internal authorities of the religious orders, pending textual expurgation (*donec expurgantur*). Since the Congregation of the Index postponed this internal survey in view of the forthcoming expurgatory index, the functionaries were interested not in the titles of the collections but in the editions themselves; thus, the complete lists of the individual and communal collections of monks, friars, vassals, and religious institutions had to indicate not only the owner and the location of each collection, but also the authors, title, provenance, publisher, year of publication, and format of each edition included.

During the three years that it took the Congregation of the Index to collect the supposedly complete documentation from the dioceses of the subalpine region, the religious orders responded fairly consistently to the Dicastery's requests. Objective limitations were represented by the different levels of literacy that characterized each group of religious in a vast and diverse territory such as the Italian peninsula. A lower level of literacy sometimes affected the quality of the lists drawn up locally. Another element of dispersion of information is the documented loss of part of the archival material of the Roman Congregation when part of its archive was transferred from Rome to Paris during the Napoleonic era.³ Nevertheless, the documentation preserved today in the *Codices Vaticani latini* 11266-11326 represents the most valuable source of information on the book holdings of the Italian regular clergy (with the sporadic addition of nuns, secular clergy, and secular individuals) for the late Renaissance.

This valuable corpus of sources has been the subject of a praiseworthy digitization effort since the early 2000s, as part of the RICCI research project (Ricerca sull'Indagine della Congregazione dell'Indice).⁴ The resulting open-access bibliographic database will be used to survey the distribution of Sacrobosco's *Sphaera mundi* on the Italian peninsula and to profile the cultural and social status of its collectors, along with their detectable common patterns of book acquisition.

2. Methodology

The process of data harvesting in the RICCI database was carried out to gather individual book collections that contained one or more editions of the *Sphaera*. In order to trace the distribution of copies of the book circulating in the peninsula at the end of the sixteenth century, the selection was limited to those editions that were classified as identified in the RICCI database (i.e.,

³ For a detailed account of the dispersion of the archive of the Sant'Uffizio and the Congregation of the Index, see Tedeschi (2003, 35-46).

⁴ RICCI: <https://ricci.vatlib.it/>. Henceforth, single entries derived from this database will be indicated either with ELE (followed by a number, refers to a single collection entered in the repository) or BIB (followed by a number refers to a single edition entered in the database). This documentation features an extensive bibliography tentatively summarized here: (Adoriso 1987; Ciccarelli 1990; 1991; Criscuolo 1986; Dahnk Baroffio 1986; 1990; 1992; 1993; 1994; 1995; 1996; Gargan 1975; Masetti Zannini 1970; Pagano 1986; Sbordone 2001; 2002; 2003; Barzazzi 1995; 2001; Compare 2000; 2002; 2003a; 2003b; Granata 2006; 2016a; 2016b; 2024; Laudadio 2005; Lebreton and Fiorani 1985; Rozzo 1998a; 1998b; Rusconi 2004a; 2004b; Rusconi and Borraccini 2006; Zanot 2003; Zardin 1992; 1999).

bibliographic records included in the *Codices Vaticani latini 11266-11326* that have a matching edition in common bibliographic resources: e.g., EDIT16, SBN, or similar catalogs). Including unidentified editions would have introduced into the sample a number of records with inaccurate or incomplete information about the place of print, publisher, and year of publication of the editions under consideration. The choice was primarily motivated by the fact, as mentioned above, that the *Sphaera* had a rather dynamic publishing history, and each edition has its own editorial tradition that is worth tracing with precision (Valleriani et al. 2022).

By selecting those individual lists found in the *Codices Vaticani latini 11266-11326* that account for at least one identified copy of the *Sphaera* in their holdings, a total of 149 editions from 1490 to 1594 were gathered.

In order to find common patterns in the book acquisitions of the collectors of the *Sphaera*, the decision was made to gather all the book lists found in the RICI database that contained one or more editions of Sacrobosco's work. However, a further selection criterion was applied to exclude all lists of common and institutional collections (e.g., conventual or monastic libraries). As Section 5 will show, a chief purpose of this paper is to lay out the intellectual morphology of the book collections gathered in the sample, along with the social and professional standing of their declared owners. Hence, the preference for clearly defined individual collections was deemed key to gathering a cohesive and coherent sample. In fact, collective collections were by definition gathered to respond to multiple uses of a given community of users rather than responding to the defined intention of individuals with a defined intellectual agenda in mind. Furthermore, the cultural interests of the purchasers were meant to be placed in the territorial context in which the collector operated (e.g., the distance from a major urban center or a relevant printing center). Institutional collections may have falsified this view in light of the fact that the capability of a convent or monastery to cater to their needs in terms of books acquisition was largely different from that of an individual member of the clergy, whether he or she was operating in a rich urban context or a resourceless peripheral village. A religious house could rely on a considerably greater wealth and a network of contacts and distribution channels of material commodities than a single individual. This would have not only spoiled the data projection found in Section 5, but it would have also affected the purpose of outlining the capacity of the single editions to navigate the Italian market from their place of origin (whether they were Italian cities or ultramontane centers) to the various locations in the peninsula.

The qualitative analysis of the book holdings of the collectors of the *Sphaera* will take into account the well-known *caveat* that ownership does not equate to readership, but it will do so with some moderation. First, it will consider the particular relationship that Renaissance book collectors had with their books, especially those outside the ranks of the nobility. Far from being an object of status or a tool to be used by others, the private collections of clerics and secular professionals (lawyers, physicians, and notaries) of the late sixteenth century were primarily intended to be repositories of operational knowledge. Previous studies of these collections have shown how closely Baroque book collections reflected the professional training and everyday practice of their owners (Ago 2002). With this in mind, the presence of a given book in a private collection will be regarded as a meaningful effort to serve an intellectual or operative purpose. Hence, this paper will not conceptually or semantically cross the threshold between ownership and readership. Rather, an intersectional space will be explored. This will also be done because

owning a given book was a cultural statement, especially in the pre-bourgeois and pre-industrial age of printing. Moreover, in post-Tridentine Catholic dioceses, under the dictated agenda of cultural reform encouraged by the Council, owning the seminal tools of pastoral practice and the rudiments of knowledge was considered a duty for those willing to respond actively to the remarks often repeated by Protestant propaganda regarding the crass ignorance of the Catholic clergy. Moreover, exponents of the high professions of the late Renaissance were engaged in a refined literary context where prescribed reading lists for active professionals, e.g., *De modo studendi in iure* or *De modo studendi seu legendi in medicina*, made the stacking up of a respectable book collection a common professional statement. The very existence of such a debate during this fervent period of intellectual engagement allows the building of an ideal bridge between, at the very least, the material possession of a book and the display of an individual cultural agenda. This liminal space is believed to be capable of revealing the intellectual orientations of individuals who, by owning a copy of an entry-level university book, demonstrated an interest in the academic curriculum of their time. The assumption that the ownership of the *Sphaera* isolates a middle- to high-ranking stratum of the reading elite of the clergy can at times be supported not only by the quantitative and qualitative structure of their book collections. In fact, in the sources at hand, i.e., the individual lists submitted to the Congregation of the Index between 1599 and 1601, the social and professional status or the ownership of an academic degree is occasionally declared by the owner at the incipit of the lists.

By applying the above-mentioned filters to the selection of book owners found in the *Codices Vaticani latini* 11266-11326, a total of 159 individual collections have been isolated. The size of each collection varies, but the total number of bibliographic entries gathered amounts to 16,550 records. This number of records, however, includes several internal and often recurring elements (e.g., author, title, place of printing, and publisher). The total number of unique records managed for this project is 63,474.

Such a large amount of data requires quantitative analysis. Moreover, the quality of the available data, which includes objects (books), persons (be they owners, authors, or publishers), places (whether printing centers or the actual locations where the collections were found), and collective institutions and communities of people (convents, monasteries, parish churches, or rural settlements, i.e., *nullius diocesis*), suggests that a profitable path should be that provided by social network analysis.

To follow this path, the data derived from the *Codices Vaticani latini* 11266-11326 via the RICI database were processed through a digital platform that could best manage this amount of data. The choice fell on Nodegoat, a digital tool developed by the Dutch company LAB1100.⁵ A diagram of the database created in Nodegoat to process the data is visible in Fig. 1.

⁵ The Nodegoat platform allows users to create their own database model to customize the potentials of the platform and adapt them to the research purpose of users. The model was developed by Andrea Ottone and implemented by both authors of this paper. Both authors take the opportunity to show their gratitude to Giacomo Stroiazzo for his technical aid management of the dataset and to Alberto Campillo Pardo for his advice for the enhancement of the visual aids proposed in this paper.

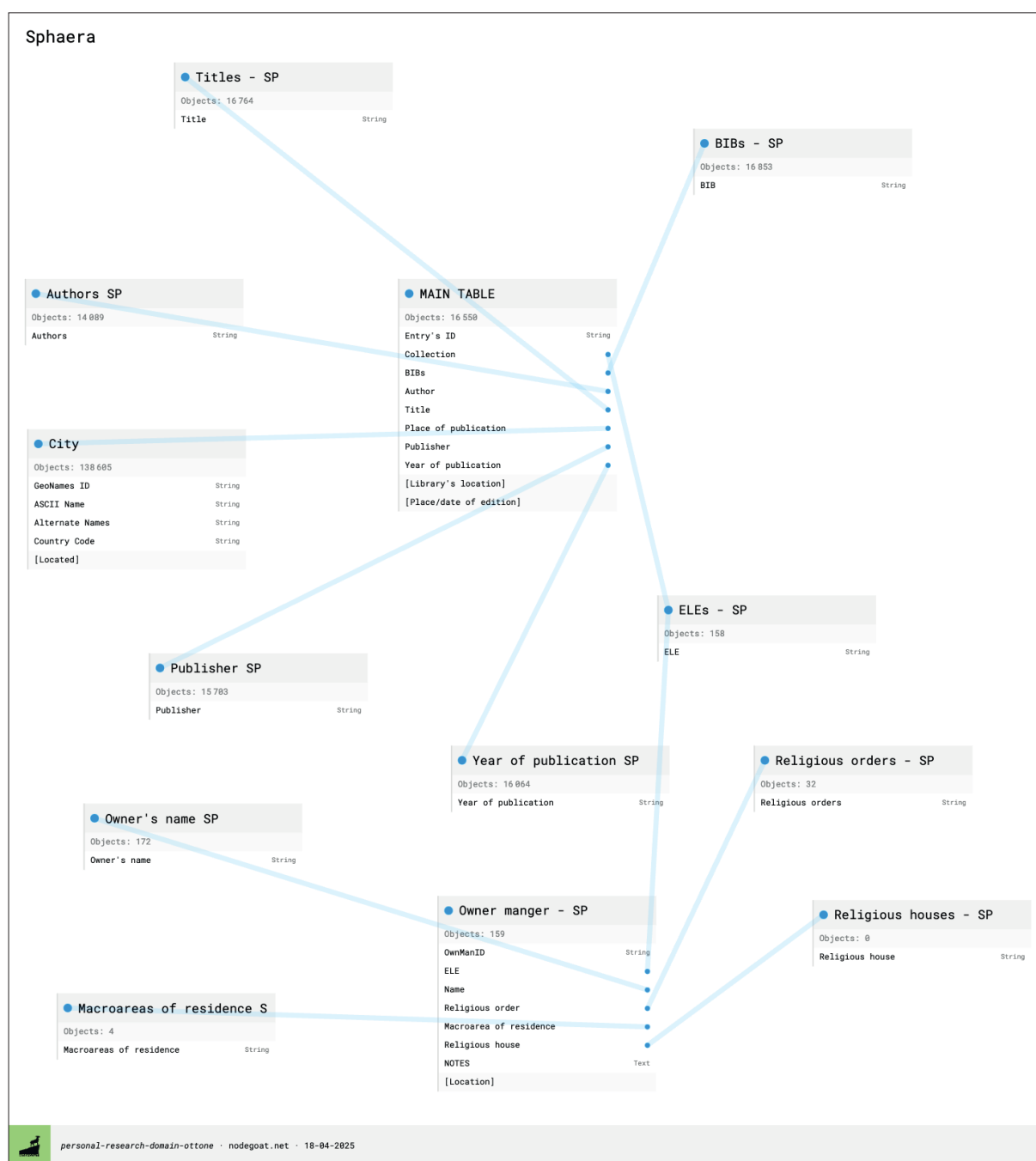


Figure 1. Logical schema of the relational database assembled for the study of the distribution of *Tractatus de sphaera* in the Italian peninsula in the late sixteenth century.

The 63,474 unique elements thus became as many nodes in a relational network with a total of 112,839 links (see Fig. 2).

Section 1 will discuss the geographic distribution of the editions of the *Sphaera* found in the corpus *Codices Vaticani latini 11266-11326* that are currently available in the RIC database. Each edition (RIC element: BIB) will be distributed on the map by the settlement, town, or city accounted in the source of origin. Moreover, it will take into account the origin of each edition of the *Sphaera*

found in the dataset (i.e., the place of publication). Section 2 will describe the cultural morphology of the book collections where the *Sphaera* has been found. For this purpose, the global network represented in Fig. 2 has been filtered to isolate only the collectors who clearly declared an academic standing or for whom an academic standing could be implied from external evidence. The data have been further filtered to isolate the authors of the books enlisted in the book collections selected. The authors have been used as an indicator of a cultural and intellectual orientation of the single book collections.

Ultimately, the present paper uses network analysis to highlight redundancies of bibliographic data to single out common patterns in the circulation of the *Sphaera mundi* in both its production and consumption modes.⁶

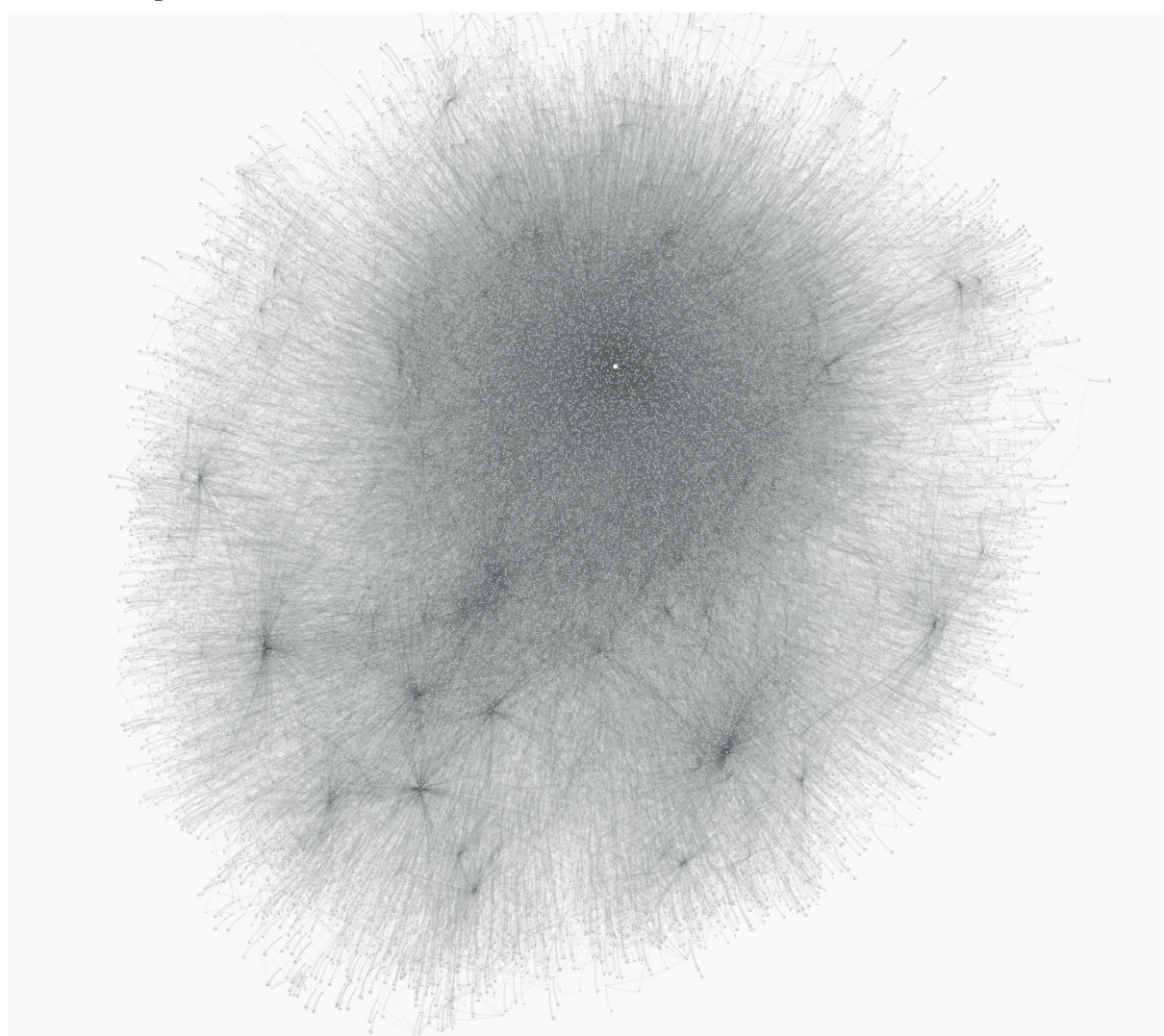


Figure 2. Global network generated by the book collections of the owners of Sacrobosco's *Sphaera* selected from the *Codices Vaticanus latini 11266-11326* via the RICCI database.

⁶ For a concise literature on the theory and methodology of social network analysis, see Ahnert et al. (2020); Brughmans (2016); Kerschbaumer (2020); Wetherell (1998).

3. The social geography of the *Sphaera*

The first observation to be drawn from the available data concerns the overall territorial distribution of copies of Joannes de Sacrobosco's *Tractatus de Sphaera*. As Fig. 3 shows, the ownership of one or more copies of this astronomical treatise was mainly, with a few exceptions, an urban phenomenon in late Renaissance Italy. Moreover, it was primarily a northern and central Italian phenomenon. This is where the density of known copies is greatest, and this is also where the *Tractatus* could be found in smaller settlements and rural areas. Several driving forces seem to be at play: the greater density of large urban centers, location of some relevant *studia* (e.g., Padua, Pavia, and Bologna) and the proximity to what is clearly the major engine for of the dissemination of the *Sphaera* in print in the subalpine market: Venice (see Fig. 4). Thus, it may be inferred that the *Sphaera mundi* followed secular market trajectories, even when viewed from the distinctive perspective of the Italian regular clergy. If the network of the regular orders could at times function as an internal distribution network for the members of individual religious orders, this does not appear to be the case for the *Sphaera*.⁷ Rather, it seems that the members of religious orders, primarily those more deeply integrated into secular society and engaged in pastoral duties and teaching, would find their way to serve their own interests in their book acquisitions, including the *Sphaera*.

The *Tractatus* had a large presence in northern Italy (mainly modern Lombardy and Veneto, with some surprising absences in modern Turin and Genoa) but rare instances in Tuscany; some copies can be found in Rome (many fewer than one would expect in such a relevant city for Renaissance culture), and its presence is episodic in southern Italy, then ruled by the Viceroyalty of Naples. The only exceptions are the cities of Naples and nearby Salerno and its surrounding area. This asymmetric distribution of *Sphaera* confirms how the work was driven by the two main forces of a consistent book market and the presence of a local university.

This pattern is strangely broken in the region of present-day Calabria and Sicily, where copies of the *Tractatus* can be found in rather small settlements. With regard to Sicily, and Messina in particular, the circulation of the *Sphaera* can be explained according to a logic that may be foreign to modern trajectories of the circulation of learned material but was more common in the early modern period: that is, the presence of a schoolmaster like Francesco Maurolico (1494-1575) on the island. Maurolico was not only an established mathematician of the Renaissance, in contact with intellectuals like the Jesuit Christophoro Clavio; he was also one of the many commentators on the *Sphaera*. His presence and relatively recent activity in Sicily were enough to generate interest in the *Tractatus* due to his intellectual legacy in the land, for instance, mingled with a hint of local cultural pride.

Overall, instead, the highlands of Abruzzo and most of the southern Adriatic peninsula do not seem to have been favorable places for Sacrobosco's work, thus confirming how the interest towards the *Tractatus* was linked more to the activities of busy commercial centers and well-established courts where bureaucrats needed to acquaint themselves with the culture of the quadrivium and beyond.

The map proposed in Fig. 3 must, however, be taken with a reasonable grain of salt, as it is far from exhaustive: the available data do not allow us to argue that there could not have been found a

⁷ Dynamics of internal transnational distributions can be spotted within the Carthusian order (Ottone 2013, 238).

copy of the *Tractatus* in the empty areas; similarly, in the areas where a few instances of its presence can be found, there were probably many more. The data available are mainly representative of the Italian clergy and are themselves unlikely to be exhaustive of this social sample alone (it is known, for example, that Dominicans and Jesuits did not participate in the assembling of the precious book lists that comprise the bulk of evidence employed here). With regard to seculars, who were also likely active collectors of the *Sphaera*, it must be pointed out that exactly those areas where the *Tractatus* seems to be most absent are also the areas where the sources in use account for most of the lists of secular individuals living in *nullius diocesis*. Regardless of its probable limits, this map is useful for showing a tendency.

Ultimately, if the presence of one or more copies of the *Sphaera* in an individual library can be taken as an indication of the owner's interest in academic writings (as this paper argues), then a very asymmetrical picture emerges between North and South regarding the Italian clergy's interest in quadrivium education.

