

## “Hence that sense of new air”. Archival debate and historiographical developments in Italy between the early and late 20<sup>th</sup> century

Raffaele Pittella<sup>(a)</sup>

a) University of Rome3, <https://orcid.org/0000-0001-9927-2038>

Contact: Raffaele Pittella, [raffaeleantoniosimo.pittella@uniroma3.it](mailto:raffaeleantoniosimo.pittella@uniroma3.it)

Received: 31 January 2024; Accepted: 04 March 2024; First Published: 15 May 2024

### ABSTRACT

In the second half of the twentieth century it became generally accepted that the goal of the archivist's work should be not only to reconstruct the history of the object's producer, but also to question its genesis, preservation and the use of documentary nuclei over time. The result was a revisiting of the principles of historical method, thanks in part to the fruitful debate that was established between the community of archivists and that of historians.

### KEYWORDS

Archivistics; Historiography; Archival research methodology; Historical research methodology.

## “Di qui quel senso di aria nuova”. Dibattito archivistico e sviluppi storiografici in Italia tra primo e secondo Novecento

### ABSTRACT

Nella seconda metà del Novecento andò consolidandosi l'idea che obiettivo del lavoro dell'archivista dovesse essere non solo quello di ricostruire la storia del soggetto produttore, ma interrogarsi sulla genesi, conservazione e utilizzo nel tempo dei complessi documentari. Ne derivò una rivisitazione dei principi del “metodo storico”, grazie anche al fecondo dibattito che si stabilì fra la comunità degli archivisti e quella degli storici.

### PAROLE CHIAVE

Archivistica; Storiografia; Metodologia della ricerca archivistica; Metodologia della ricerca storica.

## 1. Un libro da leggere come “introduzione alla *Guida*”

Nel 1987 Isabella Zanni Rosiello pubblicava per i tipi del Mulino un volume destinato da subito a suscitare un acceso dibattito nella comunità degli archivisti, e ad attirare l'interesse degli storici su di una materia che esulava dalla storia istituzionale in senso stretto, poiché incentrata sulle modalità di formazione, uso e trasmissione degli archivi tra medioevo ed età contemporanea (Zanni Rosiello 1987). Il libro in questione, dal titolo *Archivi e memoria storica*, se infatti incontrò il favore di quella parte dell'archivistica italiana secondo la quale occuparsi di archivi non poteva ridursi alla mera descrizione delle scritture sfuggite alla distruzioni del tempo e degli uomini – o alla sola storia del soggetto produttore – di contro generò dubbi e perplessità fra chi si dichiarava fedele a quell'idea dell'archivistica “pura” teorizzata negli anni Venti del Novecento da Eugenio Casanova,<sup>1</sup> e tra quanti si sentivano i custodi del pensiero di Giorgio Cencetti, per il quale era persino inesatto dire che l'archivio rispecchia l'ente, perché è l'ente medesimo (Cencetti [1935-1963] 1970, 38-69). Tra coloro che in queste pagine avvertirono “un senso di aria nuova” vi fu Filippo Valenti<sup>2</sup>, che dedicò al lavoro della Zanni Rosiello una lunga e articolata recensione, che, senza nascondere i punti ritenuti deboli, ne esaltava non tanto il valore innovativo – sebbene anche questo gli fosse per alcuni versi riconosciuto – quanto piuttosto la capacità di aver dato voce a quanti, operando sugli archivi e negli archivi, avevano avuto modo di sperimentare, nella pratica quotidiana dell'ordinamento e dell'inventariazione, che la morfologia dei complessi documentari non si conserva necessariamente immutata né sempre riflette il volto dell'istituto che lo ha prodotto, ma spesso si trasforma e cambia in riferimento al diverso utilizzo che di uno stesso nucleo di carte si fa nel tempo (Valenti [1989] 2000e). L'impressione che tale lettura suscitò nello studioso modenese fu quella di una sorta di inversione dottrinaia, che ribaltava e metteva in discussione i principi dell'ortodossia archivistica, sui quali nel corso della prima metà del secolo tale materia era andata riconoscendosi come disciplina autonoma, avente una sua dignità accademica e un suo statuto epistemologico (Mineo 2015): “circoli viziosi da troppo tempo sclerotizzati” sembravano in queste pagine sgretolarsi dinanzi ad un'immagine dinamica e plastica dell'archivio; e la presenza di “una rinnovata luce” lasciava intravedere inusitate prospettive anche al cospetto di “argomenti i più ritriti”. Sembrava quasi che si trattasse di un involontario atto di accusa verso la tradizionale trattatistica, troppo precettistica, tutta incentrata sulle teorizzazioni, volta a fare di questo ambito di studi una materia per soli addetti ai lavori: “se si può assumere che la Zanni Rosiello – scrive Valenti – ponga implicitamente in discussione, a partire quanto meno da un certo livello, lo statuto dell'archivistica come disciplina da manuale, per risolverla [...] in un discorso aperto sugli archivi, allora il suo libro può svolgere un ruolo di notevole rilevanza: se non di rottura, certo di un sostanziale rinnovamento” (Valenti [1989] 2000e, 116). Che non si trattasse di un manuale di archivistica in senso stretto era cosa ben chiara non solo all'autrice, ma anche al pubblico cui il libro era destinato: “archivisti, storici e più in generale chi ha interesse a utilizzare fonti documentarie” (Zanni Rosiello 1987, [5]). Un libro dunque incentrato sugli archivi e sugli archivisti, e non un libro di archivistica: una parola, quest'ultima, che “non appare praticamente mai nelle centocinquanta