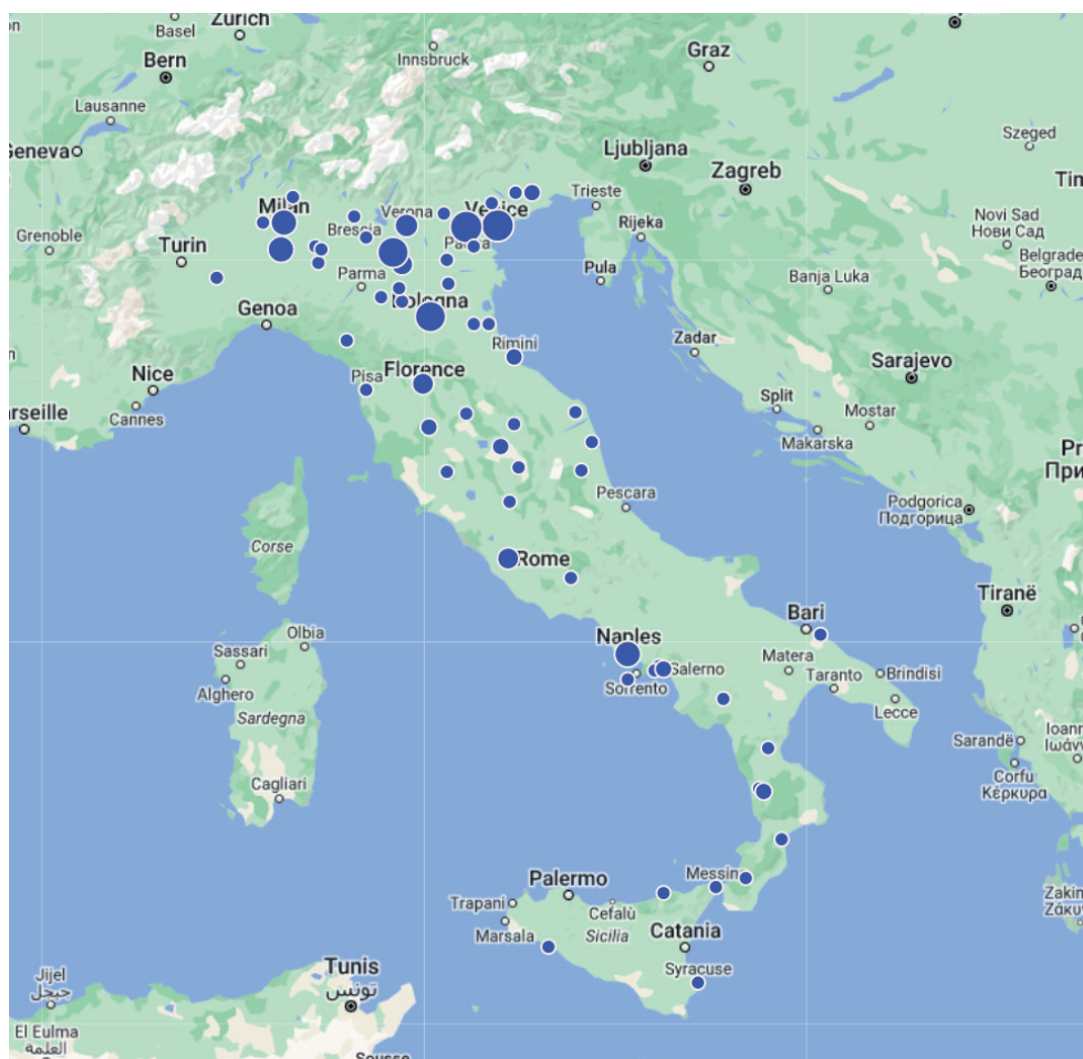


Figure 3. Territorial distribution of the single copies of the *Sphaera mundi* in 1599–1601. Each dot on the map accounts for the presence of at least one edition of the *Sphaera*. The dimensional gradient of the dots varies when multiple editions are present in the same location. Locations are meant as cities or settlements (not as single convent or monastery).

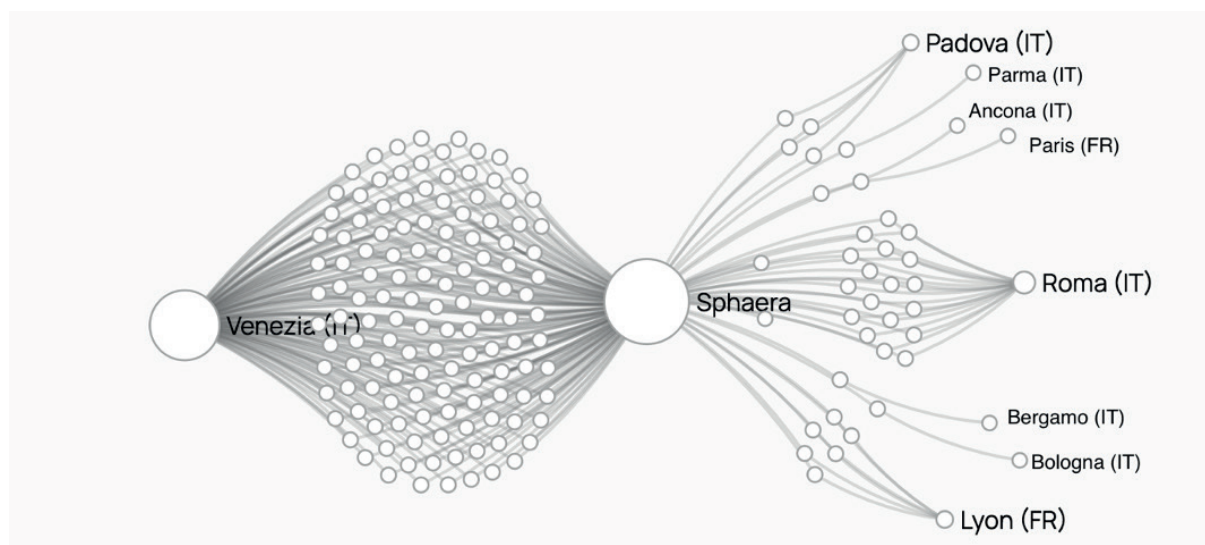


Figure 4. Provenance of the *Sphaera* copies detected in the sample. Editions of the *Sphaera* present in the global sample are grouped by their corresponding city of publication. The unlabeled connecting dots between the element “*Sphaera*” at the center and the city of print correspond to the single editions listed in the corpus *Codices vaticani latini* 11266-11326.

4. The cultural morphology of the collections

The assertion that the ownership of the *Sphaera* can be used to identify book collections that lean towards scholarly literature remains to be tested, and this section will be devoted to that task. In addition, this section will examine the cultural morphology of the collectors as revealed by their holdings and will try to find any common patterns. In particular, the research question here will be: What were the book acquisitions of individual Italian clergymen at the time? Did it make a difference if the declared status of the owner was postgraduate or pre-graduate? Moreover, was there any difference in the book diet of the few seculars selected in the sample?

In order to answer this question, thirteen collections were compared; the selection was made based on the social ranking of the owner stated in the *incipit* or *explicitly* in the list submitted to the Congregation of the Index, except for Arcangelo Giani, whose academic achievements are otherwise known.⁸ The resulting sample is the following:

ELE442	“Valeriano, baccelliere”
ELE5575	“Ioannes Baptista Vissani, maestro”
ELE6288	“Arcangelo Giani”
ELE3973	“Tomasso Felice da Ascoli, maestro”
ELE3133	“Amedeo “de Obizis,” magister”
ELE2798	“Hippolitus Mantuanus, lettore”
ELE2428	“Franciscus ab Atteste, lettore”
ELE1719	“Leonardus de Veritate, predatore e lettore”
ELE1893	“Franciscus a Vicentia, prelettore teologo”

⁸ For a biography of Arcangelo Giani, see Busolini (2000). On his book collection see Bruni (2013).

ELE2321	“Io. Bonifatio Sperello da Moro de Valli, maestro e priore”
ELE1850	“Girolamo da Teramo, baccelliere”
ELE1519	“Gio. Andrea Prigisano, medico”

This group is composed of a number of trained professionals (lecturers, masters, theologians and one medical doctor) and two trainees (*baccellieri*: students with a first degree). In Fig. 5, individual collections (ELE) are cross-referenced with the authors of the books listed in the collection: authors are used here as an identifier of the scholarly orientation of the collections and as a general indicator of the literary genre of the books listed. The diagram shows collections of different sizes; not surprisingly, the larger collections show great independence in terms of book acquisitions: in this respect, the unconnected nodes serve as a disclaimer of independence within this integrated system. On the contrary, intersectional areas are an indicator of common cultural patterns. The nodes with higher intersectionality have been placed at the center of the network, and the corresponding authors' names were made visible.

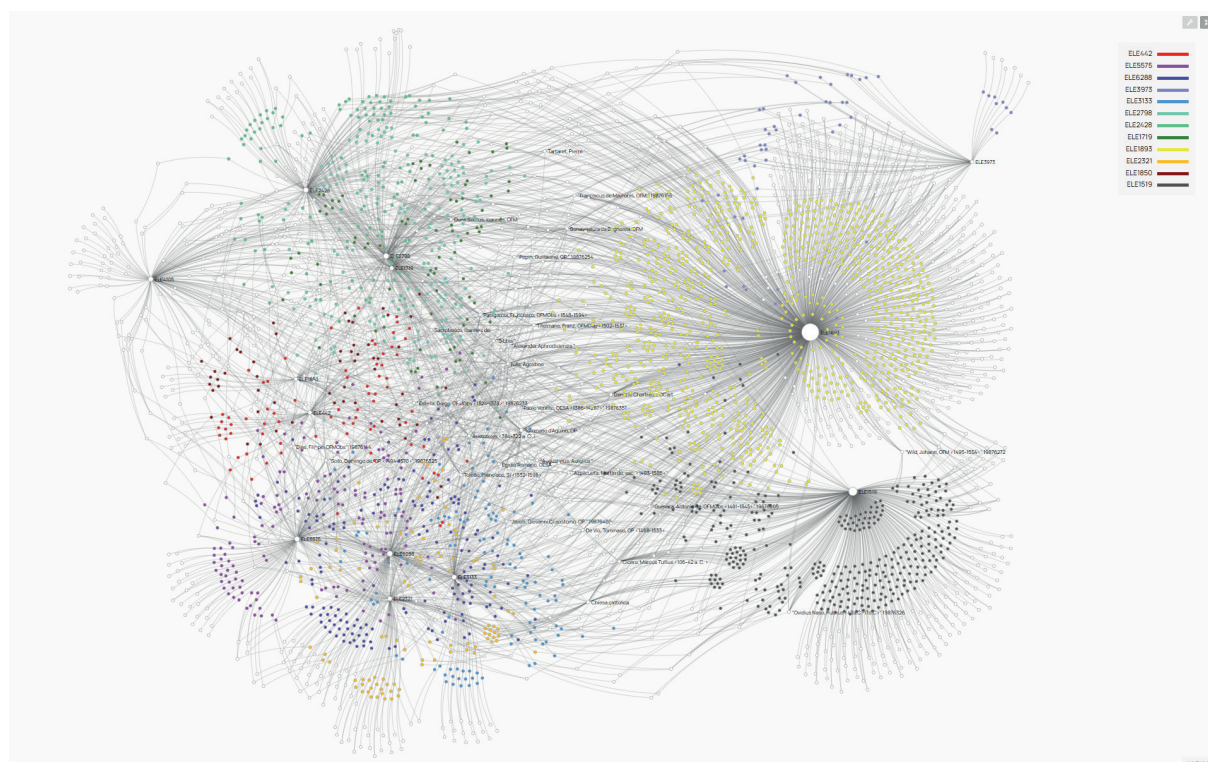


Figure 5. Comparison of twelve collections of university graduates or students. The corresponding owners in the RIC database are: ELE442, ELE5575, ELE6288, ELE3973, ELE3133, ELE2798, ELE2428, ELE1719, ELE1893, ELE2321, ELE1850, ELE1519. Along with the owner, the network comprises the following elements: “Main table” (i.e., its ID code) and the element “Author” derived from the single editions (RIC element: BIB) enlisted in the individual libraries (RIC element: ELE)

St. Bonaventure da Bagnorea (OFM), Denis le Chartreux (OCart), Francesco Panigarola (OFMObs), Franz Titelmans (OFMCap), and Francisco Toledo (SI) are all authors who refer to the pastoral duties of the book collectors. These book collectors are almost exclusively clerics with a high standing in their communities. Likewise, a group of liturgical texts (here indicated under the collective authority file “Catholic Church”) is an indicator of an active pastoral practice of the group of book collectors under scrutiny. Aristotle, Ioannes Duns Scotus (OFM), Egidio Romano (OESA), Agostino Nifo, and St. Thomas Aquinas (OP) provide the intellectual basis for the pastoral activity of the book collectors under examination, but they are also a reminder of the university curriculum of the past, or for the two bachelor graduates, they represent the present. The *Sphaera mundi*, a manual of astrology, sits at a distance from most of these authors, but it still shares the shelves with the elite of Christian theology and Western philosophy. The *Sphaera*’s closest companion is, paradoxically, Marcus Tullius Cicero, a guide to Latin eloquence for many educated professionals, including clerics, but also material for entry-level courses of eloquence, much like Sacrobosco’s *Tractatus* was a textbook for entry-level classes of applied mathematics. Like Cicero’s elegant prose, the medieval textbook of geocentric astrology remains an object to hold on to after a university degree was earned. It must have been perceived by its owners as a valuable collectable to display some erudition or a valuable tool to go back to in order to rehearse valuable knowledge such as applied mathematics and chronology.

When trying to find a common cultural model within a group of book collectors, it is dangerously easy to get lost in large libraries. It is in small book collections that some answers may emerge more easily. ELE 3973 (see Fig. 6), the collection of Tomasso Felice da Ascoli, MA graduate, counts a total of 54 editions, many of them by authors shared with other colleagues in this selected network. These common authors, one should assume, represent the basic educational foundation of a clergyman of some intellectual standing. Here are the authors that most relate him to his peers in the network: Nicolas Denisse (OFMObs),⁹ Franciscus de Mayronis (OFM),¹⁰ Bartolomeo Fumo (OP),¹¹ Guillaume Pepin (OP),¹² Bartolomeo Sibilla (OP),¹³ Pierre Tartaret,¹⁴ Agostino Trionfo (OESA),¹⁵ Juan Viguier (OP).¹⁶ These scholars were all authors of the theology books suitable for pastoral practices and that had little in common with Sacrobosco’s *De Sphaera*.

⁹ *In quatuor libros Sententiarum opus resolutio theologorum inscriptum* (BIB 126).

¹⁰ *Santuario* (BIB 30130).

¹¹ *Summa aurea Armilla nuncupata* (BIB 3624).

¹² *Opus admosum insigne de adventu domini* (BIB 31302).

¹³ *Speculum peregrinarum quaestionum* (BIB 478).

¹⁴ *Expositio in Summulas Petri Hispani* (BIB 2675).

¹⁵ *Summa de ecclesiastica potestate* (BIB 16110).

¹⁶ *Institutiones ad Christianam theologiam* (BIB 412).

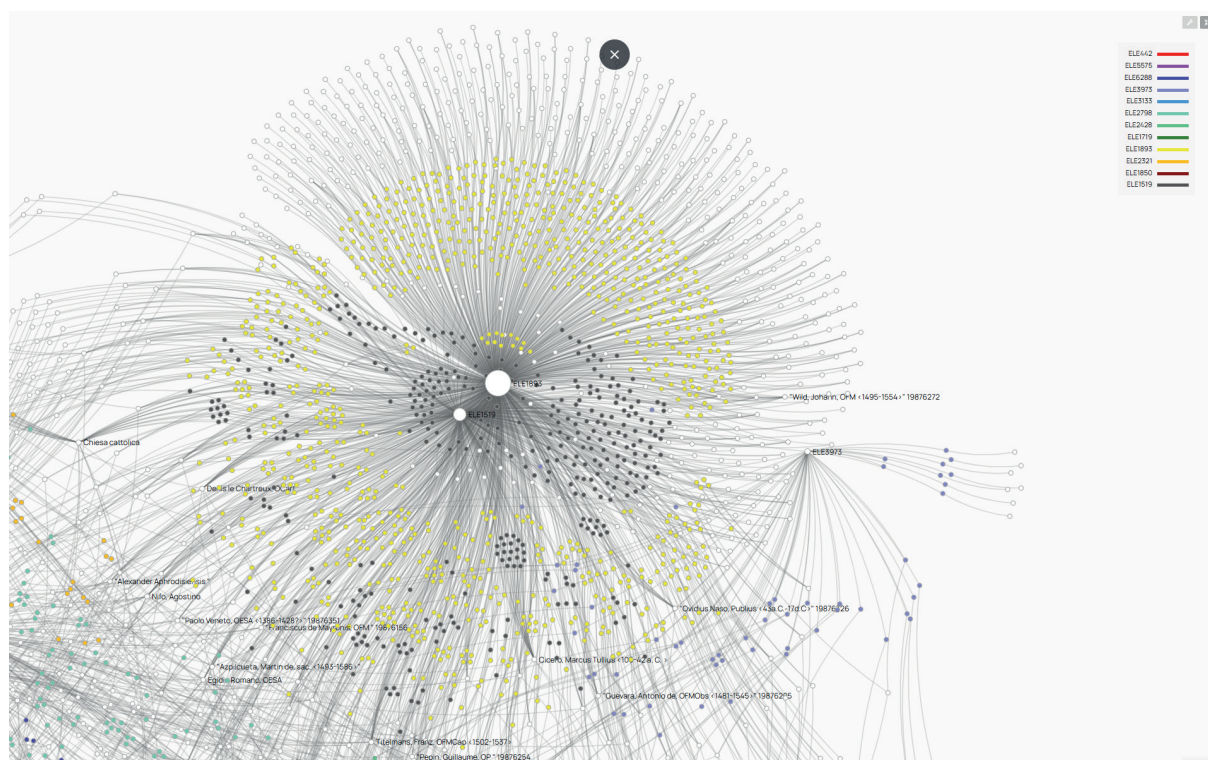


Figure 6. Close-up of Fig. 5 highlighting the book collection of Tomasso Felice da Ascoli (ELE3973) in relation to his peers.

At the other end of the network displayed in Fig. 5 are the collections of Gio. Andrea Prigisano (ELE 1519), a medical doctor, and Franciscus a Vicentia (ELE 1893), a teacher and theologian (see Fig. 6). The former is the only secular book owner singled out in the network isolated in Fig. 5; the latter owned by far the largest collection of the group. Both collections, for different reasons, show the greatest independence in the dataset (see Fig. 5).

Franciscus a Vicentia was a book owner who, as his name states, was born in the vicinity of one of the most productive sites of the European book market, Venice, where, internal evidence shows, he moved later in his life and had few problems in meeting the needs of his book acquisitions. In fact, he displays the largest collection of the group (941 items). Qualitatively, his collection shows an articulated interest in the field of philosophy, with particular reference to logic. He seems interested in argumentative theology, as suggested by the presence of two contemporary works (or nearly so in the case of Granada's book): Luis de Granada's *Ecclesiasticae rhetoricae sive De ratione concionandi libri sex* and Agostino Valier's *De rhetorica ecclesiastica*. Ironically, Franciscus a Vicentia declared ownership of a work by the very person (Agostino Valier) who, in the role of high functionary of the Congregation of the Index, had urged him to draw up his own inventory of books in the first place. Also, in the book collection of Franciscus a Vicentia, Sacrobosco's *De sphaera* is the single work related to the field of cosmology. The only ideal alliance that the *Tractatus* could be in this individual collection would be with works of Aristotelian physics, e.g., Pliny's *Historia naturale* that Franciscus owned in a vernacular translation.¹⁷

¹⁷ BIB 15893.

Gio. Andrea Prigisano, a secular medical doctor, owned a collection of books whose morphology set him apart from the rest of the group of owners of the *Sphaera* discussed here. The common denominator with Franciscus a Vicentia is a rich collection on eloquence and a similarly prominent selection of Aristotelian philosophy. Philosophical studies were indeed propaedeutic to a doctorate in medicine. What distinguishes the collection of Aristotelian philosophy of a clergyman from that of a physician is that if the former is mainly interested in logic, the latter is visibly more interested in physics.

However, the real difference in cultural attitude between an active priest like Franciscus a Vicentia and a secular physician like Gio. Andrea Prigisano is that the latter, despite having a relatively smaller collection, shows a more pronounced freedom in choosing books outside his own profession. His respectable collection of Christian literature feeds his piety and is the main link between this secular book collection and that of the clergy gathered in this sample. Perhaps the only piece of religious literature that could interfere with his practice is the *Conforto degli afflitti* (i.e., “Comfort of the Suffering”) by Gaspar de Loarte (SI), which could help him deal with difficult situations of suffering and death that his medical activity brought him close to.¹⁸ But some of his acquisitions probably simply fed his curiosity and helped him escape his daily practice: Iacopo Sannazzaro’s *Arcadia*,¹⁹ Diogenes Laertius, *De vita et moribus philosophorum libri decem* (a book consistent with his philosophical studies but also an example of virtuous living and a repository of learned secular knowledge),²⁰ an unidentified collection of proverbs,²¹ Dante Alighieri’s *Commedia*²² and Antonio Cornazzano’s *Dell’arte militare*.²³

5. Conclusions

The *Sphaera mundi* has proved its value in the discovery of learned collections in the Italian dioceses. It demonstrates that these collections stand out quantitatively and qualitatively from those of other, more modest clerics who populate the *Codices Vaticani latini* 11266-11326. Sacrobosco’s *Tractatus* had an uneven territorial distribution: more prominent near university centers and more scattered, but not absent, in rural areas. The most important conclusion to be drawn from this sample is that the *Sphaera* was virtually never located within book collections deeply invested in cosmological investigations. Rather, it was a stand-alone work used to acquire the rudiments of applied mathematics during entry-level studies by individuals with modest to significant academic and professional ambitions. Nonetheless, the *Sphaera* proved to be a valuable collectible; a book to boast of in a private collection or the reminder of a course of study that earned the owner his college degree.

¹⁸ BIB 4236.

¹⁹ BIB 17304.

²⁰ BIB 17355.

²¹ BIB 17394.

²² BIB 17404.

²³ BIB 17459.

References

- Adoriso, Antonio Maria. 1987. "Sui cataloghi perduti delle biblioteche cistercensi e fiorenti calabresi." *Rivista cistercense* 4: 161-4.
- Ago, Renata. 2002. "Collezioni di quadri e collezioni di libri a Roma tra XVI e XVIII secolo." *Quaderni storici* 110: 379-404.
- Ahnert, Ruth, Sebastian Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. 2020. *The Network Turn. Changing Perspectives in the Humanities*. Cambridge: Cambridge University Press.
- Barzazzi, Antonella. 1995. "Ordini religiosi e biblioteche a Venezia tra Cinque e Seicento." *Annali dell'Istituto storico italo-germanico in Trento* 21: 141-228.
- Barzazzi, Antonella. 2001. "La memoria di Savonarola. Testi savonaroliani nelle biblioteche dei religiosi alla fine del Cinquecento." In *Girolamo Savonarola da Ferrara all'Europa*, edited by Gigliola Fragnito and Mario Miegge, 269-84. Firenze: SISMEL - Edizioni del Galluzzo.
- Brughmans, Tom, Anna Collar, and Fiona Coward. 2016. *The Connected Past. Challenge to Network Studies in Archaeology and History*. Oxford: Oxford University Press.
- Bruni, Flavia. 2013. "The Book Inventories of Servite Authors and the Survey of the Roman Congregation of the Index in Counter-Reformation Italy." In *Documenting the Early Modern Book World*, edited by Malcolm Walsby and Natasha Constantinidou, 207-30. Leiden: Brill.
- Busolini, Dario. 2000. "Giani, Arcangelo." In *Dizionario Biografico degli Italiani* 54. https://www.treccani.it/enciclopedia/arcangelo-giani_%28Dizionario-Biografico%29/.
- Ciccarelli, Diego. 1990. "Libri francescani conventuali sardi della fine del sec. XVI." *Biblioteca francescana sarda* 4: 47-59.
- Ciccarelli, Diego. 1991. "Le biblioteche francescane conventuali nell'indagine del 1598-1603." In *Archivi, biblioteche, beni e centri culturali Atti del convegno. Assisi, 19-21 settembre 1990*, edited by Gino Zanotti, 285-303. Assisi: Studi C.I.M.P.
- Citron, Nana. 2025. *Spinning the Cosmos. Volvelles in the Early Modern Commentary*.
- Compare, Carmela. 2000. "Biblioteche monastiche femminili aquilane alla fine del XVI secolo." *Rivista di storia della Chiesa in Italia* 54: 469-516.
- Compare, Carmela. 2002. "I libri delle clarisse osservanti nella Provincia seraphica S. Francisci di fine '500." *Franciscana* 4: 169-372.
- Compare, Carmela. 2003. "Il clero e la *istruzione* delle religiose: avvertimenti monacali, esercizi particolari e pratica spirituale." In *Per il Cinquecento religioso italiano. Clero cultura società: atti del convegno internazionale di studi. Siena 27-30 giugno 2001*, edited by Maurizio Sangalli, 443-54. Roma: Edizioni dell'Ateneo.
- Compare, Carmela. 2003. "Inventari di biblioteche monastiche femminili alla fine del XVI secolo." *Genesis. Rivista della Società Italiana delle Storiche* 2: 220-32.
- Criscuolo, Vincenzo. 1986. "La biblioteca dei cappuccini di Amalfi alla fine del Cinquecento (cod. Vat. lat. 11325)." *Rassegna del Centro di cultura e storia amalfitana* 6: 65-104.

Dahnk Baroffio, Emilia. 1986. "Biblioteche religiose novaresi verso il 1600 nel censimento della Congregazione dell'Indice." *Novarien* 16: 138-47.

Dahnk Baroffio, Emilia. 1990. "Biblioteche religiose novaresi verso il 1600 nel censimento della Congregazione dell'Indice (II): La biblioteca degli Eremiti di San Girolamo dell'Osservanza." *Novarien* 20: 185-89.

Dahnk Baroffio, Emilia. 1992. "Biblioteche religiose novaresi verso il 1600 nel censimento della Congregazione dell'Indice (III): La biblioteca del monastero benedettino vallombrosano di San Bartolomeo." *Novarien* 22: 245-73.

Dahnk Baroffio, Emilia. 1993. "Biblioteche religiose novaresi verso il 1600 nel censimento della Congregazione dell'Indice (IV): La biblioteca del collegio dei Barnabiti di San Marco." *Novarien* 23: 143-58.

Dahnk Baroffio, Emilia. 1994. "Biblioteche religiose novaresi verso il 1600 nel censimento della Congregazione dell'Indice (V): La biblioteca del convento dei Carmelitani." *Novarien*, 24: 189-210.

Dahnk Baroffio, Emilia. 1995. "Biblioteche religiose novaresi verso il 1600 nel censimento della Congregazione dell'Indice (VI): La biblioteca del convento degli agostiniani di S. Nicola." *Novarien* 25: 289-318.

Dahnk Baroffio, Emilia. 1996. "Biblioteche religiose novaresi verso il 1600 nel censimento della Congregazione dell'Indice (VII): Le biblioteche dei conventi francescani di San Francesco e di San Nazzaro." *Novarien* 26: 145-68.

Fragnito Gigliola. 2006. "L'Indice clementino e le biblioteche degli Ordini religiosi." In *Libri, biblioteche e cultura degli ordini regolari nell'Italia moderna attraverso la documentazione della Congregazione dell'Indice: Atti del convegno internazionale: Macerata, 30 maggio-1 giugno 2006*, edited by Rusconi, Roberto and Rosa Marisa Borraccini Verducci, 37-59. Città del Vaticano: Biblioteca apostolica Vaticana.

Granata, Giovanna. 2006. "Le biblioteche dei religiosi in Italia alla fine del Cinquecento attraverso l'inchiesta della Congregazione dell'Indice. A proposito di libri «scomparsi»: il caso dei francescani Osservanti di Sicilia." In «*Ubi neque aerugo neque tinea demolitur*». *Studi in onore di Luigi Pellegrini per i suoi settanta anni*, edited by Maria Grazia Del Fuoco, 329-406. Napoli: Liguori.

Granata, Giovanna. 2016a. "Books without Borders: The Presence of the European Printing Press in the Italian Religious Libraries at the End of the Sixteenth Century." In *International Exchange in the Early Modern Book World*, edited by Matthew McLean and Sara K. Barker, 214-38. Leiden: Brill.

Granata, Giovanna. 2016b. "On the Track of Lost Editions in Italian Religious Libraries at the End of the Sixteenth Century: A Numerical Analysis of the RICCI Database." In *Lost Books: Reconstructing the Print World of Pre-Industrial Europe*, edited by Flavia Bruni and Andrew Pettegree, 324-44. Leiden: Brill.

Granata, Giovanna. 2024. "Italian Religious Orders and Their Books at the End of the Sixteenth Century." In *Early Modern Catholicism and the Printed Book*, edited by Justyna Kiliańczyk-Zięba and Magdalena Komorowski, 199-219. Leiden: Brill.

- Gutierrez, David. 1975. "Leonardo Coqueau O.S.A. y su biblioteca en el año 1602." *Analecta Augustiniana* 38: 5-62.
- Kerschbaumer, Florian, Linda von Keyserlingk-Rehbein, Martin Stark, and Marten Düring. 2020. *The Power of Networks. Prospects of Historical Network Research*. Londres: Routledge.
- Laudadio, Rosella. 2005. "La provincia dei frati Minori dell'Osservanza di Trinacria e i suoi libri alla fine del Cinquecento." *Franciscana* 7: 209-99.
- Lebreton, Maria Magdalena, and Aloisius Fiorani. 1985. *Codices Vaticani Latini: Codices 11266-11326: Inventari di biblioteche religiose italiane alla fine del Cinquecento*. Città del Vaticano: Bibliotheca Vaticana.
- Masetti Zannini, Gian Ludovico. 1970. "Libri di fra Paolo Sarpi e notizie di altre biblioteche dei Servi (1599–1600)." *Studi storici dell'ordine dei Servi di Maria* 20: 174-200.
- Ottone, Andrea. 2013. "Pastoral Care and Cultural Accuracy: Book Collections of Secular Clergy in Three Southern Italian Dioceses." In *Documenting the Early Modern Book World*, edited by Malcolm Walsby and Natasha Constantinidou, 231–60. Leiden: Brill.
- Pagano, Sergio. 1986. "Le biblioteche dei barnabiti italiani nel 1599." *Barnabiti studi* 3: 7-102.
- Rozzo, Ugo. 1998a. "Le biblioteche dei cappuccini nell'inchiesta della Congregazione dell'Indice (1597–1603)." In *Girolamo Mautini da Narni e l'Ordine dei Frati minori cappuccini fra '500 e '600*, edited by Vincenzo Criscuolo, 57-101. Roma: Istituto Storico dei Cappuccini.
- Rozzo, Ugo. 1998b. "Le biblioteche dei cappuccini nell'inchiesta della Congregazione dell'Indice (1597-1603)." In *Bartolomeo Barbieri da Castelvetro (1615–1697), un cappuccino alla scuola di san Bonaventura nell'Emilia del '600*, edited by Andrea Maggioli e Pietro Marinesi, 57-101. Roma: Istituto Storico dei Cappuccini.
- Rusconi, Roberto. 2004a. "I libri dei religiosi nell'Italia di fine '500." *Accademie e biblioteche d'Italia* 77: 19-40.
- Rusconi, Roberto. 2004b. "Le biblioteche dei monasteri e dei monaci della congregazione dei Celestini alla fine del secolo XVI." In *Mediterraneo, Mezzogiorno, Europa. Studi in onore di Cosimo Damiano Fonseca*, edited by Giancarlo Andenna and Hubert Houben, 961-87. Bari: Adda.
- Rusconi, Roberto, and Rosa Marisa Borraccini Verducci, eds. 2006. *Libri, biblioteche e cultura degli ordini regolari nell'Italia moderna attraverso la documentazione della Congregazione dell'Indice: Atti del convegno internazionale: Macerata, 30 maggio-1giugno 2006*. Città del Vaticano: Biblioteca apostolica Vaticana.
- Sbordone, Sara. 2001. "Le biblioteche dei cappuccini italiani nel Cinquecento. Campania (Parte prima)." *Studi e ricerche francescane* 30: 51-304.
- Sbordone, Sara. 2002. "Le biblioteche dei cappuccini italiani nel Cinquecento. Campania (Parte seconda)." *Studi e ricerche francescane* 31: 3-239.
- Sbordone, Sara. 2003. "Le biblioteche dei cappuccini italiani nel Cinquecento. Campania (Parte terza)." *Studi e ricerche francescane* 32: 3-198.

Tedeschi, John A. 1997. *Il giudice e l'eretico: Studi sull'Inquisizione Romana*. Milano: Vita e Pensiero.

Tradition of Johannes de Sacrobosco's De sphaera. Dordrecht: Springer Nature.

Ulla Lorenzo, Alejandra. 2022. "The Iberian and New World Circulation of Sacrobosco's Sphaera in the Early Modern Period." In *Publishing Sacrobosco's De sphaera in Early Modern Europe*, edited by Matteo Valleriani and Andrea Ottone, 225-54. Cham: Springer.

Valleriani, Matteo, Malte Vogl, Hassan el-Hajj, and Kim Pham. 2022. "The Network of Early Modern Printers and Its Impact on the Evolution of Scientific Knowledge: Automatic Detection of Awareness Relationships." *Histories* 2 (4): 466-503.

Valleriani, Matteo, and Andrea Ottone. 2022. "Printers, Publishers, and Sellers: Actors in the Process of Consolidation of Epistemic Communities in the Early Modern Academic World." In *Publishing Sacrobosco's De sphaera in Early Modern Europe*, edited by Matteo Valleriani and Andrea Ottone, 1-23. Cham: Springer.

Wetherell, Charles. 1998. "Historical Social Network Analysis," *International Review of Social History*, 43 : 125-44.

Zanot, Massimiliano. 2003. "Le biblioteche del Terzo Ordine della Regolare Osservanza di san Francesco in Italia alla fine del Cinquecento (1596-1600)." *Franciscana* 5: 263-341.

Zardin, Danilo. 1992. "Mercato librario e letture devote nella svolta del Cinquecento tridentino. Note in margine ad un inventario milanese di libri di monache." In *Stampa, libri e letture a Milano nell'età di Carlo Borromeo*, edited by Nicola Raponi and Angelo Turchini, 135-246. Milano: Vita e pensiero.

Zardin, Danilo. 1999. "Libri e biblioteche negli ambienti monastici dell'Italia del primo Seicento." In *Donne, filosofia e cultura nel Seicento*, edited by Pina Totaro, 347-83. Roma: Consiglio nazionale delle ricerche.

The Filing plan. Legal, Technical, and Practical Issues

Gianni Penzo Doria^(a)

a) University of Insubria, <https://orcid.org/0000-0002-7654-6638>

Contact: Gianni Penzo Doria, gianni.penzodoria@gmail.com

Received: 30 December 2024; **Accepted:** 10 March 2025; **First Published:** 15 May 2025

ABSTRACT

The new Guidelines on the creation, management, and preservation of digital documents, approved by the Italian Digital Agency (AgID) and effective as of January 1, 2022, place particular emphasis not on individual records, but rather on fundamental archival units and series.

To adopt AgID's terminology, the focus of document management and preservation has shifted to "records aggregations", namely series and files, accompanied by all the relevant finding aid.

Within public administrations, however, compliance can only be achieved following the preparation of the table of administrative procedures (along with related matters and activities). It is, indeed, essential to bear in mind that within the realm of administrative actions, four key regulatory frameworks converge: those concerning transparency, understood as total accessibility; the exercise of the right of access (to the file, rather than just individual documents); and, within the scope of digital administration, the digital file.

This study, following an analysis of the current regulatory framework, aims to elucidate the drafting methodologies for the Filing Plan (the *nomen juris* prescribed in the Guidelines) and the associated metadata set. The underlying logic, accordingly, ensures continuity between the creation and preservation phases of records.

KEYWORDS

Classification; Filing Plan; Records aggregations; Archival series; AgID Guidelines.

Il piano per la fascicolatura. Questioni giuridiche, tecniche e applicative

ABSTRACT

Le nuove *Linee Guida sulla formazione, gestione e conservazione dei documenti informatici* approvate dall'Agenzia per l'Italia digitale (AgID) e in vigore dal 1° gennaio 2022, pongono la massima attenzione non tanto sul singolo documento, quanto piuttosto sulle unità archivistiche di base e sulle serie.

Volendo utilizzare il lessico di AgID, ora al centro della gestione e della conservazione dei documenti ci sono le "aggregazioni documentali", cioè serie e fascicoli, con tutti i mezzi di corredo di riferimento.

All'interno delle amministrazioni pubbliche, tuttavia, non è possibile adempiere se non dopo la redazione della tabella dei procedimenti amministrativi (unitamente ad affari e attività). Non è opportuno, infatti, dimenticare che nell'ambito dell'azione amministrativa vengono a intersecarsi quattro norme molto importanti sulla trasparenza intesa come accessibilità totale, sull'esercizio del diritto di accesso (al fascicolo, non più soltanto ai documenti) e, nell'ambito dell'amministrazione digitale, sul fascicolo informatico.

Il saggio, a valle di una disamina dell'ordinamento vigente, ha come obiettivo l'illustrazione delle modalità redazionali del *Piano per la fascicolatura* (*nomen juris* previsto nelle LLGG) e del set di metadati ad esso connesso. La logica sottesa, coerentemente, evita le soluzioni di continuità dalla formazione alla conservazione dei documenti archivistici.

PAROLE CHIAVE

Classificazione; Fascicolatura; Aggregazioni documentali; Serie archivistiche; Linee guida AgID.

1. La fascicolatura come fulcro della gestione documentale

Il titolare è soltanto uno dei mezzi di corredo – nemmeno quello più importante – per la gestione documentale di un soggetto produttore. Pur rappresentando un'attività oggi ineludibile, la classificazione priva di fascicolatura di per sé non è sufficiente a garantire efficacia all'azione amministrativa, al punto da rendere il *records management* da un lato formalmente e giuridicamente corretto, dall'altro monco di un elemento fondamentale e imprescindibile.¹

Ci sono uffici, ad esempio, che scelgono e fanno propri in via esclusiva uno o due gradi divisionali del titolare e, indipendentemente dalle funzioni esercitate, continuano – con pervicacia meccanica – a classificare tutti i documenti della propria unità organizzativa in maniera acritica e monotematica, nella deleteria logica adempimentale. Il DPR 445/2000 – norma dai tantissimi pregi – stabilisce che registrazione, classificazione e segnatura sono “operazioni necessarie e sufficienti”.² Non è arduo dimostrare l'insufficienza tecnica di tale statuizione, poiché il fulcro della gestione documentale (quindi del protocollo manageriale, non del protocollo notarile), è rappresentato non tanto dalla classificazione, quanto piuttosto dalle operazioni connesse alla fascicolatura. Infatti, la classificazione da sola è inefficace, poiché ogni documento deve essere fascicolato, cioè ricondotto all'unità archivistica della propria classe estrema (l'ultimo grado divisionale utilizzato). Un'identica concettualizzazione, riferita ai “livelli gerarchici inferiori” del titolare, si trova oggi nelle *Linee guida* AgID sui documenti informatici, laddove il *Piano di organizzazione delle aggregazioni documentali* (poi *Piano per la fascicolatura*) è definito come lo “Strumento integrato con il sistema di classificazione a partire dai livelli gerarchici inferiori di quest'ultimo e finalizzato a individuare le tipologie di aggregazioni documentali (tipologie di serie e tipologie di fascicoli) che devono essere prodotte e gestite in rapporto ai procedimenti e attività in cui si declinano le funzioni svolte dall'ente”.³

2. Due strumenti per la dichiarazione di contesto: titolare e fascicolo

È opportuno, prima di proseguire, riprendere la differenza sostanziale tra classificazione e fascicolatura come due distinte dichiarazioni di contesto. Classificare un documento significa identificare l'ambito funzionale nei gradi divisionali in cui si articola il titolare e cui fa riferimento il contenuto – *res* giuridica, amministrativa, gestionale – in ossequio al principio della funzione esercitata.⁴ Fascicolare, invece, significa identificare l'azione amministrativa cui il documento fa riferimento per il rispettivo affare, attività o procedimento amministrativo e, come diretta conseguenza, mantenere stabili le evidenze del vincolo archivistico.

Nel primo caso siamo di fronte a una dichiarazione di contesto funzionale. Essa, dunque, è intrinseca rispetto al contenuto, cioè alla *res* sulla quale incide l'atto, anche in potenza. Nel caso del fascicolo, invece, si tratta di una dichiarazione di contesto procedimentale (o di affare o di attività)

¹ Tralascio il tema dell'applicazione formale e non sostanziale del titolare. Il tema era stato denunciato, tra i primi, da Antoniella (1999) e da Giuva (1999).

² DPR 28 dicembre 2000, n. 445, art. 56, comma 1.

³ Agenzia per l'Italia digitale, determina del Direttore generale 17 maggio 2021, n. 371, *Linee guida sulla formazione, gestione e conservazione dei documenti informatici*, integrativa della determina 9 settembre 2020, n. 407.

⁴ Per la classificazione in base al principio della funzione esercitata (Penzo Doria 2022a).

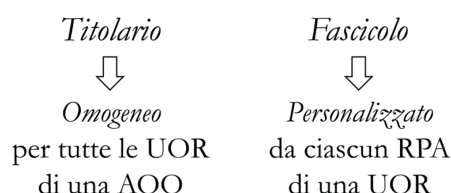
estrinseca, in quanto dipendente dal modello di organizzazione, ma indipendente dal contenuto dei documenti, attuato dal responsabile del procedimento e dalle varie regole interne al soggetto produttore, ad esempio, il Manuale di gestione documentale.

In altre parole, il titolare è una suddivisione convenzionale, logica, astratta e in potenza rispetto agli ambiti funzionali di un soggetto produttore, tanto da risultare un funzionigramma rappresentato in forma gerarchica. Per tale ragione, possiamo definire il titolare come una *tassonomia gerarchica funzionale*, priva di correlazione con la tecnostruttura, cioè con l'organigramma. L'azione amministrativa, infatti, si sedimenta come residuo documentale di efficacia probatoria nei fascicoli istruiti seguendo il piano per la fascicolatura, come possiamo evincere nell'immagine seguente.



Infatti, in un anno determinato alcuni gradi divisionali del titolare possono risultare privi di documenti da classificare. Di contro, tutti i fascicoli prodotti, in ossequio ai principi di tipicità e di determinatezza dell'azione amministrativa (Tessaro e Bertin 2021; Caringella 2022) sono sempre legati alla concreta attività pratica posta in essere da un soggetto produttore tramite il responsabile del procedimento (RPA o anche del RUP).⁵

Dunque, mentre il titolare è omogeneo per tutte le unità organizzative responsabili (UOR) afferenti a un'area organizzativa omogenea (AOO), il fascicolo è appannaggio dei vari RPA, i quali ne personalizzano l'istruzione anche in sottofascicoli e inserti in base alle proprie esigenze pratiche, come si evince nell'immagine seguente.



⁵ È improbabile che la funzione *Stemma, gonfalone e sigillo* sia utilizzata correntemente o annualmente. Essendo funzione, pur non trovando ricorrenze documentali frequenti, non può mancare nel titolare. L'acronimo RUP (responsabile unico del procedimento) è stato ridenominato in *responsabile unico di progetto* dal D.Lgs. 36/2023, art. 15.

Possiamo ora enucleare per la classificazione una funzione distinta rispetto alla fascicolatura. La prima tende a una tassonomia concettuale e non applicativa, senza la possibilità di individuare in maniera univoca un documento. Su questo punto, anzi, esiste una sopravvalutazione dottrinale del titolario, che si riscontra non raramente in dottrina archivistica, addirittura nello straordinario saggio “spartiacque” di Pavone. Infatti, pur avendo chiara la portata dello strumento, asseriva: “L’introduzione dei titolari spezzò programmaticamente questo rapporto ingenuo ed aurorale fra archivio ed istituto. Il titolario mirava a rendere facile e pronto, ai fini del miglior funzionamento dell’istituto o ufficio, il reperimento di un singolo atto in mezzo alla mole sempre crescente di tutti gli altri, basandosi soprattutto su un criterio classificatorio delle competenze” (Pavone 1970).⁶ La fascicolatura, invece, assume una funzione aggregativa (fisica o logica), tanto da organizzare i documenti classificati in maniera omogenea in unità archivistiche o in serie (fascicoli, sottofascicoli e inserti, definiti anche “aggregazioni documentali”, al pari delle serie) secondo le necessità specifiche del rispettivo responsabile del procedimento amministrativo. Infatti, nelle amministrazioni pubbliche, il titolario è utilizzato indistintamente da tutti gli uffici in maniera omogenea come mezzo di corredo condiviso dell’archivio corrente, mentre il fascicolo afferisce alle competenze di organizzazione documentale proprie del responsabile del procedimento amministrativo. Tale autonomia procedimentale è prevista espressamente dall’art. 6, comma 1, della legge 241/1990. Semplicemente scorrendo i verbi elencati a capoverso (valutare, accertare, proporre, curare e adottare) risulta evidente come le azioni in capo al responsabile del procedimento amministrativo siano di natura documentale. Ciò avviene anche in ossequio al principio di documentalità, che impone – in armonia con l’art. 2699 del Codice civile – la redazione di un documento per ogni atto pubblico adottato.

Dunque, sostenere che il titolario sia lo strumento principe dell’archivio corrente è fuorviante perché, se lasciato avulso dal contesto delle aggregazioni documentali, l’applicazione diviene fine a se stessa. Infatti, la gestione documentale, soprattutto in ambiente digitale, è totalmente inefficace se priva della fascicolatura e dell’organizzazione delle serie.⁷

Inoltre, il titolario – in astratto e per quanto potrebbe risultare poco incline all’ortodossia un’affermazione del genere – pur essendo dal punto di vista della normativa attuale una condizione per la fascicolatura, non è in realtà archivisticamente indispensabile, né una condizione irrinunciabile e bloccante. Fosse il contrario – e non rileva la produzione alluvionale cui assistiamo da almeno mezzo secolo – potremmo sostenere che gli archivi antecedenti al XIX secolo fossero disorganizzati?

C’è di più e qui ci soccorre la storia degli archivi. Mentre la classificazione troverà spazio solo a partire dalla fine del sec. XVIII, le aggregazioni documentali erano ben note ai soggetti produttori, i quali, fin dal XII secolo, avevano organizzato i documenti per tipologia documentale, per forma, per magistratura o per contenuto, in “fascicoli”, serie, filze, indipendentemente da una classificazione funzionale, che avrebbe visto la luce soltanto secoli dopo.

⁶ Per le unità archivistiche restano tuttora di riferimento Carucci (1987) e Duranti (1997), mentre per la crisi dei mezzi di corredo per la gestione documentale Guercio (2002). Più di recente, anche per le affinità e per le differenze tra mezzi di corredo e strumenti di ricerca Nimore (2023). Sull’opera di Pavone, si veda di recente Di Marcantonio (2019), Valacchi (2019), Di Marcantonio (2022) e la monografia che chiude un percorso di ricerca sul grande studioso Di Marcantonio (2023).

⁷ Ottimo il lavoro di Alfier (2021) e, sotto il profilo più informatico, Giacalone (2024).

2.1 Tre operazioni per la gestione documentale efficace

La sedimentazione dell'archivio si attua attraverso tre operazioni fondamentali, in armoniosa sequenza:

- a) *registrazione* (a protocollo o a repertorio), con conseguente attribuzione di data e di provenienza certe, laddove “certe” significa anche certificate dal pubblico ufficiale incaricato dell'operazione;
- b) *classificazione*, con conseguente identificazione dell'ambito funzionale-amministrativo cui essi appartengono;
- c) *fascicolatura*, con conseguente aggregazione alle altre unità documentali e attribuzione della responsabilità giuridica da parte di un modello organizzativo (unità organizzativa responsabile – UOR) e di una persona fisica (responsabile del procedimento amministrativo – RPA), in base al principio di responsabilità.

Mentre la classificazione coinvolge un aspetto esclusivamente gestionale, la registrazione dei documenti e la sedimentazione efficace del fascicolo hanno ricadute anche sul fronte delle plurime responsabilità nei confronti delle discipline in materia di trasparenza amministrativa (*accountability*)⁸ e di protezione dei dati personali.⁹

Analogo ragionamento risulta possibile anche rispetto all'esercizio del diritto di accesso, dal momento che l'ostensione dei documenti non è limitata, ma estesa. Infatti, si può accedere all'unità archivistica e non soltanto al singolo documento, seguendo la regola sandulliana dell'ostensibilità del fascicolo, intesa come *actio ad exhibendum in extensum* (Sandulli 1984, 667). Non a caso, non di sola archivistica vive il fascicolo, intriso com'è di diritto (civile e amministrativo, in particolare), protezione dei dati personali e di informatica (soprattutto informatica giuridica).¹⁰

Il fascicolo, inoltre, non è uno strumento classificatorio, ma di organizzazione coerente rispetto alla classificazione dei documenti che ne fanno parte, nella misura in cui ogni documento assume una rilevanza – giuridica, amministrativa, fiscale, archivistica, etc. – solo se connesso alla rispettiva unità archivistica o alla rispettiva serie, cioè ai vari contesti di riferimento, da cui possiamo evincere la definizione di *archivistica come scienza del contesto*.

Con riferimento agli albori dell'introduzione del titolario, la pietra angolare – ma non la sola – fu la sistematizzazione asburgica della gestione documentale in virtù delle *Istruzioni per il Governo* del 1815, di fatto assimilabili alle odierne *regole tecniche* o *linee guida* della normativa italiana, ma anche con molti spunti per un moderno *manuale di gestione ante litteram*. Pubblicate a stampa, contengono – in pieno Congresso di Vienna – alcune regole antesignane di gestione documentale, tanto per la classificazione, quanto per la fascicolatura.¹¹

⁸ Art. 5, commi 1 e 3, del D.Lgs. 33/2013, e l'art. 6, comma 1 del D.Lgs. 97/ 2016. Noto il commento sul burocrate di Zanni Rosiello (2015).

⁹ Regolamento (UE) 2016/679 (GDPR), sul quale cfr. nota 18, in particolare all'art. 24, soprattutto per le misure che il responsabile del trattamento deve mettere in atto nel caso in cui il fascicolo conservi documenti contenenti dati cd. “sensibili” (oggi declinati dall'art. 9 come *categorie particolari di dati personali*).

¹⁰ Su questo punto, il regolamento attuativo della legge 241/1990, contenuto nel DPR 184/2006, art. 7, comma 2, è limpido: si accede al fascicolo, non al documento.

¹¹ Il regolamento è stato pubblicato nello studio di Tonetti (2000). La novità rispetto al titolario delle Prefetture napoleoniche del 1803 è rappresentata dall'apparato minuzioso di regole e di procedure (Spaggiari 2022, 161-71), mentre molto importante risultano i contributi di Martorano (2020) e di Brunetti (2016; 2022); infine, per un commento al regolamento asburgico si veda Penzo Doria (2022b).

Qualche decennio più tardi, le regole asburgiche saranno fonte di grande ispirazione – se non addirittura di mutazione – per la normativa italiana del 1897 e del 1900.¹²

3. Un esempio: il Comune di Milano

Prendiamo come esempio il titolario per gli archivi comunali utilizzato dal Comune di Milano nella versione del dicembre 2005. La funzione inerente a “Concorsi, selezioni e colloqui” trova accoglienza nel titolo III, classe 1. Ciò significa che i documenti inerenti al reclutamento tramite procedure concorsuali assumono classificazione normalizzata in III/1.¹³

Nel corso del 2021 il direttore dell’Area acquisizione risorse umane bandì un centinaio di concorsi, tra i quali uno per 6 posti di dirigente tecnico (DirT, con determina 9 novembre 2021, n. 9658), una selezione per un autista e una per un muratore (B3CST, entrambe con determina 15 novembre 2021, n. 9897).¹⁴ Se ipotizziamo una trentina di domande per ciascuno dei cento concorsi, siamo di fronte a 3.000 documenti in arrivo solo per le istanze di partecipazione, senza contare le dichiarazioni di non incompatibilità, le determine di nomina delle commissioni e di integrazione di membri dimissionari, i verbali e i provvedimenti di approvazione atti.

Ebbene, per ogni concorso è istruito un fascicolo procedimentale nell’anno 2021, dando atto che i documenti esaminati finora hanno *omogenea* classificazione in III/1. Ed ecco la cartina di tornasole: ciò che contraddistingue i fascicoli in maniera univoca non è l’indice di classificazione, ma la segnatura archivistica.



Se supponiamo che il concorso per dirigente tecnico sia il 79°, per autista l’81° e per muratore il 94°, a parità di anno di istruzione e di indice di classificazione, avremo tre segnature archivistiche differenti. Seguendo la sintassi per la quale l’anno di istruzione è separato dall’indice di classificazione da un trattino (-), i gradi divisionali da una barretta (/) e le unità archivistiche da un punto (.), con l’oggetto racchiuso tra virgolette caporali (« »), otterremo:

- 2021-III/1.79 «Concorso DirT – Selezione per sei posti di dirigente tecnico»
- 2021-III/1.81 «Concorso B3CST – Selezione per un posto di autista»
- 2021-III/1.94 «Concorso B3CST – Selezione per un posto di muratore»

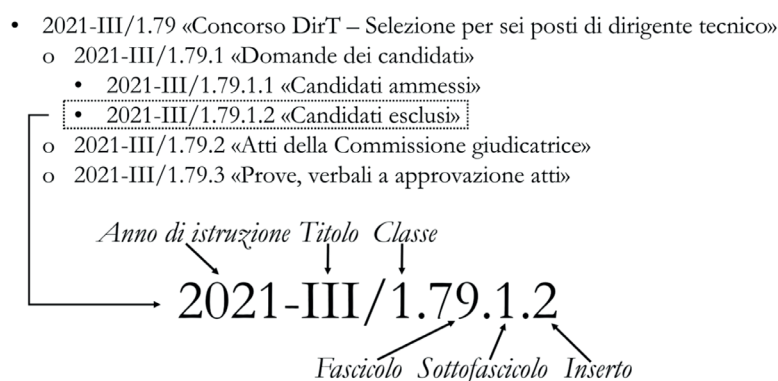
¹² Ministero dell’Interno, 1° marzo 1897, n. 17.000-2, detta *Circolare Astengo* e RD 25 gennaio 1900, n. 35.

¹³ Per lo strumento si vedano Bonfiglio Dosio e Pavone (2007), mentre per una panoramica in prospettiva storica Brunetti (2012). Il titolario del Comune di Milano è incluso nel Manuale di gestione, adottato dalla Giunta comunale, delibera 17 febbraio 2017, n. 209.

¹⁴ Comune di Milano, sezione trasparenza: <https://www.comune.milano.it/comune/amministrazione-trasparente/bandi-di-concorso> (verificato il 15 novembre 2024).

Ipotizziamo ora che il RPA, ai sensi del § 10.33.1 del Manuale di gestione del Comune di Milano, abbia deciso di ripartire il fascicolo in tre sottofascicoli e che il primo sottofascicolo sia a sua volta articolato in due inserti, l'ultimo dei quali dedicato ai documenti inerenti ai candidati esclusi, avremo: 2021-III/1.79.1.2 «Candidati esclusi».

Ora inseriamo la sequenza di aggregazioni documentali nell'immagine seguente.



La stringa ha il pregio di prestarsi a una lettura bustrofedica e, pertanto, di potersi leggere da destra a sinistra, così come da sinistra a destra. Cominciamo la lettura da sinistra: fascicolo istruito nell'anno 2021, sotto il titolo III, nella classe 1, con il numero di repertorio 79, di sottofascicolo 1 e inserto 2. Viceversa: inserto 2, del sottofascicolo 1, del fascicolo 79, della classe 1, sotto il titolo III, istruito nell'anno 2021.

Dunque, l'organizzazione in fascicoli, sottofascicoli e inserti non dipende dal titolario, cioè dalla classificazione sistematica in potenza sottesa alle funzioni esercitate, ma dall'organizzazione delle aggregazioni documentali in atto, che sfuggono a una logica tassonomica preordinata, anche se quest'ultima resta importante, ancorché non imprescindibile.

4. Fascicolatura o fascicolazione?

Il termine *fascicolatura* – come abbiamo visto *supra* – ricorre al § 3.3.2, ma anche più volte nel set di metadati previsti dall'allegato 5 delle *Linee guida* AgiD. Pertanto, *fascicolatura* è *nomen juris*.

Al di là di stucchevoli questioni nominalistiche, è importante ribadire che i termini con suffisso in *-tura*, quindi con morfema utilizzato spesso per sostantivi femminili derivati da verbi indicanti un'azione attiva, sono largamente utilizzati nel lessico archivistico presente nelle tre età, dal protocollo all'archivio storico: campionatura, cartellinatura, marcatura, punzonatura, registratura, schedatura, segnatura, spolveratura, timbratura e così via.

Quindi, fascicolatura o fascicolazione? Non sussistono ragioni decisive in favore dell'una o dell'altra forma, in virtù del fatto che l'italiano è una lingua viva e dinamica. Il termine utilizzato dalla burocrazia asburgica è fascicolatura, al pari di registratura, come citato anche negli studi di Tonnelli (2000; 2003, 124).

Pertanto, in aderenza alle *Linee guida* AgID, è preferibile utilizzare il termine *fascicolatura*, anche come effetto giuridico dell'azione del fascicolare, non prima di aver chiarito che il termine *fascicolazione* non risulta erraneo.

4.1 Fascicolatura, information retrieval e Piano delle performance (PIAO)

L'esperienza sul campo ci insegna quanto i boiardi siano riottosi alla corretta gestione documentale. Due sono le eccezioni principali che essi pongono di fronte alle proprie inadempienze. In primo luogo, la carenza di tempo da spendere in adempimenti ritenuti "burocratici". In seconda battuta, la convinzione che, nell'era dell'*information retrieval*, operazioni come la classificazione e la fascicolatura siano ormai superate dalla tecnologia.

A chi eccepisce in questo modo, giova ribadire che quei pochi secondi impiegati nella fascicolatura saranno di gran lunga restituiti in luogo delle defatiganti e costosissime (anche in maniera indiretta, in termini tempo e di personale) ricerche di documenti nel rispettivo contesto (e non isolati). Nella realtà, un dirigente può toccare con mano a cosa porta il disordine documentale, non solo in termini di perdita di tempo, ma anche di valutazioni improprie, in quanto non supportate da un'istruttoria tassonomica, di censure giudiziali per un negato o mal articolato riscontro a una richiesta di accesso o all'evasione di una pratica senza qualche evidenza documentale, che spesso si rivela decisiva.

Certo che Google vincerà sempre contro la classificazione e la fascicolatura qualora si cercasse un solo documento determinato e ben definito. Tuttavia, Google risulterebbe estremamente pericoloso e fuorviante nella ricerca di tutti i documenti di una determinata pratica. Inoltre, mentre la ricerca archivistica è neutra e non colleziona documenti in base alla volontà del ricercatore o alla sua posizione geo-referenziata, Google modifica i criteri in base a profilature di contesto (luoghi, interessi personali, precedenti ricerche e così via), tant'è che una stringa di ricerca identica potrebbe ottenere risultati molto differenti se la ricerca parte da Roma o da New York.

Non ultimo, la fascicolatura consente, grazie alla dichiarazione di contesto che si esprime attraverso il vincolo archivistico, tutelato proprio dalla fascicolatura, di ritrovare anche documenti registrati male o in modo approssimativo (con oggetti non parlanti, con errori di battitura nei campi a testo libero, come mittente/destinatario etc.). Infine, gli strumenti archivistici garantiscono nel tempo la persistenza giuridica e la certezza di azioni e della loro rappresentazione sotto forma di documenti, o meglio: di poter capire le ragioni di certe azioni o decisioni. Garantiscono, in una parola, la "memoria illuminata e affidabile".

Google, di contro, fa il collezionista di informazioni sparse nella rete, ma prive di una visione organica di tipo procedimentale, prive del filo conduttore ineludibile del procedimento che le unisce (lo potremmo definire *vincolo procedimentale*). Insomma, l'esatto contrario di quello che serve a un cittadino per dialogare con un'amministrazione pubblica o a uno studioso per la ricerca storica.

Mentre gli strumenti archivistici sedimentano la memoria con cura tecnico-giuridica, conservando tutti i documenti prodotti di un procedimento determinato, Google è un raffazzonatore occasionale, inidoneo a garantire l'esistenza o meno di un documento determinato (magari privo di oggetto o con metadati sbagliati o incompleti, pur appartenendo a un fascicolo procedimentale). Inoltre, dal momento che i burocrati agiscono soltanto in presenza di una norma che preveda un adempimento, è possibile ricordare che il *Codice dell'amministrazione digitale*, obbliga le amministrazioni pubbliche a istruire un fascicolo per ciascun procedimento amministrativo.¹⁵

¹⁵ D.Lgs. 82/2005, art. 41, comma 2.

Non bastasse, oltre al già citato DPR 445/2000 e alle già richiamate *Linee Guida* AgID, si può far riferimento puntuale alla previsione contenuta nell'art. 35 del *Decreto trasparenza*, ricco di adempimenti documentali e procedurali inerenti proprio alla classificazione e alla fascicolatura.¹⁶ Infine, tra gli obiettivi che l'organo di vertice può assegnare nel *Piano delle performance* (oggi PIAO) ai propri dirigenti e quest'ultimi agli uffici di riferimento, si rinviene agevolmente anche il piano per la fascicolatura, con effetto l'assenza al 31 dicembre di ogni anno di documenti sciolti, cioè privi della rispettiva dichiarazione di contesto procedimentale nel sistema documentale o, più banalmente, nel protocollo informatico.¹⁷

5. Il metodo redazionale

Per fare in modo che il piano per la fascicolatura non sia avulso o considerato estraneo ai sistemi direzionali apicali e agli uffici trasversali o di staff, risulta opportuno che il sistema archivistico sia integrato con gli altri sistemi utilizzati dall'apparato amministrativo. Conviene riflettere, dunque, non solo in base alla dottrina archivistica, ma anche in virtù delle informazioni previste dalla normativa vigente in materia di trasparenza amministrativa, procedimento amministrativo e aree di rischio corruttivo, senza dimenticare la disciplina sulla protezione dei dati personali.¹⁸ In questo caso, è possibile fare riferimento, in particolare, alla legge 241/1990, alla legge 190/2012, al D.Lgs. 33/2013 e alle norme di riforma dell'amministrazione pubblica, compresa l'anticorruzione.¹⁹

Il tracciato dei dati da inserire nel piano per la fascicolatura è semplice, perché nasce dagli elementi obbligatori per il protocollo già previsti da un quarto di secolo dal DPR 445/2000 (in realtà dal previgente DPR 428/1998), dalle *Linee guida* AgID e dagli elementi ai fini della trasparenza amministrativa previsti dal D.Lgs. 33/2013. Quest'ultimo, infatti, ha novellato all'art. 23, comma 2, che: "Per ciascuno dei provvedimenti compresi negli elenchi di cui al comma 1 sono pubblicati il contenuto, l'oggetto, la eventuale spesa prevista e gli estremi relativi ai principali documenti contenuti nel fascicolo relativo al procedimento. La pubblicazione avviene nella forma di una scheda sintetica, prodotta automaticamente in sede di formazione del documento che contiene l'atto".

Nella disposizione appena richiamata la triade perfetta per la trasparenza amministrativa è costituita da documenti, fascicoli e procedimenti amministrativi. Poi il D.Lgs. 33/2013 ha enucleato minuziosamente, al successivo art. 35, comma 1, l'elenco dei dati da pubblicare. Tolti gli elementi non strettamente rispondenti al nostro scopo, possiamo a questo punto tracciare gli elementi del

¹⁶ D.Lgs. 33/2013, art. 35, comma 1.

¹⁷ D.Lgs. 150/2009, art. 10, comma 1. Oggi è confluito nel *Piano integrato di attività e organizzazione* – PIAO (D.Lgs. 80/2021, art. 6).

¹⁸ *Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE* (Regolamento generale sulla protezione dei dati - GDPR) e, in ambito nazionale, D.Lgs. 30 giugno 2003, n. 196, *Codice in materia di protezione dei dati personali, recante disposizioni per l'adeguamento dell'ordinamento nazionale al regolamento (UE) n. 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE*. Il titolo è stato così modificato dall'art. 1, comma 1, D.Lgs. 10 agosto 2018, n. 101. Il titolo originario, meno esteso, era infatti *Codice in materia di protezione dei dati personali*.

¹⁹ ANAC, *Piano nazionale anticorruzione* – PNA e sulle aree di rischio corruttivo, Delibera 21 luglio 2021.

piano per la fascicolatura come segue, non prima di aver ricordato che questo strumento soffre di obsolescenza organizzativa e, quindi, dev'essere aggiornato con un incarico *ad hoc*:

- Denominazione del soggetto produttore
- Identificativo del procedimento, affare, attività
- Titolo
- Descrizione
- Normativa nazionale e comunitaria
- Normativa comunale
- Classificazione
- Tipologia di fascicolo
- Sintassi di descrizione del fascicolo
- Unità organizzativa responsabile (anche altre UOR in Conferenza di servizi)
- Area
- Responsabile del procedimento
- Recapiti del RPA
- Area di rischio (compresa la presenza di dati sensibili, *rectius*, “categorie particolari di dati personali”)
- Tempo di esecuzione
- Tempo di conservazione
- Motivazione (se non a lungo termine)
- Note
- Documenti principali del procedimento

6. Le cinque tipologie di fascicolo

Prima di concludere la nostra indagine con l'esame di casi concreti, è opportuno ribadire la possibilità per le amministrazioni pubbliche di enucleare cinque tipologie di fascicolo: a) affare, b) attività, c) procedimento, d) persona fisica ed e) persona giuridica.²⁰

Come mai è necessario distinguere tali tipologie? Si tratta di un'analisi fine a se stessa o deriva la propria ragione d'essere nella gestione documentale efficace? Con ogni evidenza, la rassegna delle diverse tipologie risponde a una necessità di strutturare la fascicolatura per una gestione documentale complessiva, dall'archivio corrente all'archivio storico, in maniera puntuale, nella misura in cui è possibile guidare gli uffici e il personale a una sedimentazione ordinata dei documenti nelle rispettive unità archivistiche, come si può evincere dalla seguente immagine.

²⁰ Sono interessanti le riflessioni di Giorgi e Mineo (2009), e sulle cinque tipologie si veda anche Penzo Doria (2007).

	AFFARE	ATTIVITA'	PROCEDIM
Contiene atti esclusivamente vincolati o proceduralizzati?	☐	☒	☐
Ha contenuto esclusivo e peculiare?	☒	☐	☒
Ha contenuto tipico?	☐	☒	☒
Ha un tempo certo per la chiusura?	☐	☒	☒
Può cambiare il modello di fascicolatura in base alla mole?	☐	☒	☐
Può essere iterato anno dopo anno? È periodico?	☐	☒	☐
Si chiude con un atto finale?	☐	☐	☒

Infatti:

- le modalità di istruzione possono cambiare in base al procedimento, anche per una maggiore efficienza nella gestione documentale (il responsabile del procedimento di una selezione/concorso può istruire le unità e le sottounità in modo da favorirne la selezione, ad es., organizzando i documenti dei candidati non ammessi in sottofascicoli destinati allo scarto una volta spirato il termine per i ricorsi);
- la responsabilità in ordine all'azione trattata, alla durata, alle modalità di sedimentazione e di conservazione dipende dalla tipologia di fascicolo; ad esempio, la durata del fascicolo di procedimento è stabilita dalla legge, diversamente da quello di attività (che dipende dalla mole di documenti, mensile, semestrale, annuale) o da quello di affare (che dipende dal modello organizzativo scelto dal RPA);
- l'istruzione dei fascicoli dipende dal fine istituzionale dell'ente e dalle tipologie di azioni amministrative poste in essere (come vedremo negli ultimi due esempi in tema di scarto dei documenti dichiarati inutili, lo stesso procedimento ha due modalità di fascicolatura differenti tra l'organo di sorveglianza e il soggetto produttore sorvegliato);
- le regole per l'istruzione possono variare in base alla tipologia (un fascicolo di attività è dimensionabile in base al volume dei documenti, pure restando il fascicolo un'entità logica e non fisica); ad esempio, i pareri legali in un'Avvocatura distrettuale dello Stato possono essere fascicolati per Ente, per tematica (non materia) o, più semplicemente, per data di rilascio, contrariamente ai fascicoli dell'avvocatura interna a un ente, che seguono il procedimento o una cadenza per tematica;
- l'esercizio del diritto di accesso ha regole di sedimentazione differenti rispetto al fascicolo soggetto (l'istruzione del fascicolo contenente l'istanza di accesso ha regole differenti da quelle del fascicolo procedimentale di cui è oggetto l'istanza medesima)
- la tutela dei dati personali, sensibili e giudiziari avviene anche grazie alla corretta descrizione del fascicolo (il livello di riservatezza, compito essenziale del responsabile del procedimento, deve essere dichiarato), così come la consultabilità per scopi storici tanto nell'archivio corrente quanto nell'archivio di deposito.

Inoltre, mentre la procedura e il processo possono indistintamente essere realizzati da un ente pubblico o da un ente privato, il procedimento amministrativo appartiene alla sfera esclusiva delle amministrazioni pubbliche e dei soggetti che agiscono alla loro stregua. Questo fattore determina anche l'obbligo di stabilire i tempi entro i quali deve concludersi, nonché di attribuire la responsabilità organizzative, fisiche, giuridiche e individuabili come UOR e RPA nella *Tabella dei procedimenti amministrativi*. Quest'obbligo preciso si riverbera nella gestione documentale con l'obbligatorietà di istruire e di chiudere il fascicolo procedimentale entro un termine prestabilito dall'ordinamento interno o generale.²¹

Infine, risulta opportuna una osservazione di metodo redazionale a favore del soggetto produttore. La scelta di limitare il piano per la fascicolatura alla sola area archivistica potrebbe rivelarsi penalizzante. Uno dei punti di forza dell'archivio è di rappresentare una funzione trasversale a beneficio di tutte le unità organizzative responsabili. Per tale ragione, è bene che esso sia integrato nella gestione dei procedimenti amministrativi, nelle aree di rischio e in tutti gli adempimenti in tema di trasparenza.

In una visione dinamica dei mezzi di corredo è possibile inserire nel piano per la fascicolatura anche il tempo di conservazione delle unità archivistiche e tutte le informazioni che la disciplina vigente riversa sui responsabili dei procedimenti amministrativi e, più in generale, sulla dirigenza pubblica. In questo modo avremo coniugato diverse aree di interesse dell'amministrazione attiva, sfruttando appieno le potenzialità di una gestione documentale efficace e non avulsa – come nicchia – rispetto agli altri modelli organizzativi. L'archivio, dunque, deve essere catapultato dal responsabile (o dal coordinatore) della gestione documentale nell'agorà di tutte le funzioni cruciali che, di norma, gli organi di vertice preferiscono tenere in staff.

7. Due esempi di piano per la fascicolatura

Per restare nell'alveo della prassi, prendiamo come esempio due procedimenti con il rispettivo piano per la fascicolatura, scegliendoli per differente tipologia di soggetto produttore, con un esempio procedimentale omogeneo e con un piano per la fascicolatura differente. Esaminiamo, dunque, quelli previsti per una Università (riferito all'anno 2016) e per un Comune (2020), avendo chiara l'integrazione tra archivistica, informatica giuridica e diritto amministrativo, soprattutto relativamente al fascicolo procedimentale.

7.1 Primo scenario: il reclutamento dei disabili all'Università degli Studi dell'Insubria

Denominazione del soggetto produttore

Università degli Studi dell'Insubria

Identificativo del procedimento

n. 7612/2016

²¹ Si vedano Romano (2006); Molè (2007) e *La dirigenza degli enti locali* (2022).

Titolo

Personale tecnico amministrativo – Avviamento a selezione numerica (cat. B, disabili)

Descrizione

Ai sensi dell'art. 35, comma 1, lettera b) del D.Lgs. 165/2001, l'assunzione di categorie per le quali è previsto il solo requisito della scuola dell'obbligo (posizione iniziale cat. B) avviene mediante avviamento degli iscritti nelle liste di collocamento (ex L. 56/1987). L'Ateneo effettua tali assunzioni tramite selezione tra gli iscritti nelle liste formate dai Centri per l'Impiego provinciali (CIP). I lavoratori sono avviati numericamente secondo l'ordine di graduatoria risultante dalle liste dei Centri per l'impiego territorialmente competenti. Ai sensi dell'art. 35, comma 2 del D.Lgs. 165/2001, le assunzioni obbligatorie ex L. 68/1999 (persone disabili) per la copertura della quota d'obbligo di categorie per le quali è previsto il solo requisito della scuola dell'obbligo vengono effettuate per chiamata numerica tramite l'Ufficio Categorie Protette della Provincia. Le medesime assunzioni obbligatorie, ma per la copertura della quota d'obbligo di categorie superiori (B3, C1, D1, EP1), avvengono tramite selezioni pubbliche riservate o prevedendo una riserva di posti nelle selezioni pubbliche. In questi casi, la selezione – sia a tempo indeterminato, sia determinato – viene effettuata nel rispetto delle norme generali (D.Lgs. 165/2001, D.P.R. 487/1994), del CCNL e dei Regolamenti di Ateneo. Le selezioni (pubbliche con riserva o interamente riservate a persone disabili) sono pubblicate in Gazzetta Ufficiale (per esteso o nella forma di avviso) e all'Albo Ufficiale di Ateneo.²²

Normativa nazionale e comunitaria

Legge 28 febbraio 1987, n. 56; L. 29 marzo 1985, n. 113; L. 23 marzo 1999, n. 68; D.Lgs. 30 marzo 2001, n. 165, art. 35 DPR 9 maggio 1994, n. 487

Normativa comunale

Regolamento concorsi di Ateneo

Classificazione

VII/1 (Personale / Reclutamento e selezioni)

Tipologia di fascicolo

Fascicolo per procedimento amministrativo (tipo C)

Sintassi di descrizione del fascicolo

La denominazione del fascicolo riporta la dicitura: “Selezione numerica per...”, con la struttura seguente: 1) Parte generale, che si apre con il decreto di approvazione del bando e si chiude con il decreto di approvazione degli atti e comprende anche la documentazione generale; 2) Domande dei candidati; 3) Commissione giudicatrice, comprendente i documenti riferiti alla commissione, cioè nomina dei componenti, surroghe, dichiarazioni di incompatibilità, ecc.; 3) Verballi della

²² La descrizione è riferita al 2016, ma la materia è stata rinnovata alla luce della Direttiva n.1/2019 del Dipartimento Funzione Pubblica. Pertanto, un piano per la fascicolatura deve seguire l'evoluzione normativa, a pena dell'obsolescenza.

Commissione giudicatrice, comprese le lettere di trasmissione degli stessi. Supponendo che si tratti del 5° fascicolo della classe 1 del titolo VII istruito nell'anno 2016, la struttura del fascicolo e dei sottofascicoli sarà la seguente:

2016-VII/1.5 «Personale tecnico amministrativo - Selezione numerica per due unità - cat. B anno 2016»

2016-VII/1.5.1 «Parte generale»

2016-VII/1.5.2 «Domande dei candidati»

2016-VII/1.5.3 «Commissione giudicatrice»

2016-VII/1.5.4 «Verbali»

Unità organizzativa responsabile

Servizio Personale tecnico amministrativo

Area

Risorse umane e finanziarie

Responsabile del procedimento amministrativo

Maria Rossi

Recapiti del RPA

Via Ravasi, 2 - 21100 Varese

Tel: 0332 219090

Email: rossi.maria@uninsubria.it

Area di rischio

Elevato

Tempo di esecuzione

180 giorni dalla data di conclusione del termine per la presentazione delle candidature

Tempo di conservazione

Illimitato

Motivazione (se non a lungo termine)

Nulla

Note

Nulla

Documenti principali del procedimento

- Comunicazione posto vacante di cat. B, pos. Economica B1, nell'ambito della programmazione annuale oppure Accertamento dell'obbligo di assunzione di un centralinista non vedente
- Mobilità ai sensi degli artt. 30 del D.Lgs. 165/2001

- Decreto del Direttore Generale per avviso pubblico di selezione
- Pubblicità dell'avviso – Richiesta di pubblicazione in Gazzetta ufficiale e all'Albo on-line di Ateneo
- Richiesta di avviamento a selezione numerica alla Provincia
- Comunicazione dei nominativi dei lavoratori avviati da parte della Provincia
- Richiesta di disponibilità di aule e richiesta di ausili ex art. 20 della legge
- Richiesta di autorizzazione di far parte della Commissione giudicatrice per componente esterni all'Ateneo
- Autorizzazione a far parte della Commissione giudicatrice per componenti esterni all'Ateneo
- Decreto del Direttore Generale di nomina della Commissione giudicatrice
- Trasmissione del Decreto del Direttore Generale di nomina ai componenti della Commissione giudicatrice
- Trasmissione del Decreto del Direttore Generale di nomina dei componenti della Commissione giudicatrice alla Consigliera di fiducia
- Convocazione dei candidati alle prove
- Verbale della Commissione giudicatrice
- Decreto del Direttore Generale di approvazione atti concorsuali
- Richiesta di pubblicazione all'Albo ufficiale di Ateneo dell'esito della selezione dei lavoratori avviati
- Comunicazione dell'esito della selezione alla Provincia e richiesta nominativa di avviamento
- Comunicazione dell'esito della selezione ai partecipanti
- Pagamento di gettoni e missioni a cura del Servizio

7.2 Secondo scenario: lo scarto dei documenti dichiarati inutili al Comune di Trieste

Denominazione del soggetto produttore
Comune di Trieste

Identificativo del procedimento
n. 3421/AAGG

Titolo
Scarto dei documenti dichiarati inutili

Descrizione

L'art. 68 del DPR 445/2000 prescrive per tutte le amministrazioni pubbliche la redazione di un piano di conservazione o, più correttamente, di un massimario di selezione. Tale strumento contiene la descrizione dei tempi di conservazione o di scarto dei documenti dichiarati inutili. Periodicamente, la posizione organizzativa dell'Ufficio Archivio e protocollo, su delega dirigenziale come posizione organizzativa, adotta una determinazione di proposta di scarto. Tale proposta, ai sensi dell'art. 21 del D.Lgs. 42/2004, deve ottenere l'autorizzazione preventiva da parte del Ministero della Cultura, per il tramite della Soprintendenza archivistica del Friuli-Venezia Giulia. La

proposta contiene l'elenco di fascicoli e serie per i quali si propone lo scarto, unitamente a una relazione che descrive la motivazione per la quale si valuta inutile l'ulteriore conservazione.

Normativa nazionale e comunitaria

DPR 28 dicembre 2000, n. 445, art. 68; D.Lgs. 22 gennaio 2004, n. 42, art. 21.²³

Normativa comunale

Regolamento per la disciplina dei protocolli e degli archivi comunali, D.C. 16 marzo 1995, n. 46, artt. 43-45; Regolamento sul procedimento amministrativo, D.C. 17 novembre 1997, n. 115; Manuale di gestione documentale (2003), § 5.2.

Classificazione

I/6 (Amministrazione / Archivio generale)

Tipologia di fascicolo

Fascicolo per procedimento amministrativo (tipo C)

Sintassi di descrizione del fascicolo

È istruito un fascicolo procedimentale per ogni scarto. La sintassi del fascicolo è la seguente: <anno_apertura> - I/6.<num_fasc>, «Denominazione del fascicolo». La denominazione del fascicolo riporta la denominazione del procedimento, e la descrizione puntuale dell'intervento con riferimento alla unità o alle unità organizzative per le quali si propone e gli anni di riferimento. Ad esempio, supponendo che si tratti del quattordicesimo fascicolo istruito sotto la classe 6, del titolo I, nel 2021, ricaveremo la stringa seguente:

2021-I/6.14 «Scarto dei documenti dichiarati inutili – Ufficio ragioneria – Serie dei mandati e delle reversali dal 2001 al 2010».

Unità organizzativa responsabile

Archivio e protocollo

Area

Innovazione e servizi generali

Responsabile del procedimento

Bianchi Bruno

²³ D.Lgs. 22 gennaio 2004, n. 42, *Codice dei beni culturali e del paesaggio, ai sensi dell'articolo 10 della legge 6 luglio 2002, n. 137*, con particolare riferimento all'art. 21, comma 1, lett. d), sono subordinati all'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, con l'eccezione prevista all'articolo 10, comma 2, lettera c), e delle biblioteche private per le quali sia intervenuta la dichiarazione ai sensi dell'articolo 13". Per gli archivi statali, le Commissioni di sorveglianza sono tuttora regolate dal DPR 8 gennaio 2001, n. 37, *Regolamento di semplificazione dei procedimenti di costituzione e rinnovo delle Commissioni di sorveglianza sugli archivi e per lo scarto dei documenti degli uffici dello Stato* (n. 42, allegato 1, della legge n. 50/1999).

Recapiti del RPA

Piazza dell'Unità d'Italia, 4
34121 Trieste (TS)
Tel: 040/6787555
mail: bianchi.bruno@comune.trieste.it

Area di rischio

Basso

Tempo di esecuzione

30 giorni dalla data di presentazione della richiesta di autorizzazione alla Soprintendenza archivistica del Friuli-Venezia Giulia

Tempo di conservazione

Illimitato

Motivazione (se non a lungo termine)

Nulla

Note

Nulla

Documenti principali del procedimento

- Elenco descrittivo di fascicoli e serie da destinare allo scarto
- Richiesta di autorizzazione allo scarto alla Soprintendenza archivistica
- Autorizzazione (o diniego) della Soprintendenza archivistica del Friuli-Venezia Giulia
- Provvedimento di scarto
- Verbale di distruzione o di cessione a CRI o ONLUS (se autorizzato)
- Invio del verbale alla Soprintendenza archivistica

Riferimenti bibliografici

- Alfier, Alessandro. 2021. *Il sistema di documentazione digitale*. Milano: Editrice Bibliografica.
- Antoniella, Augusto. 1999. "Attualità degli strumenti dell'archivio e del protocollo." In *Titulus 97, verso la creazione di un sistema archivistico universitario nazionale. Atti della 1^a Conferenza organizzativa degli archivi delle università italiane (Padova, 22-23 ottobre 1998)*, 69-76. Padova: CLEUP.
- Bonfiglio-Dosio, Giorgetta, e Valeria Pavone. 2007. *Il piano di classificazione (titolario) per i documenti dei comuni*. Padova: Regione del Veneto.
- Brunetti, Dimitri. 2012. *Norme sabaude per gli archivi dei comuni*. Torino: Centro Studi Piemontese.
- Brunetti, Dimitri. 2016. *L'archivio comunale dall'Unità al 1897. I manuali per i segretari comunali, i modelli di classificazione e la Circolare 17100-1 del 1885*. Roma: Ministero dei beni e delle attività culturali e del turismo.
- Brunetti, Dimitri. 2022. *L'archivio comunale dalla Circolare Astengo al 1915. I manuali e i modelli di classificazione*. Roma: Ministero della Cultura, Direzione Generale Archivi.
- Caringella, Francesco. 2022. *Manuale di diritto amministrativo. Parte generale e parte speciale*. Roma: Dike giuridica editrice.
- Carucci, Paola. 1987. *Il documento contemporaneo. Diplomatica e criteri di edizione*. Roma: Nuova Italia Scientifica.
- Di Marcantonio, Giorgia. 2019. "Il fondo Pavone presso l'Archivio Centrale dello Stato. Prime note." *Parole-chiave* 1-2: 81-92.
- Di Marcantonio, Giorgia. 2022. "E se l'archivio non rispecchia l'istituto? Pavone e il rispecchiamento: analisi di una bozza preliminare." *AIDAinformazioni* 1: 51-68.
- Di Marcantonio, Giorgia. 2023. *Pavone nello specchio delle sue carte. Il fondo documentario presso l'Archivio Centrale dello Stato*. Milano: Editrice Bibliografica.
- La dirigenza degli enti locali alla prova della contrattazione collettiva*. 2022. A cura di Mario Cerbone e Rosario Santucci. Torino: Giappichelli.
- Giacalone, Paolo. 2024. *Il ciclo di vita del documento informatico. Gestione e aspetti normativi*. Milano: Franco Angeli.
- Giorgi, Andrea, e Leonardo Mineo. 2009. "Su alcune modalità di gestione dei fascicoli: serie tipologiche e/o "iper-fascicoli". Note in margine a recenti esperienze trentine." *Archivi IV* (2): 21-34.
- Giuva, Linda. 1999. "Gli strumenti archivistici per la gestione dei documenti: la registrazione di protocollo, la classificazione, i piani di conservazione." *Rassegna degli Archivi di Stato* LIX (1-2-3): 128-39.
- Guercio, Maria. 1992. *Archivistica informatica. I documenti in ambiente digitale*. Roma: Carocci: 155-84.
- Martorano, Annantonia. 2020. *Classificare il potere: dal Prospetto delle materie del 1803 alla gestione documentale delle Prefetture*. Torre del Lago: Civita editoriale.

- Molè, Giorgio. 2007. "Profili di responsabilità della dirigenza pubblica tra legalità e trasparenza, logica di mercato ed etica pubblica. Riflessioni sulla direttiva 1° febbraio 2007 del dipartimento della funzione pubblica." *Il lavoro nelle pubbliche amministrazioni* 2: 559-64.
- Nemore, Francesca. 2023. *Cercando il bandolo della matassa. Teorie, usi e prassi degli strumenti di ricerca archivistici*. Roma: Bulzoni editore.
- Pavone, Claudio. 1970. "Ma è poi tanto pacifico che l'archivio rispecchi l'istituto?." *Rassegna degli Archivi di Stato* XXX (1): 145-9.
- Pavone, Claudio. 2004. "Ma è poi tanto pacifico che l'archivio rispecchi l'istituto?." In *Intorno agli archivi e alle istituzioni. Scritti di Claudio Pavone*. A cura di Isabella Zanni Rosiello, 71-5. Roma: Ministero per i beni e le attività culturali.
- Penzo Doria, Gianni. 2007. "Il fascicolo archivistico: le cinque tipologie e i modelli organizzativi." *Archivi & computer. Automazione e beni culturali* XVII (2-3): 22-49.
- Penzo Doria, Gianni. 2022a. "L'archivio e il principio della funzione esercitata." *Atlanti. Rivista internazionale di teoria e pratica archivistica moderna* 32 (1-2): 87-94.
- Penzo Doria, Gianni. 2022b, "La gestione documentale sotto l'aquila asburgica (1815 e 1817)." *Le carte e la storia*, 1: 133-51.
- Romano, Filippo. 2006. "La responsabilità del pubblico funzionario nel procedimento amministrativo, nel trattamento dei dati e nella tenuta del protocollo informatico." *Comuni d'Italia* 10: 23-37.
- Sandulli, Aldo Maria. 1984. *Manuale di diritto amministrativo*. Napoli: Jovene.
- Spaggiari, Angelo. 2022. *Archivi e istituti dello stato unitario. Guida ai modelli archivistici*, Modena. Modena: Archivio di Stato, Scuola di archivistica, paleografia e diplomatica.
- Tessaro, Tiziano, e Margherita Bertin. 2021. *Delibere, ordinanze, decreti, determinazioni*. Santarcangelo di Romagna: Maggioli.
- Tonetti, Eurigio. 2000. *Minima burocratica. L'organizzazione del lavoro negli uffici del Governo austriaco nel Veneto, con appendice documentaria e atlante diplomatico*. Venezia: Istituto veneto di scienze, lettere ed arti.
- Tonetti, Eurigio. 2003. "Strutture d'archivio e classificazione degli atti negli uffici del Governo austriaco del Veneto (1816-1848). Spunti da una ricerca in corso." *Archivi per la storia* XVI (2): 119-27.
- Valacchi, Federico. 2019. "Ripartire da Pavone. Spunti di archivistica." *Parole-chiave* 1-2: 161-75.
- Virga, Pietro. 2001. *Diritto amministrativo. Atti e ricorsi*. Vol. 2. Milano: Giuffrè.
- Zanni Rosiello, Isabella. 2015. "Del linguaggio burocratico." *Le carte e la storia* XXI (2): 11-24.
- Zanni Rosiello, Isabella. 2019. "Su Claudio Pavone che ha preso molto sul serio il lavoro archivistico." *Parole-chiave* 1-2: 19-34.

Why should I be a librarian?*

Mauro Guerrini^(a)

a) University of Florence, <https://orcid.org/0000-0002-1941-4575>

Contact: Mauro Guerrini, mauro.guerrini@unifi.it

Received: 29 March 2025; Accepted: 01 April 2025; First Published: 15 May 2025

ABSTRACT

Who is a librarian? The question has been asked many times over time, and the answers have always been different; none of them satisfactory or lasting. The question is still open: the figure of the librarian, in fact, escapes categorical definitions, or is drawn with a discouraging superficiality even by scholars, a sign of a profession lacking widespread social recognition. Yet it is librarians who make libraries and transform them into living organisms. Librarians belong to a professional community recognized by virtue of the specific skills acquired in a training and certified path. Does the figure of the librarian still exist today in an era characterized by disintermediation? Choosing to be a librarian and valorizing the profession within society is a civil commitment, as well as a cultural activity, which represents a strategic challenge for those who wish to work towards building a more equitable and inclusive future.

KEYWORDS

Librarian.

Perché mai dovrei fare il bibliotecario?*

ABSTRACT

Chi è un bibliotecario? La domanda è stata posta numerose volte nel tempo e le risposte sono sempre state diverse e nessuna soddisfacente o duratura. La questione è tuttora aperta: la figura del bibliotecario, infatti, sfugge da definizioni categoriche, o viene disegnata con una superficialità sconcertante anche da letterati, segno di una professione priva di un riconoscimento sociale diffuso. Eppure sono i bibliotecari a fare le biblioteche e a trasformarle in organismi vivi. I bibliotecari appartengono a una comunità professionale riconosciuta in virtù delle competenze specifiche acquisite in un percorso formativo e certificato. La figura del bibliotecario esiste ancora oggi in un'epoca caratterizzata dalla disintermediazione? Scegliere di fare il bibliotecario e valorizzare la professione all'interno della società è un impegno civile, oltreché un'attività culturale, che rappresenta una sfida strategica per chi desidera operare alla costruzione di un futuro più equo e inclusivo.

PAROLE CHIAVE

Bibliotecario.

* Questo contributo è la rielaborazione della Lectio Magistralis “A difesa della conoscenza critica: le responsabilità degli archivisti e dei bibliotecari”, tenuta insieme al Prof. Federico Valacchi a Firenze il 4 marzo 2025.

Una domanda preliminare: la figura del bibliotecario ha ancora senso nell'era digitale e in un'epoca caratterizzata dalla disintermediazione bibliografica? Non è una domanda provocatoria, bensì un interrogativo dalla risposta complessa e non scontata.

Alcuni tra i principali protagonisti della letteratura biblioteconomica internazionale e italiana¹ si distinguono nelle loro opere per un'apertura di visione, problematica e comprensiva. In particolare, Melvil Dewey, nel primo fascicolo di *Library journal* del 1876, *annus mirabilis* della biblioteconomia statunitense, scrive orgogliosamente nel suo articolo intitolato *The profession*:

È finalmente giunto il momento in cui un bibliotecario può, senza supposizioni, parlare della sua professione come professione. [...] I migliori bibliotecari [...] sono] personaggi positivi, aggressivi, in prima fila tra gli educatori delle loro comunità, fianco a fianco con i predicatori e gli insegnanti (Dewey 1876/77, 5).²

Dewey, fin da giovane, ritiene rilevante il possesso di competenze tecniche per svolgere la professione, da apprendere in apposite scuole. La sinergia tra tecnica, vocazione per il servizio e missione educativa della biblioteca è alla base del successo della professione negli Stati Uniti, paese in cui il bibliotecario ha un'elevata considerazione sociale. A metà Novecento, Jesse H. Shera elabora la proposta di un'epistemologia sociale, nella cui complessità trova spazio e coerenza per le tecnologie, la storia delle biblioteche e della professione. Negli anni Trenta, il filosofo Ortega y Gasset invoca tecniche e linguaggi biblioteconomici raffinati, quindi un *di più* di professione consapevole tecnicamente, socialmente e culturalmente. A metà secolo scorso, Ranganathan ritiene che il bibliotecario svolga un servizio sociale rilevante perché funzionale alla diffusione della conoscenza e del sapere, lo spirito fondamentale dell'educazione. Egli considera tre elementi – il libro, il lettore, la biblioteca – come ugualmente importanti in un sistema di autoregolamentazione e di ricerca d'equilibrio. Ranganathan inserisce il bibliotecario nel modello triadico, dialogico, nel contesto dell'interrelazione tra collezione e lettore: è solo l'esistenza contemporanea di questa *triade* – collezione, bibliotecario e lettore – a costituire una biblioteca e a definire il bibliotecario. L'autore, in *L'organizzazione delle biblioteche* del 1963, spiega il concetto con la suggestiva metafora dell'elettromagnete.

160 Analogia dell'elettromagnete

Sembra prevalere nella gente l'idea anacronistica che una biblioteca sia un'istituzione che ha il compito di conservare una collezione di libri; per il pubblico, infatti, una biblioteca è semplicemente una collezione di libri. Tuttavia, il termine 'biblioteca' non si applica a una semplice collezione di libri più di quanto non si applichi a un semplice gruppo di lettori o al semplice insieme del personale. Una collezione di libri diventa una biblioteca se, e soltanto se, esiste del personale che aiuta i lettori a trovare e utilizzare i libri. Un elettromagnete è una barra di ferro dolce avvolto da una spirale di filo elettrico in cui passa l'elettricità. Le proprietà specifiche dell'elettromagnete non risiedono nel ferro, nel filo o

¹ Solo per citarne alcuni: Melvil Dewey, Charles Sustrac, José Ortega y Gasset, S.R. Ranganathan, Jesse H. Shera, Michel Melot, Luigi De Gregori, Francesco Barberi, Luigi Balsamo, Virginia Carini Dainotti, Carlo Revelli, Luigi Crocetti, Alberto Petrucciani, Paolo Traniello.

² Crocetti parla di Dewey come del «nuovo bibliotecario professionista, esperto di gestione e di tecniche, devoto in primo luogo all'abilità e all'organizzazione» (Crocetti 2003, 9).

nell'elettricità presi separatamente. Esse iniziano a esistere – l'elettromagnete stesso esiste – soltanto quando la corrente elettrica passa attraverso il filo avvolto attorno al nucleo. Così è per la biblioteca: inizia a esistere soltanto quando lettori, libri e personale funzionano insieme. Lettori, libri e personale formano la triade di una biblioteca (Ranganathan 2023, 74).

La frase finale del bibliotecario e studioso indiano esprime un equilibrio fra ciascun elemento. La contemporanea interazione tra lettori, libri e personale non è uno stato acquisito, bensì è un processo che sempre si reinventa. Ciò segna il passaggio decisivo che trasforma il bibliotecario da guardiano dei libri a professionista al servizio del lettore in una biblioteca accogliente, in cui i libri sono disposti a scaffale aperto e non più conservati incatenati ai plutei o racchiusi entro scaffalature protette da reti metalliche, una consapevolezza che ha richiesto molto tempo. Scrive Ranganathan:

Fino a quando l'obiettivo principale di una biblioteca fu la conservazione dei libri, tutto quello che si pretese dal suo personale fu che fosse costituito da guardiani capaci di combattere i quattro nemici dei libri: fuoco, acqua, parassiti e uomini. Non era strano che un posto di lavoro in biblioteca rappresentasse il rifugio possibile per le persone incapaci di fare altri lavori. Ci volle davvero molto tempo perché si comprendesse che era necessario un bibliotecario professionale (Ranganathan 2010, 40-41).

È ciò che afferma con garbo Luigi De Gregori nel saggio *Il bibliotecario* del 1947:

Chi concepisce ancora la figura del bibliotecario secondo l'abusato cliché d'un occhialuto vecchio in papalina assorto da mattina a sera nella cura della conservazione e dell'incremento della sua Biblioteca, costui non conosce neppure di vista una biblioteca moderna, la sua funzione, i suoi bisogni, i suoi fini (De Gregori 1947, 5).

In passato, infatti, la figura di bibliotecario – intesa come professionista laureato, prevalentemente in ruolo presso pubbliche amministrazioni – è stata frequentemente vissuta come un ripiego dall'insegnamento liceale o accademico piuttosto che come vocazione o missione. Ne è testimonianza la polemica fra Francesco Barberi e Maria Ortiz, sostenitrice, a metà Novecento, di una differenziazione qualitativa del lavoro all'interno di una biblioteca: «lavori umili» vs compiti di «più alta sfera». Barberi risponde che dai primi ci si riscatta approfondendoli umanisticamente e non già eseguendoli più o meno meccanicamente e cercando soddisfazione nei secondi (Barberi 1984). Il bibliotecario romano è tra i primi, nell'Italia del Secondo dopoguerra, a porre la figura del bibliotecario nei termini di un professionista specializzato a servizio del lettore; in lui il bibliotecario supera la connotazione sia di un personaggio erudito e studioso di formazione umanistica, sia di un ruolo lavorativo come una *sinecura* di retaggio ottocentesco, che consente di sbarcare il lunario e di dedicarsi a studi e interessi personali. Il bibliotecario ideale si caratterizza per due componenti che costituiscono le facce della stessa medaglia: la competenza istituzionale e il servizio al lettore; Barberi, infatti, auspica che

le sorti delle biblioteche e della stessa professione siano affidate, oltretutto allo zelo del bibliotecario nel disimpegnare il suo ufficio quotidiano, anche al suo rendimento scientifico: due attività, d'altronde, difficilmente separabili, data la singolare natura del nostro lavoro [...]; cosicché può dirsi che trattasi piuttosto di due aspetti, integranti e soccorrenti a vicenda, di una attività unica (Barberi 1967, 311).

All'inizio della sua carriera, nel 1934, Barberi scrive una nota nel suo diario *Schede di un bibliotecario (1933-1975)*, edito nel 1984, distinguendo fra due interpretazioni del ruolo interne alla professione:

Ci sono due diversi modi di fare il bibliotecario (ma non solo il bibliotecario): l'uno è di chi si limita a eseguire, più o meno diligentemente, il lavoro assegnatogli senza preoccuparsi del resto; l'altro di chi non resiste al bisogno di allargare lo sguardo intorno a sé per vedere come vanno le cose in biblioteca, magari in più biblioteche; di approfondirle, discuterne coi colleghi, intervenire (Barberi 1967, 13).

Sempre in questo libro, denso di notizie e di riflessioni, descrive la sua interpretazione di bibliotecario:

Il bibliotecario è geloso custode dell'eredità letteraria del passato e fautore del suo accrescimento; mette in valore, umanisticamente, i pensieri espressi e facilita il nascere di nuovi. La conservazione dei libri-oggetti può ostacolare il più largo sfruttamento, e quindi la piena valorizzazione, del loro contenuto di pensiero, e viceversa: il bibliotecario è in questi casi vittima della sua difficile missione, apparentemente contraddittoria. A lui, anello di una tradizione antica quanto la civiltà scritta, ancora oggi, come al monaco medievale, molti autori e testi debbono la propria sopravvivenza (77).

In questo senso, il bibliotecario è un *intellettuale tecnico* (Prandstraller 1972), come Barberi lo definisce in più occasioni, fra le quali in *Intellettuali bibliotecari*, edito nel 1981 (Barberi 1981, [255]-284, [295]-306).

Chi è un bibliotecario? Una domanda che si ripropone nel tempo

Dunque: *chi è un bibliotecario?* La domanda è stata posta numerose volte nel tempo, a cominciare da Charles Sustrac oltre cent'anni fa (Revelli 2007, 68). Le risposte sono sempre state diverse e nessuna soddisfacente o duratura. La questione è tuttora aperta:³ la figura del bibliotecario, infatti, sfugge da definizioni categoriche, sguscia come un pesce dalle mani o viene disegnata con una superficialità sconcertante anche da letterati, segno di una professione priva di un riconoscimento sociale diffuso. Pertanto, come osserva Paolo Traniello,

cosa debba precisamente fare un bibliotecario non è cosa determinabile a priori, ma dipende da molte variabili, relative al contesto sociale in cui opera e al tipo di biblioteca, tenendo conto che la fisionomia degli istituti, i loro compiti e le relative competenze si pongono oggi in un quadro tipologico talmente vario da legittimare la domanda se possa esistere veramente una sola istituzione sociale denominata biblioteca (Traniello 2017, XXII).

³ La domanda – ricorda Carlo Revelli – era già stata posta in passato da Anne Kupiec, *Qu'est-ce qu'un(e) bibliothécaire?* «Cos'è un bibliotecario? Perché porre nuovamente questa domanda? È stato formulato molte volte e sono state date molteplici risposte. Non c'è alcuna pretesa di fornire, oggi, una risposta definitiva e perentoria. Si tratta piuttosto di ipotizzare che la quasi permanenza della questione deriverebbe da una difficoltà nel definire lo statuto della biblioteca e, di conseguenza, di coloro che ad essa sono legati: i bibliotecari. Questa difficoltà non deriverebbe dalla strategia di rottura da tempo adottata dai bibliotecari dinamici e dalla tentazione della modernità a cui sono sensibili?» (Kupiec 2003, 5).

Biblioteche e bibliotecari, infatti, con le parole di Luigi Balsamo, presentate al convegno *Biblioteche e sviluppo culturale* del 1977,

non solo si condizionano a vicenda in maniera determinante quanto a funzionalità dei servizi, ma per certi aspetti giungono addirittura ad identificarsi; cosicché parlare di bibliotecari è la stessa cosa che parlare di biblioteche e viceversa (Balsamo 1978, 79).

Giulio Einaudi, fondatore della biblioteca di Dogliani (Cuneo), aperta nel 1963 quale donazione al Comune in ricordo del padre Luigi, afferma che dipende in larga misura dalla preparazione, sensibilità ed efficienza del bibliotecario

se le biblioteche nuove e vecchie saranno luoghi vivi di cultura o morti depositi di carta stampata.⁴

Concorda Virginia Carini Dainotti, in *La Biblioteca pubblica in Italia tra cronaca e storia* (1947-1967) del 1969, secondo cui:

sono i bibliotecari a fare le biblioteche (Carini Dainotti 1969, 501).

Il bibliotecario: un caleidoscopio di attività

Maria Stella Rasetti, in *Bibliotecario, il mestiere più bello del mondo* del 2014, scrive:

Se nell'immaginario collettivo i bibliotecari si somigliano un po' tutti, la realtà quotidiana ci restituisce un vero e proprio caleidoscopio di differenze (Rasetti 2014, 8).

Questo caleidoscopio può sopraffare la definizione stessa di bibliotecario? Si domanda Alberto Salarelli in *Il bibliotecario sopraffatto* del 2005:

Dunque, come prima causa di sopraffazione, c'è l'eccesso di informazione: 'la messe è tanta ma gli operai sono pochi'. Seconda causa di sopraffazione: i ruoli che oggi un bibliotecario deve rivestire per essere considerato tale. Qualche etichetta? (Salarelli 2005).

Per Giovanni Solimine, nella relazione *Il bibliotecario: una professione, molte fisionomie*, presentata al convegno di *Biblioteche oggi*, di Châtillon nel 1987,

se il bibliotecario va considerato come un professionista responsabile della struttura e dei servizi che gestisce, e se la sua professionalità è proporzionale alla complessità della struttura in cui opera, alla qualità dei servizi che offre, ai risultati che riesce a raggiungere; e se la biblioteca ha oggi la necessità di specializzarsi, indirizzandosi a precise fasce di utenza o a chiari ambiti disciplinari, cercando di saldarsi fortemente ai propri fini istituzionali, rispondendo ai bisogni dell'ente di cui è emanazione o del territorio sul quale insiste; se in tutto questo possiamo riconoscere una soluzione al quesito su quale sia oggi l'identità del bibliotecario e della biblioteca, allora il discorso deve necessariamente spostarsi

⁴ Archivio Einaudi, Segreteria editoriale, Corrispondenza, Corrispondenza con autori e collaboratori italiani, m. 74.2, fasc. 1118.4. Documento non datato, in (Faggiolani 2020, 164)

sulla definizione delle diverse fisionomie del bibliotecario, come sono e come dovrebbero essere, e sulla possibilità o meno di ritrovare a monte di queste fisionomie una matrice unitaria (Solimine 1988, 87).

Una possibilità che esiste, come afferma Alberto Petrucciani in *Biblioteca e professione bibliotecaria: quale prospettiva per oggi e per domani?* del 2007: fra tutte le varie fisionomie vi è una

unitarietà sostanziale della professione bibliotecaria (Petrucciani 2007, 207-212).

Conoscenze diversificate, afferma ancora Petrucciani in *Discipline del libro, identità professionale, ricerca scientifica* del 1991, le cui «basi e le sue articolazioni fondamentali sono unitarie ed omogenee»:

Che l'attività del bibliotecario richieda, per essere efficace, un bagaglio vasto e profondo di conoscenze a me pare ovvio, e la realtà delle cose ce ne fornisce continue conferme tanto in positivo quanto in negativo. Altrettanto ovvio è che questo sapere non avrà lo stesso contenuto per tutte le figure e le situazioni professionali: questo è quanto avviene, per esempio, anche per un medico, un giurista, un ingegnere. Forse meno ovvio, per molti, è che le basi di questo sapere e le sue articolazioni fondamentali sono unitarie ed omogenee, mentre le differenze, che pure esistono e sono importanti, costituiscono piuttosto delle specializzazioni o sono semplicemente relative ai diversi ambienti, o organizzazioni, nell'ambito dei quali si va ad operare (Petrucciani 1991, 36).

Dunque, chi è il bibliotecario?

Carlo Revelli, in *Compiti e caratteristiche del bibliotecario* del 1995 (Revelli 1995, 51), menziona *Le bibliothèques françaises au XX^e siècle* di Martine Poulain.

L'intero secolo è scandito da questi dibattiti e preoccupazioni: il bibliotecario è uno studioso? Un conservatore? Un tecnico? Un ingegnere? Un mediatore? Un facilitatore? Un manager? Una "specie in via di estinzione"? Ciascuna di queste denominazioni ha avuto la sua accoglienza nel corso del secolo. Tutte sono portatrici di un'immagine che ciascuno cerca di promuovere e, quindi, riguardano sia questioni corporative sia strettamente professionali. Tutte, infatti, evidenziano concezioni diverse (e perché non contrastanti?) dell'eccellenza professionale e sentono generalmente il bisogno di fare riferimento ad altre professioni, vicine e concorrenti, professionalmente o simbolicamente: documentalisti, curatori di musei, informatici, insegnanti, ecc. Tutto questo parlare di competenza, di distinzione, di differenza, si palesano soprattutto nei dibattiti sull'attuazione dei percorsi di formazione professionale (Poulain 1994, 6).

Michel Melot, in *La saggezza del bibliotecario* del 2004, presenta una riflessione affascinante. Il bibliotecario è, in genere, un uomo riservato, poiché la biblioteca è il luogo di quello che definisce un anonimato pubblico; il bibliotecario ama i libri come il marinaio ama il mare. Un marinaio non è necessariamente un bravo nuotatore, ma sa navigare e sa che non è a nuoto che si va più lontano.

L'oceano del sapere che inebria tutti i sapienti rende umile il bibliotecario (Melot 2005, 5).

La saggezza è la capacità con cui il bibliotecario gestisce le proprie competenze a servizio del variegato pubblico dei lettori.

Luigi Crocetti, in *Bibliothecarius technologicus* del 1998, rinvia a una metafora suggestiva, *un cartografo dell'informazione*, esaltando lo *stile* rigoroso del bibliotecario, legato alla precisione e congruità dei dati forniti (Crocetti 1988, 223-228),⁵ unico metodo per offrire un servizio efficace.

Chi era e chi è il bibliotecario? A consultare dizionari ed enciclopedie c'è da rimanere delusi. In pratica vi dicono che è chi lavora in una biblioteca. Avremmo bisogno [...] di qualcosa di più e meglio. Non so se di più e meglio sia dire che è uno che lavorava e lavora nel mondo dell'informazione; ricorriamo ancora una volta a una metafora chiamandolo un cartografo dell'informazione. Con la peculiarità che la sua cartografia si applica all'informazione rappresa nei documenti: un tempo limitati ai libri e ad altra roba manoscritta o stampata e ora senza più limiti, estesa a qualsiasi documento recante informazioni. Ma la sua funzione rimane di cartografo. Come nella cartografia antica e moderna, i suoi dati continuano a pretendere l'esattezza e a voler essere offerti in forma congrua; i nomi e le configurazioni cercate debbono essere cercabili e trovabili secondo sistemi predisposti d'identificazione e di localizzazione (Crocetti 1998, 6-7).

Il bibliotecario come parte di una comunità professionale

Giovanni Di Domenico ritiene che i bibliotecari appartengano a una comunità professionale che, come le altre del lavoro intellettuale, possiede un'idea di sé che la differenzia dalle categorie solo impiegate; pertanto, più che un mestiere è (o dovrebbe essere) un'attività istituzionalmente riconosciuta in virtù delle competenze specifiche acquisite in un percorso formativo e certificato.

I bibliotecari fanno comunità professionale, interagiscono fra loro, si rapportano con le altre organizzazioni professionali e con le istituzioni che governano la politica bibliotecaria (Di Domenico 2009, 29-30).

Michael Gorman disegna alcune caratteristiche per la definizione generale del profilo professionale in *I nostri valori, rivisti*, edito in italiano nel 2018 (Gorman 2018, 12).⁶ Egli basa i suoi principi sui valori fondamentali liberali, democratici e umanistici che caratterizzano la cultura europea e americana contemporanea. È fortemente preoccupato del futuro della professione, sempre più attratta dalle sirene di un approccio superficiale ai problemi biblioteconomici e catalografici. Sostiene che i valori fondamentali dei bibliotecari agevoleranno la crescita personale, delle istituzioni in cui lavorano e dell'intera società. La sua è una sintesi appassionata, colta, viva, di una necessità di armonizzare tanti aspetti della tradizione bibliotecaria; è un'*advocacy* della professione. L'autore elenca nove valori (otto nella prima edizione del libro, edito sempre dall'ALA del 2000 (Gorman 2022), aggiunto l'ultimo nella seconda del 2015): capacità di gestione, servizio, libertà intellettuale, razionalità, alfabetismo e apprendimento, equità d'accesso alla conoscenza e all'informazione, privacy, democrazia, bene superiore. Non si tratta di enunciati

⁵ «Esattezza e precisione dovrebbero essere lo stile della biblioteca cioè, ancora una volta, la sua cultura: la cultura altamente formale di un'istituzione che opera in una disciplina pervasiva come, adoperando una parola molto amata dalla lingua egemone della biblioteconomia, potremmo chiamarla. E chi scrive cartelli, accudisce a giardini, confeziona frontespizi e compila indici si convinca di non essere la prima tigre» (Crocetti 1988, 228).

⁶ Si veda, inoltre, (Gorman 2004).

astratti, bensì di valori-azione, oggetto di un esercizio quotidiano, in cui si realizza quella vocazione all'apertura, al confronto che è propria di ogni vera biblioteca e di ogni vero bibliotecario contemporaneo. Il concetto di valore è strettamente legato a quello di etica, al modo di agire del bibliotecario; avere valori significa essere schierati e Gorman non perde occasione per affermare il sistema politico della democrazia.

È quanto sottolinea l'*IFLA Code of ethics for librarians and other information workers* del 2012:

Il mestiere del bibliotecario si basa essenzialmente su un'attività etica, su un trattamento professionale delle informazioni caratterizzato dal rispetto di valori fondamentali. La necessità di scambiarsi idee e informazioni assume un'importanza sempre più rilevante in una società che da secoli si fa continuamente più complessa. Questa evoluzione forma la base per le biblioteche e per la prassi professionale bibliotecaria.⁷

La transizione bibliografica che stiamo vivendo dovuta alle esperienze e alle riflessioni bibliotecarie va in direzioni molto diverse fra loro. Negli Stati Uniti e in Italia, per esempio, gli studi più avanzati di biblioteconomia sociale cercano di definire un modello di biblioteca inteso come *hub di comunità*, in quanto struttura aperta che dialoga con il territorio, il cui fine è migliorare le competenze dei cittadini e far comprendere loro le modalità migliori per l'accesso all'informazione e per valutare e scegliere testi di qualità. In questo contesto estremamente variegato, la biblioteconomia e la figura del bibliotecario mantengono la loro identità e sono vitali proprio perché stanno evolvendosi con lo sviluppo dei bisogni informativi, delle nuove opportunità tecnologiche e della ricerca. Su questo approccio pragmatico s'inserisce la riflessione teorica che contribuisce, e spesso non poco, a raffinare la portata dell'innovazione e a definire nuovi modelli concettuali e organizzativi.

Oggi si fa riferimento a un bibliotecario che opera e in taluni casi compete con altri protagonisti che insistono nello stesso alveo. Da tempo, infatti, l'informazione bibliografica non viene più trasmessa esclusivamente dai libri e dalla parola oggettivata sulla carta; ciò rende obsoleta la definizione di bibliotecario come mediatore tra la raccolta e i lettori. Essa è superata da una connotazione professionale mirata a realizzare servizi che favoriscano l'accesso all'informazione e alle risorse bibliografiche, nonché alla qualità dell'interazione e integrazione con il contesto territoriale e culturale di riferimento. La norma UNI 11535, dedicata alla *Figura professionale del bibliotecario*, parla, infatti, di «contesti organizzativi» entro cui il bibliotecario opera che contribuiscono a definire la sua figura, non più legata esplicitamente alle biblioteche e al libro. Si può essere bibliotecari anche lavorando per un'agenzia bibliografica (per esempio, Casalini Libri), per una casa editrice o per un'organizzazione la cui finalità è l'*advocacy* della biblioteca e della professione (per esempio, IFLA, AIB) (Manoni et al. 2015; Ponzani, Manoni, e Mazzola Merola 2023). Ciò è un segno di vitalità e contribuisce a rendere il bibliotecario, come tutti i professionisti, una figura in costante divenire, la cui definizione si evolve nel tempo.

Nonostante ciò, la figura del bibliotecario è, ancora oggi, spesso narrata tramite immagini romantiche o stereotipate: il custode del sapere, il silenzioso ordinatore di libri, il promotore della lettura

⁷ https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/publications/IFLA%20Code%20of%20Ethics%20-%20Long_0.pdf.

ecc.; in molte rappresentazioni, dall'Antichità all'Ottocento e oltre, il bibliotecario è l'erudito. Dietro queste visioni limitate e, talvolta, distorte, si cela una professione sconosciuta quanto necessaria. Molte definizioni apparse negli anni hanno retto lo spazio di un mattino, manifestazioni di una moda effimera. Poche si sono affermate nel tempo e sono quelle che testimoniano ed esprimono i cambiamenti epocali, com'è stato, per esempio, il passaggio dal paradigma convenzionale a quello digitale nell'organizzazione dei servizi interni ed esterni della biblioteca. La definizione di bibliotecario è sempre problematica e aperta, mai stabile, perché è perennemente in divenire, legata a una complessità di fattori, risultato di tante componenti; varia in relazione ai differenti ambiti culturali, territoriali e organizzativi.

Il cambio di paradigma nell'era digitale: la disintermediazione e la metadatazione

Nell'era digitale si assiste alla disintermediazione nei processi di trasmissione della conoscenza registrata, con la ridefinizione dei confini e degli ambiti delle biblioteche, delle competenze del bibliotecario e della terminologia utilizzata. È un cambio di paradigma.

Il tema della disintermediazione diminuisce il ruolo del bibliotecario? È innegabile che il nuovo contesto culturale e tecnologico abbia spinto, e continui a spingere, verso un ripensamento generalizzato delle modalità di reperimento o scoperta delle informazioni, e, di conseguenza, verso nuovi strumenti, quali, per esempio, i motori di ricerca in rete o i *web-scale discovery services*. In biblioteca, la compresenza di risorse analogiche e digitali ha portato a un'evoluzione graduale ma costante dei servizi a esse legati; tuttavia, il problema della disintermediazione, ovvero il poter fare a meno dei cataloghi e dei repertori e di coloro che li creano, li organizzano e li gestiscono, è solo apparente. Il bibliotecario, accettando la sfida posta dal web e dal sempre più ampio universo bibliografico, continua a svolgere il ruolo che tradizionalmente ha rivestito, ossia quello di mediatore tra le collezioni e il lettore. Ciò che cambia è il contesto, che è diventato quello digitale, della rete e, di conseguenza si è trasformato anche il ruolo del bibliotecario, che, accanto all'*information literacy*, dovrà guidare l'utente nella *media literacy*, ossia introdurlo al linguaggio tipico dell'era digitale. Il bibliotecario, pertanto, non è destituito dal suo ruolo guida. Molti lettori, infatti, – seppure sempre più indipendenti –, non sanno usare autonomamente e consapevolmente gli strumenti di ricerca oggi a disposizione, costantemente più sofisticati e complessi. Ci sarà, pertanto, ancora bisogno di un mediatore che dovrà essere capace di connettere, fornire collegamenti, *link*, far capire l'architettura dell'informazione culturale, spiegare i percorsi cognitivi, la cultura del web, la cultura digitale; occorre, cioè, ribadire la necessità dell'alfabetizzazione, della *media literacy* per salvaguardare la stessa democrazia delle istituzioni.

La metadatazione (in inglese: *metadata creation*) è parte della realtà bibliotecaria dai primi decenni del secolo in Europa, negli Stati Uniti e in altre parti del mondo. La descrizione e l'indicizzazione compiute con l'ausilio della macchina non devono far pensare a una deresponsabilizzazione o a una modalità aggiornata di mattanza del catalogatore. La tecnologia è un aiuto nel proprio lavoro, non sostituisce la competenza e il ruolo del bibliotecario; sono gli agenti umani che inquadrano il dato nel giusto contesto e, tramite standard e linguaggi codificati, esplicitano attributi e relazioni che permettono alla macchina di 'comprendere' il dato. L'ausilio del robot è forse da interpretare come un depauperamento della professionalità del chirurgo? Non è, invece, un'occasione di cre-

scita professionale? E soprattutto non è uno strumento meno invasivo di altre metodiche per intervenire all'interno del corpo umano per asportare una parte malata? La metadattazione, pertanto, corrobora il rigore dell'analisi bibliografica (analisi formale, analisi letteraria, analisi concettuale) necessaria per ottenere una descrizione di qualità, permettendo così una migliore identificazione e, di conseguenza, un miglior reperimento delle risorse (Guerrini 2022a; 2022b).

È nato ormai da anni il bibliotecario dei dati, specializzato nella gestione dei metadati, seppure al momento, sia una figura “di nicchia” presente in alcune realtà accademiche e di ricerca (ma scarsamente in Italia). Uno tra i suoi compiti è istruire gli utenti sulla natura e il reperimento consapevole dei dati (Rosati 2019). L'alfabetizzazione applicata ai dati risulta fondamentale poiché, mentre l'informazione arriva al lettore con un contesto (il libro, la rivista, la pagina web), i dati sono destrutturati e gli utenti hanno bisogno di comprenderli, contestualizzarli, individuarne limiti e potenzialità (Hirsh 2024).⁸

La figura del bibliotecario, dunque, esiste ancora oggi in un'epoca caratterizzata dalla disintermediazione?

Michael Gorman chiude *La biblioteca come valore* con un *Epilogo* denso di speranza:

I nostri tempi, come i passati, sono interessanti e molti bibliotecari si sentono avversati dal cambiamento, alienati da un aspetto o dall'altro delle biblioteche moderne, e minacciati dal drago a due teste della riduzione delle risorse e delle richieste della tecnologia. C'è, ne sono convinto, solo un modo per recuperare la felicità nel nostro lavoro e l'equilibrio nelle nostre vite ed è guardare a dove siamo adesso e dove eravamo in passato con chiarezza di visuale e di agire sulla base della cognizione che quella chiarezza porta. Comprendere i processi in corso e le forze che plasmano le nostre vite lavorative significa prendere il controllo di entrambe (Gorman 2004, 201-2).

C'è, dunque, un futuro per il bibliotecario? È la domanda che si pone la redazione del *Bollettino AIB* nel 2011:

Ma oggi, che percezione abbiamo del futuro che ci attende? Il domani che ci apprestiamo a vivere sarà quello di una biblioteca non solo senza pareti ma anche senza bibliotecari? (*Il futuro in biblioteca* 2011, 313)

In passato bibliotecari illuminati e guru improvvisati hanno prospettato scenari che poi si sono rivelati solo elaborazioni della loro illusione. Impossibile ipotizzare cosa sarà il bibliotecario o come si chiamerà in futuro. Si possono percepire unicamente tendenze. In alcuni Paesi, come l'Italia, ciò che altrove è realtà, come l'ambiente della biblioteconomia dei dati, è letto come futuro nella stragrande maggioranza dei casi ed è concretezza e quotidianità solo in situazioni molto avanzate, ovvero situazioni in linea con il palcoscenico internazionale.

Sul sito dell'AIB appare una definizione del 3 giugno 2024 quanto mai pertinente e sintetica; il primo paragrafo si chiude con uno sguardo verso la preparazione dei bibliotecari del futuro.

⁸ Suggestivi i brevi contributi apparsi in (Hirsh 2024); i partecipanti – tra cui R. David Lankes, Kelvin Watson, Annie Norman, Miguel Figueroa e Nicole Cooke e altri 25 dirigenti di biblioteca – ipotizzano il futuro della biblioteca e della professione bibliotecaria.

La professione del bibliotecario è sempre stata caratterizzata da competenze diversificate tese a rispondere alle diverse esigenze del pubblico e dei cittadini e alle molteplici attività svolte nelle biblioteche. La trasformazione e l'evoluzione delle competenze richieste nelle biblioteche sono state guidate in parte da fattori endogeni, corrispondenti al progressivo affinamento delle tecniche biblioteconomiche, ma soprattutto dallo sviluppo delle tecnologie disponibili per la gestione delle informazioni. Questo fattore è stato particolarmente rilevante negli ultimi 60 anni, durante i quali la progressiva adozione delle ICT per la gestione delle informazioni bibliografiche e delle operazioni amministrative ha comportato una profonda modifica delle competenze di gran parte dei bibliotecari e lo sviluppo, all'interno della professione del bibliotecario, di nuove figure professionali specializzate su temi quali la *digital library*, la metadattazione e conservazione di documenti digitali, la raccolta e elaborazione dei dati di utilizzo delle risorse elettroniche, l'educazione degli utenti allo sviluppo di *digital e information literacy*, i servizi a distanza, le varie modalità di lettura e così via.⁹

Ciò che è certo è che la professione bibliotecaria rappresenta una componente necessaria dell'infrastruttura culturale di ogni Paese, perché, con le parole di Michael Gorman

tutti i valori e le idee che dominano la teoria e la pratica bibliotecaria appartengono anche alla democrazia: la libertà intellettuale, il bene comune, lo spirito di servizio verso gli altri, la trasmissione delle conoscenze umane alle generazioni future, il libero accesso alle informazioni e alla conoscenza, la non discriminazione e così via. Un bibliotecario che non sia un democratico è del tutto inconcepibile; le biblioteche sono cresciute e fiorite sul terreno della democrazia e il nostro destino vi è indistricabilmente connesso (Gorman 2018, 167-8).

Commenta Paolo Traniello:

Ciò sul presupposto che la democrazia consista essenzialmente nel fornire a tutti i cittadini uguali possibilità in tutti i campi, quindi anche a livello culturale, possibilità che spetterà poi agli stessi individui far fruttare come meglio avranno saputo (teoria delle 'pari opportunità') (Traniello 2014, 144).¹⁰

Maria Stella Rasetti, evidenziando la finalità di *Bibliopride*, la Giornata nazionale delle biblioteche e dell'orgoglio bibliotecario, promossa dall'Associazione italiana biblioteche, tenuta per la prima volta a Napoli il 13 ottobre 2012, scrive parole efficaci: l'iniziativa si rivolge ai cittadini, alle istituzioni, ai politici, alle persone che amano le biblioteche

per offrire loro l'opportunità di scoprire che le biblioteche rappresentano una risorsa straordinaria per lo sviluppo civile, sociale ed economico del Paese: uno dei pilastri, assieme alla scuola e alla cultura, su cui fondare non solo una società più equa e una cittadinanza più consapevole, ma anche un'economia più competitiva e perciò una democrazia più solida.¹¹

Una risorsa così importante che, per Peter Lor, le biblioteche e i bibliotecari possono svolgere un ruolo anche a favore della pace:

⁹ Dal sito AIB, <https://www.aib.it/notizie/mercato-lavoro-professioni-bibliotecarie/>, 3 giugno 2024.

¹⁰ Si veda: (Guerrini e Neri 2024; Rasetti 2014, 26).

¹¹ Dal sito AIB, <https://www.aib.it/notizie/mercato-lavoro-professioni-bibliotecarie/>, 3 giugno 2024.

Di per sé le biblioteche e i bibliotecari non sono in grado di assicurare una maggiore uguaglianza, ridurre la povertà, proteggere l'ambiente, promuovere i diritti umani, generare mutuo rispetto tra persone con punti di vista e bagagli culturali diversi. Tuttavia è possibile dare un contributo – e questo contributo, ne sono convinto, può essere significativo. In questo modo i bibliotecari impegnati e creativi contribuiscono alla pace nel mondo (Lor 2024, 25).¹²

Aspetti negativi: il mancato riconoscimento professionale e la crescente burocratizzazione

Tutto bene? Sono stati privilegiati gli aspetti positivi, ma non mancano (e sono parecchi) quelli negativi, dal mancato riconoscimento professionale alla crescente burocratizzazione. Per esempio, Sven Kuttner, in *Beruf. Die Wieder-Buzás-Kontroverse 1959 bis 1962* del 2009 (Kuttner 2009, 384-98), ricostruisce dettagliatamente, nella prima parte del saggio, l'acceso dibattito scaturito da due articoli di Joachim Wieder e Ladislaus Buzás (Wieder 1959, 132-65; Buzás 1960, 81-104) circa la caratterizzazione del bibliotecario: “studioso” o “di servizio”? Il pressante ritmo di lavoro tipico dell'epoca moderna è causa di preoccupazione per il declino intellettuale e sociale della professione bibliotecaria. Infatti, la massificazione, la meccanizzazione delle procedure bibliografiche e gestionali e i crescenti compiti amministrativi compromettono la natura scientifica della professione, che si è tradizionalmente distinta dallo studio delle collezioni, e trasformano il bibliotecario in un burocrate soggiogato dalla *routine*.

In Italia, la burocratizzazione inizia dalle modalità di reclutamento dei bibliotecari, una questione complessa e annosa. Una domanda preliminare: i titoli che vengono richiesti per iscriversi a certi concorsi per funzionario bibliotecario rappresentano già una selezione? Qualcuno si è posto il problema di quanto costa la gestione di un concorso? È insufficiente la preparazione di un dottore di ricerca e di chi ha conseguito una specializzazione o un master biennale?

Conclusione

La conclusione della riflessione riprende le parole di Raffaele de Magistris:

Il bibliotecario potrà svolgere una funzione democratica di orientamento e di organizzazione dell'informazione e della conoscenza più efficacemente se riuscirà a introdursi, e a introdurre i metodi e valori professionali (trasparenza, rigore, equidistanza, pluralismo ecc.) nei meccanismi di funzionamento di quella grande biblioteca virtuale, di quel gran catalogo, che servirà a presentare e descrivere la realtà e di cui oggi assistiamo solo ai prodromi.¹³

Le biblioteche sono componenti essenziali di una società democratica; esse, che vivono di cultura, rappresentano una risorsa per l'intera organizzazione sociale, culturale ed economica di un Paese,

¹² Non è stato sempre così; Petrucciani (2003) ricorda che un alto numero di bibliotecari o di coloro che diverranno bibliotecari era favorevole all'interventismo e si arruolava volontariamente per partecipare alla Prima Guerra mondiale.

¹³ Raffaele De Magistris, email del 28 gennaio 2025.

e il bibliotecario, con la sua attività, partecipa all'affermazione della democrazia, considerata alla base di tutti i valori della teoria e della pratica bibliotecaria. Scrive Michael Gorman:

Bibliotecari e archivisti (che ritengo membri della stessa chiesa, anche se spesso coinvolti in uno scisma!) hanno un ruolo unico nel preservare e trasmettere i documenti dell'umanità per le generazioni future. Non uso la parola 'unico' casualmente: molti dei nostri compiti e valori sono condivisi da altri gruppi con interessi diversi, ma siamo i soli a dedicarci alla conservazione della conoscenza e dell'informazione: la conoscenza registrata (Gorman 2018, 68).

Pertanto, scegliere di fare il bibliotecario e valorizzare la professione all'interno della società è un impegno civile, oltretutto un'attività culturale, che rappresenta una sfida strategica per chi desidera operare alla costruzione di un futuro più equo e inclusivo.

Riferimenti bibliografici**

Archivio Einaudi, Segreteria editoriale, Corrispondenza, Corrispondenza con autori e collaboratori italiani, m. 74.2, fasc. 1118.4.

Balsamo, Luigi. 1978. *Situazione e formazione del personale delle biblioteche*. In *Biblioteche e sviluppo culturale: atti del Convegno, Milano, 3-5 marzo 1977*, a cura del Comune di Milano, materiali raccolti da Anna Maria Rossato. Roma: Editori Riuniti.

Balsamo, Luigi. 2024. *Scritti di biblioteconomia*. A cura e con un saggio introduttivo di Alberto Salarelli, prefazione di Mauro Guerrini. Firenze: Firenze University Press.

Barberi, Francesco. 1967. "Gli studi del bibliotecario". In Francesco Barberi, *Biblioteca e bibliotecario*, 311-6. [Bologna]: Cappelli.

Barberi, Francesco. 1981. *Biblioteche in Italia. Saggi e conversazioni*. Firenze: Giunta regionale toscana; La nuova Italia.

Barberi, Francesco. 1984. *Schede di un bibliotecario (1933-1975)*. Roma: Associazione italiana biblioteche.

Battaglia, Maddalena. 2019. "La visione della biblioteca attraverso i bibliotecari: metodi narrativi per le organizzazioni." In *Conoscere gli utenti per comunicare la biblioteca*, a cura di Chiara Faggiani, 253-91. Milano: Editrice Bibliografica.

Buzás, Ladislaus. 1960. "Berufssorgen des wissenschaftlichen Bibliothekars. Ein Diskussionsbeitrag." *Libri* 10: 81-104.

Carini Dainotti, Virginia. 1969. *La biblioteca pubblica in Italia tra cronaca e storia (1947-1967). Scritti, discorsi, documenti*. Firenze: Olschki.

Crocetti, Luigi. 1987. *Lo stile della biblioteca*. In *La cultura della biblioteca. Gli strumenti, i luoghi, le tendenze: atti del Convegno di Biblioteche oggi*, realizzato in collaborazione con Comunità montana del Marmore, Comune di Châtillon, Association des bibliothécaires valdôtains, AIB Sezione Piemonte, Châtillon, 22-24 maggio 1987, a cura di Massimo Belotti, 223-28. Milano: Editrice Bibliografica.

Crocetti, Luigi. 1998. "Bibliothecarius technologicus. Rivoluzione quantitativa o nascita di una nuova specie?" *Biblioteche oggi* 16 (4): 6-7.

Crocetti, Luigi. 2003. "Un "oggi" lungo vent'anni." *Biblioteche oggi* 21 (10): 9.

De Gregori, Luigi. 1947. "Il bibliotecario." *Rivista delle biblioteche*, 1 (1): 5.

De Gregori, Luigi. 1947. *Il bibliotecario*. Roma: [s.n.].

Desideri, Laura, e Alberto Petrucciani, a c. di. 2014. *Le biblioteche di Luigi Crocetti: saggi, recensioni, paperoles (1963-2007)*. Presentazione di Stefano Parise. Roma: Associazione italiana biblioteche.

Dewey, Melvil. 1876/77. "The profession." *The American library journal* 1 (1): 5.

** Tutti siti web hanno come data di ultima consultazione il 26 marzo 2025.

Di Domenico, Giovanni. 2009. *Biblioteconomia e culture organizzative. La gestione responsabile della biblioteca*. Milano: Editrice Bibliografica.

Faggiolani, Chiara. 2020. *Come un ministro per la cultura: Giulio Einaudi e le biblioteche nel sistema del libro*. Firenze: Firenze University Press.

“Il futuro in biblioteca, la biblioteca in futuro.” 2011. *Bollettino AIB* 51 (4): 313-14.

Gorman, Michael. 2002. *I nostri valori. La biblioteconomia nel XXI secolo*. Traduzione di Agnese Galeffi, con la collaborazione di Carlo Ghilli, a cura e con presentazione di Mauro Guerrini, postfazione di Alberto Petrucciani. Udine: Forum.

Gorman, Michael. 2004. *La biblioteca come valore. Tecnologia, tradizione e innovazione nell'evoluzione di un servizio*. Traduzione di Matteo Barucci, a cura e con prefazione di Mauro Guerrini, postfazione di Alberto Petrucciani. Udine: Forum. Traduzione di: *The enduring library. Technology, tradition and the quest for balance*. Chicago: American Library Association, 2003.

Gorman, Michael. 2018. *I nostri valori, rivisti: la biblioteconomia in un mondo in trasformazione*. Traduzione di Giuliano Genetasio, prefazioni di Rosa Maiello e Alberto Petrucciani, edizione italiana a cura e con introduzione di Mauro Guerrini. Firenze: Firenze University Press.

Guerrini, Mauro. 2022a. *Dalla catalogazione alla metadattazione: tracce di un percorso*. Prefazione di Barbara Tillett, postfazione di Giovanni Bergamin. Seconda edizione, a cura di Denise Biagiotti e Laura Manzoni, con prefazione di Peter Lor. Roma: Associazione italiana biblioteche.

Guerrini, Mauro. 2022b. *Metadattazione: la dimensione contemporanea della catalogazione*. Prefazione di Diego Maltese, introduzione di Paola Castellucci e Gino Roncaglia. Milano: Editrice Bibliografica.

Kupiec, Anne. 2003. “Qu'est-ce qu'un(e) bibliothécaire?.” *Bulletin des bibliothèques de France (BBF)* 1: 5.

Kuttner, Sven. 2009. “Beruf. Die Wieder-Buzás-Kontroverse 1959 bis 1962. Ein Blick hinter die Kulissen einer Berufsbilddiskussion der späten Nachkriegszeit.” *Bibliotheksdienst* 43 (4): 384-98.

Hirsh, Sandra, ed. 2024. *Library 2035: imagining the next generation of libraries*. Lanham, Maryland: Rowman & Littlefield Publishers.

Lor, Peter Johan. 2024. “Biblioteche, internazionalismo e pace.” *Bibelot* 30 (1): 25.

Manoni, Paola, Giovanna Mazzola Merola, Flavia Cancedda, e Giovanni Michetti. 2015. “Quaderno sui profili professionali. Bibliotecario e archivista nelle norme UNI 11535:2014 e UNI 11536:2014.” *AIB studi* 55 (1): [105]-34.

Melot, Michel. 2005. *La saggezza del bibliotecario*. Milano: Sylvestre Bonnard.

Petrucciani, Alberto. 1991. *Discipline del libro, identità professionale, ricerca scientifica*. In *Università e biblioteca*, a cura di Enzo Esposito, 36. Ravenna: Longo.

Petrucciani, Alberto. 2003. “Storie di ordinaria dittatura: i bibliotecari italiani e il fascismo (1922-1942).” *Bollettino AIB* 43 (4): [417]-442.

Petruciani, Alberto. 2007. *Biblioteca e professione bibliotecaria: quale prospettiva per oggi e per domani?*. In *Le politiche delle biblioteche in Italia: la professione: atti del 53° Congresso nazionale dell'Associazione italiana biblioteche, Roma, 18-20 ottobre 2006*, 207-12. Roma: Associazione italiana biblioteche.

Ponzani, Vittorio, Paola Manoni, e Giovanna Mazzola Merola, "Il riconoscimento della figura professionale del bibliotecario e la nuova edizione della UNI 11535." *AIB studi* 3: 553-62.

Poulain, Martine. 1994. "Le bibliothèques françaises au XX^e siècle." *Documentation et bibliothèques* 40 (1): 6.

Prandstraller, Gian Paolo. 1972. *L'intellettuale tecnico e altri saggi*. Milano: Edizioni di Comunità.

Ranganathan, Shiyali Ramamrita. 2010. *Le cinque leggi della biblioteconomia*. Traduzione e note a cura di Laura Toti, saggio introduttivo di Giovanni Solimine. Firenze: Le Lettere.

Ranganathan, Shiyali Ramamrita. 2023. *L'organizzazione delle biblioteche*, edizione e traduzione a cura di Carlo Bianchini. Firenze: Firenze University Press.

Rasetti, Maria Stella. 2014. *Bibliotecario, il mestiere più bello del mondo*. Milano: Editrice Bibliografica.

Revelli, Carlo. 1995. "Compiti e caratteristiche del bibliotecario – 2. La crisi dell'identità professionale e l'impatto con le nuove tecnologie." *Biblioteche oggi* 2: 51.

Revelli, Carlo. 2007. "Il mestiere del bibliotecario." *Biblioteche oggi* 1: 68.

Rosati, Luca. 2019. *Sense-making: organizzare il mare dell'informazione e creare valore con le persone*. Roma: UXU.

Salarelli, Alberto. 2005. "Il bibliotecario sopraffatto." *Bibliotime* 8 (1) <https://www.aib.it/aib/sezioni/emr/bibtime/num-viii-1/salarell.htm>.

Solimine, Giovanni. 1988. *Il bibliotecario: una professione, molte fisionomie*. In *La cultura della biblioteca. Gli strumenti, i luoghi, le tendenze: atti del Convegno di Biblioteche oggi*, realizzato in collaborazione con Comunità montana del Marmore, Comune di Châtillon, Association des bibliothécaires valdôtains, AIB Sezione Piemonte, Châtillon, 22-24 maggio 1987, a cura di Massimo Belotti, 87. Milano: Editrice Bibliografica.

Traniello, Paolo. 2017. *Prefazione*. In Mauro Guerrini, *De bibliothecariis. Persone, idee, linguaggi*, premessa di Luigi Dei, prefazione di Paolo Traniello, presentazione di Graziano Ruffini, a cura di Tiziana Stagi, XXII. Firenze: Firenze University Press.

Wieder, Joachim. 1959. "Berufssorgen des wissenschaftlichen Bibliothekars." *Libri* 9: 132-65.

Why should I be an archivist?*

Federico Valacchi^(a)

a) University of Macerata, <https://orcid.org/0000-0003-2710-9316>

Contact: Federico Valacchi, federico.valacchi@unimc.it

Received: 25 March 2025; **Accepted:** 25 March 2025; **First Published:** 15 May 2025

ABSTRACT

The contribution examines the salient features of the archival profession with particular reference to the value system underlying the work of the archivist and the consequences of the current transformations.

KEYWORDS

Archives; Archivistics; Deontology; Digital archives.

Perché dovrei fare l'archivista?*

ABSTRACT

Il contributo prende in esame i tratti salienti della professione archivistica con particolare riferimento al sistema di valori sotteso al lavoro dell'archivista e alle conseguenze delle trasformazioni in atto.

PAROLE CHIAVE

Archivi; Archivistica; Deontologia; Archivio digitale.

* Questo contributo è la rielaborazione della Lectio Magistralis “A difesa della conoscenza critica: le responsabilità degli archivisti e dei bibliotecari”, tenuta insieme al Prof. Mauro Guerrini a Firenze il 4 marzo 2025. Nel tentativo di mantenersi fedeli a quell'impostazione si è preferito non utilizzare le note, limitandosi a citare nel corpo del testo le fonti letterarie utilizzate.

La domanda che dà il titolo a queste brevi considerazioni ha molte possibili risposte, tante quante sono le opportunità e le complessità che si nascondono nella parola archivio e che l'archivista eredita dal suo oggetto di studio e lavoro. La polifunzionalità degli archivi impone infatti una molteplicità di sguardi e di azioni e impedisce di pensare all'archivista come ad una figura professionale a senso unico. Non è concesso, o, meglio, non è più concesso, immaginarsi l'archivista imprigionato dentro ai patetici stereotipi disegnati dalla subcultura dei luoghi comuni. E non è neppure concesso agli archivisti assumere atteggiamenti che possano continuare ad alimentare la mistificante retorica dell'ultimo tra gli ultimi, che sconta la sua pena a colpi di timbro, respirando la polvere del più disagiato degli uffici. Ci sono molte buone ragioni per scegliere questa professione ma, se lo si fa, ci si deve assumere anche la responsabilità di difenderla fattivamente dalle molte false interpretazioni che se ne danno e dalle molte invasioni di campo determinate dalla convinzione che a gestire un archivio basti il buon senso e, magari, una cultura umanistica generalista.

L'archivistica non è una disciplina umanistica ma piuttosto giuridica, amministrativa e, quando i tempi sono maturi, storica. Gli archivi non sono soltanto beni culturali e non nascono come tali. Sono strumenti di vita quotidiana che solo quando svanisce quasi del tutto la ragione operativa per cui sono nati si trasformano in risorse culturali. Dire archivista significa alludere perciò alle molte specializzazioni che la parola sintetizza, dal *data curator* all'archivista storico, transitando per una serie di ulteriori sfumature che l'evoluzione dei mezzi di produzione e trasmissione impone.

Indipendentemente dalla sfera di azione e dal profilo specialistico gli archivisti condividono un sistema di valori e di competenze di base cui non si può rinunciare. Specializzarsi in ambito archivistico significa perciò imparare a declinare, e se necessario aggiornare, i fondamentali tecnici e culturali dentro a contesti mutevoli. Significa anche essere disponibili ad accettare l'idea che la formazione professionale è un percorso in costante aggiornamento. L'archivista lavora con gli archivi e gli archivi cambiano costantemente in ragione di molti fattori, da quelli tecnologici a quelli legati al diversificarsi degli utenti e della destinazione d'uso. Fare l'archivista oggi significa più che mai essere pronti a mettere in discussione i fondamenti della disciplina per difenderla nella sua essenza. Ed è proprio l'uso che se ne fa che qualifica l'archivio e il professionista che governa e organizza i dati, indirizzandoli a specifiche utilizzazioni.

L'archivistica, a ben guardare, è infatti un'impalcatura che si regge sulle parole e, preferibilmente, sulle parole scritte. Si destreggia sul filo sottile che lega gli avvenimenti alla loro memoria, in perenne equilibrio tra la storia e l'amministrazione. Scrivere significa in ogni caso registrare, fissare nel tempo e nello spazio ma, alla fine, si scrive perché poi sia possibile dimenticare, quando l'ombra delle cose ci schiaccia e l'oblio è una ragionevole speranza di venirne fuori.

Fare l'archivista significa quindi confrontarsi con le ondate documentarie che scaturiscono dalla vita quotidiana, perché la scrittura negli archivi è una conseguenza di cose fatte in quanto scritte. A monte della professione c'è un sistema di valori consolidati e il bisogno di contribuire a una società ordinata, perché notoriamente, e secondo Derrida, l'archivio è il luogo dove l'ordine è dato. Citando *Le Avventure di Archinia*, la graphic novel archivistica che chi scrive ha creato a suo tempo insieme a Salvatore Renna, si potrebbe dire, diversamente, che *«quanto agli archivi, come sempre, sono un ricordo del tempo a venire. Ci sono ma non si vedono. Vivono nei ricordi dei nostri eroi, nell'immaginazione di giovani storici pieni di speranza di poter raccontare molti domani e nei bisogni quotidiani di ogni abitante»*.

In questi archivi fatti di tempo c'è paradossalmente una dimensione fuori dal tempo che rischia di collocarli in una bolla asincrona e astratta. Il mito dell'archivio è il primo nemico dell'archivio. Conviene allora ribadire subito con forza che, pur a fronte di tante suggestioni in questo senso, l'archivista non è un pensatore o un eremita della memoria e l'archivistica non è un'arma di distrazione dalle cose del mondo. L'archivistica, al contrario, è una professione dai contorni molto netti, non un hobby colto, o, peggio ancora, una missione. È un mestiere che va fatto *con* passione ma non semplicemente *per* passione. Dell'archivistica e degli archivisti con le loro specializzazioni c'è bisogno nella concretezza del vivere. Ogni società organizzata per andare avanti ha bisogno di certificarsi e di ricordare, e questo è il senso primordiale dell'archivistica: contribuire alla vita quotidiana. Per queste ragioni, soprattutto in un mondo travolto dalle informazioni, gli archivisti non sono gli animaletti innocui e benefici cari a Benedetto Croce, e neppure sacerdoti che amministrano il culto della memoria, inevitabilmente inchiodati al passato. Sono piuttosto dei professionisti della complessità, le cui competenze aiutano a governare e comprendere il presente senza dimenticare le sue premesse e le sue conseguenze.

Il futuro, peraltro, è una questione di responsabilità ed è il tempo in cui il mestiere archivístico dà il meglio di sé. Certo, bisogna anche intendersi su cosa, dove e quando sia il futuro, perché non basta collocarlo davanti a noi per capire come meritarcelo. Bjørn Larsen, in un suo romanzo molto ben riuscito, *Filosofia del pendolare*, ci invita ad esempio a dubitare del futuro: “*prendi sempre il primo treno che arriva al binario, anche se è pieno zeppo e c'è solo posto in piedi. Del futuro non ci si può mai fidare*”. Non gli si può dare torto perché, in effetti, del futuro tutti noi, veggenti a parte, prefiguriamo l'avvento, ma non conosciamo la sostanza. Gli archivi anche in questo caso possono andare contro tendenza e sorprenderci. I complessi di documenti infatti alimentano ragionevoli speranze di immaginare il futuro, perché sono destinati a popolarlo con i loro contenuti e a dargli un inevitabile passato. La responsabilità dell'archivista sta allora proprio nel dare qualificate garanzie di affidabilità al futuro, accettando l'oneroso compito di trasferire nel tempo che verrà ciò che il passato e il presente hanno lasciato sedimentare. L'archivista in questo senso è un traghettatore di dati, una figura professionale che coi dati costruisce un'ipotesi attendibile di futuro. Questa responsabilità è ancora più marcata nei nostri faticosi anni tecnologici, quando la dematerializzazione ci chiama ad accudire con urgenza gli archivi digitali che senza la nostra attenzione andranno a fondo, spingendo verso le profondità dell'oblio ciò che noi siamo oggi. La sfida delle tecnologie (ammesso che si voglia usare questa espressione che gronda di vuota retorica), in campo archivístico pone questioni che non esiterei a definire devastanti. Nel sistema digitale, infatti, il documento e l'archivio stanno dissolvendosi e vanno perdendo quell'univocità fisica e filologica che da sempre ci ha consentito di fissarli in strutture logiche predeterminate. Il soggetto produttore si scontra nelle nebbie dell'interoperabilità e la materialità del documento è trascinata via dalle scariche elettriche che ne accompagnano la costruzione e la trasmissione. La memoria è subordinata alla memorizzazione e si fa prima di tutto immagazzinamento di dati in qualche posto. La struttura del documento, che con tanta cura abbiamo stabilizzato e studiato nel suo riprodursi analogico, diventa “una serie finita di celle, ciascuna delle quali è individuata da un indirizzo di memoria univoco e contiene un numero prestabilito di informazioni”, come recita la Treccani alla voce memorizzazione, e come ha notato Roberto Guarasci in un recente convegno. La destrutturazione programmatica che le tecnologie determinano mette in forte affanno il canone archivístico, delocalizzandone la ragion d'essere faticosamente messa a punto. Vengono messe

in discussione le stringenti relazioni fisiche su cui si è costruita nel tempo la rappresentazione possibile di un archivio sedentario e gerarchico.

Vacilla perfino l'idea di vincolo archivistico. Il vincolo, oltre che tutto il resto, è per così dire anche una necessarietà rappresentativa, un modo di immaginare l'azione amministrativa, che si sviluppa in ragione di archivi gerarchici e strutturati come i loro soggetti produttori. O, viceversa, è il risultato di un certo tipo di pensiero giuridico amministrativo che ne ha bisogno per dare sequenzialità alla sua azione. L'azione però esisteva anche prima del fascicolo: in una filza dov'è il vincolo? Forse vicino a dove lo colloca la memorizzazione, in una posizione non speculare all'azione e legata a modalità organizzative diverse e determinate da un pensiero più che da una necessarietà.

Sono questioni molto più grandi di queste pagine, ma certamente l'archivistica davanti a tutto questo è chiamata ad interrogarsi a fondo, nella speranza di trovare dentro o fuori dai suoi schemi risposte attendibili. La responsabilità di proiettare memoria nel futuro e la capacità tecnica che serve per farlo sono allora davvero buone ragioni per fare l'archivista, almeno dentro alla esasperante corsa tecnologica dentro alla quale viviamo.

Su un altro versante, poi, l'archivistica è in estrema sintesi una disciplina di educazione alla cittadinanza e l'archivio è un servizio alla collettività. Ci insegnò a suo tempo Eugenio Casanova nel suo ben noto manuale che *“la conservazione degli atti in archivio (...) ha uno scopo positivo ben determinato e tangibile, vale a dire quello di renderli utili alla generalità degli individui e agli individui stessi. Appunto in conseguenza di questo scopo, di questa utilità, l'archivio acquista un carattere squisitamente etico”*. La dimensione etica della conservazione prende corpo dalle conseguenze di precise azioni archivistiche e la cultura della gestione documentale, cui conducono le buone pratiche di archiviazione; è quasi tautologicamente cultura delle istituzioni e garanzia di trasparenza e democrazia. Protocollare, classificare, fascicolare, così come descrivere, ordinare e produrre inventari sono attività tecniche mai fini a sé stesse. Sono piuttosto azioni etiche, che garantiscono la qualità “morale” dell'archivio in quanto bene collettivo polifunzionale. La dimensione tecnica dell'archivistica, i cui precetti informano anche le norme, è l'unica garanzia di efficienza degli archivi e l'insostituibile premessa a una più diffusa funzione pubblica degli archivi stessi. Un buon archivista è anche un buon cittadino e un cittadino responsabile e consapevole ha nell'archivista un forte alleato. E anche questa è un'ottima ragione, se non la migliore, per fare l'archivista.

Gli archivi però, è bene ricordarlo ribaltando in qualche modo i termini della questione, nascono come informazione organizzata, ma finiscono il loro ciclo vitale come memoria qualificata. Agli antipodi della formazione la polifunzionalità trasfigura l'archivio e ne fa un potente serbatoio di memoria. Fare l'archivista, allora, significa anche prendersi carico del trasformismo degli archivi per contenerne le derive autorappresentative e le esuberanze narrative. L'archivio vive, e vivendo si trasforma. Come ci ricorda Gabriel Garcia Marquez nella sua autobiografia, *Vivir para contarla*, infatti, *“La vida no es la que uno vivió, sino la que uno recuerda y cómo la recuerda para contarla”*, e gli archivi non sfuggono a questa regola. La loro diacronica esistenza li plasma e ne modifica gli assetti. La sedimentazione “naturale” e spontanea strada facendo perde l'ingenuità delle origini e diventa un ricordo elaborato, finalizzato alla autorappresentazione. Gli archivisti stanno esattamente al centro di questo processo, intenti dapprima a garantire una sedimentazione organica e poi votati a dare conto delle trasformazioni intervenute, che alla fine chiamiamo contesto. Gli archivi però non sono letteratura, per quanto nella letteratura ci siano tanti archivi e viceversa. L'analisi del grande scrittore colombiano, quindi, introduce il tema ma non lo risolve. Anche in

questo passaggio ci vuole professionalità e ci vogliono strumenti metodologici, perché in un archivio niente accade per caso. Il racconto sotteso all'evoluzione di un complesso documentario verso forme di memoria organica si chiama mediazione. Ma la mediazione, a sua volta, non è una categoria dello spirito. Presuppone un duro lavoro di conoscenza dei contesti e dei contenuti, e prende corpo a partire dal metodo e da quella descrizione archivistica che sotto molti punti di vista è la più nobile delle diversificate arti archivistiche. Come ci ricorda Josè Saramago, per bocca dell'Artur Paz Semedo di *Alabarde Alabarde*, *"per fare un lavoro del genere sarà necessario un orientamento, un criterio, non può servire qualsiasi documento solo perché a me è parso importante"*. L'archivista, in quanto professionista, non può effettivamente fare a meno del metodo, perché il metodo nel suo insieme è l'unica speranza di trasformare un caos potenziale in informazione fruibile, in quanto organizzata e qualificata. Il metodo in archivistica è al tempo stesso professione di umiltà e rigoroso anelito all'oggettività. Il rispetto del metodo ci indirizza verso un'oggettività quanto più possibile neutrale rispetto agli oggetti della descrizione. La valutatività, però, è il sogno che ci tiene svegli quando descriviamo un archivio. La descrizione passa anche dalla capacità di ascoltare l'archivio con umiltà, senza pretendere che esso risponda di necessità ai nostri pregiudizi interpretativi impregnati di metodo asettico. L'oggettività nel processo di mediazione è un obiettivo cui tendere, mai una certezza, perché gli archivisti hanno a che fare con una memoria vitale non con un passato statico da riesumare.

La memoria degli archivi ha le fattezze della statua che ancora Josè Saramago evoca nei *Quaderni di Lanzarote*: *"è anche una statua di argilla. Il vento passa e a poco a poco le porta via particelle, granelli cristalli. (...) Minuto dopo minuto quello che era non è più e della statua non resterebbe altro che una sagoma informe, un impasto primario se pure un minuto dopo l'altro non continuassimo a restaurare, a memoria, la memoria. La statua si manterrà in piedi, non è la stessa ma non è un'altra"*. Questa memoria erosa dalle intemperie e faticosamente tenuta in piedi grazie anche agli archivisti, a ben guardare, è l'unica fisiologicamente possibile. La sua alternativa è una cupa staticità antiquaria, dentro alla quale si perdono i dinamismi dell'identità che il passato sa insegnarci solo se siamo a disposti a fare lo sforzo di mantenerlo vivo, senza celebrarne esequie solenni ma vuote.

La memoria di argilla, se la vogliamo etichettare così, non cerca assiomatiche certezze, ma accetta la fatica di processi in divenire ed è disponibile a fare i conti con una verità che alla fine negli archivi non esiste, almeno come rivelazione mistica ed esogena. *Lungo le strade della storia non è vera la verità, ma il suo racconto. Voi siete dentro alla storia e potete vedere ciò che agli altri è celato. Ogni dimensione ha un finale diverso, così si può scegliere*, si legge ancora nelle *Avventure di Archinia*. La verità negli archivi è quindi un'ipotesi o, meglio, un fantasma ingannevole. L'affidabilità dell'archivio, in ragione di quella fluidità che abbiamo più volte richiamato, non si traduce in affermazioni apodittiche, è a sua volta un processo in divenire, che mette in gioco una molteplicità di fattori e che può dar luogo ogni volta a soluzioni diverse.

La verità, questa verità, in fondo è una scelta, una risposta possibile alle molte domande che all'archivio si possono porre, sulla scia più di soprese serendipitose che di monumentali certezze euristiche. Ogni archivio, del resto, è un mondo intero, un sistema di relazioni interne ed esterne che si aprono sulla realtà e sulle sue molte fenomenologie. Lo spiega meglio di tutti Italo Calvino ne *Le Città invisibili*, basta sostituire archivio a città: *"Le città (cioè gli archivi) (...) sono un insieme di tante cose: di memoria, di desideri, di segni d'un linguaggio ; sono un luogo di scambio (...), ma questi scambi non sono solo scambi di merci, sono scambi di parole, di desideri, di ricordi"*. L'idea

che l'archivio sia un luogo di scambio non è solo una suggestiva capriola di senso, perché in ogni fondo archivistico l'informazione si distribuisce lungo percorsi che divergono volentieri dalla pista principale e finiscono col richiamare energie esterne, capaci di interazioni stupefacenti. Si sa da quale porta si entra ma è difficile immaginare dove e quando si troverà l'uscita.

L'archivista cerca come può di far fronte a tanta intensità, perché questo è il suo compito. Non basta la tecnica, però, per quanto sopraffina possa essere. Ci vuole fantasia, per fare l'archivista o, meglio, servono capacità mimetiche di immedesimarsi nel passato che si tenta di interpretare. Lo stesso metodo storico è in fondo una rappresentazione nella quale mettiamo in scena i soggetti produttori come protagonisti di una recita a soggetto di cui l'inventario è al tempo stesso il copione e il deus ex machina che alla fine toglie le castagne del fuoco. Il tutto mantenendo un prudente understatement, perché come ci insegna Pessoa nel *Libro dell'inquietudine*: *"i classificatori di cose, che sono quegli uomini di scienza la cui scienza consiste solo nel classificare, ignorano in genere che il classificabile è infinito e che dunque non si può classificare"*. Per la verità, con buona pace del saggio portoghese, gli archivisti a questa impossibilità di classificare ci credono molto poco, tassonomisti incalliti quali sono, e si ostinano ad impilare contesti e contenuti per farne materia informativa. C'è poco da fare, la mediazione è il loro destino, e soprattutto per questo da grande vorrei fare l'archivista.