<sup>1</sup> Sull'archivistica come disciplina tripartita in archiveconomia, archivistica pura e utilizzazione degli archivi (Casanova 1928; 1929). Su Casanova teorico dell'archivistica “pura” (Lodolini 1958; Petrucci 1978).

<sup>2</sup> Sulla concezione archivistica di Valenti e sul suo interesse per gli studi filosofici (Franzese 2023).

pagine del testo, e neppure [...] nelle ventinove della bibliografia ragionata, ove manca per di più una sezione dedicata alle opere di carattere generale sugli archivi, e dove neppure i più classici e consacrati manuali e trattati vengono menzionati come tali, ma solo eventualmente a ciò che dicono in ordine a determinati problemi” (Valenti [1989] 2000e, 115-116). Evidente e netta è quindi la distanza che intercorre tra questo nuovo modo di guardare all’archivio e quello andato profilando nelle disquisizioni dottrinarie, nella manualistica canonizzata, nelle speculazioni astratte che avevano caratterizzato la prima metà del Novecento: al rigorismo di un metodo storico incentrato sul dogma della coincidenza organica tra l’ente e l’archivio, la Zanni Rosiello contrapponeva la duttilità di un’impostazione volta a cogliere le incoerenze e le contraddizioni esistenti fra il modo in cui l’archivio è stato creato e la forma in cui è stato trasmesso, tra la storia dell’archivio e la storia dell’ente. Una prospettiva d’indagine, questa, volta cioè ad evidenziare peculiarità e differenze esistenti fra complessi documentari all’apparenza simili per soggetto produttore e struttura; in grado di spiegare il perché dei pieni e dei vuoti documentari; capace di mettere in luce le azioni volontarie o accidentali legate alla permanenza o alla distruzione di certi e non altri nuclei di scritture. Non è un caso infatti che Adriano Prosperi, nel commentare il ruolo avuto dalla *Guida generale degli Archivi di Stato* nello sviluppo degli studi storici, individui proprio nella pubblicazione della Zanni Rosiello il testo che meglio avrebbe potuto assurgere al ruolo di introduzione alla *Guida* stessa, una sorta di *vademecum* per chi voglia “entrare nella cucina della storia, un rimedio indispensabile se si vuole evitare che il passato storico diventi sempre più [...] un illustre sconosciuto” (Prosperi 1988, 252).

## 2. Sconfinamenti disciplinari

Come precisato nella breve ma densa premessa, *Archivi e memoria storica* costituiva una “raccolta” di scritti, nelle intenzioni ordinata e coerente, su particolari aspetti della produzione, conservazione ed uso dell’archivio inteso come fonte; dell’archivio osservato cioè da un’angolazione storica. Una raccolta dunque che intendeva porre l’accento su come il “passaggio del ricordo” si discosti sovente da quel legame di stretta reciprocità *archivio/ente, complesso documentale/soggetto produttore* su cui frequentemente si è insistito, trattandosi di un passaggio spesso influenzato da scelte, condizionamenti, mediazioni istituzionali che modificano le originarie geografie documentarie, ne alterano i confini, finendo così per attribuire agli archivi nuovi significati, se non addirittura per creare nuovi archivi e nuovi documenti: smembrando o accorpando, slegando o unendo carte appartenenti in principio a contesti diversi. Non è raro infatti che nella “trasmissione del ricordo” – intesa come conservazione e tradizione archivistica – intervengano “ostacoli, emarginazioni, sconfitte, sepolture, ma anche fortune, onori, splendori, resurrezioni” (Zanni Rosiello 1987, 42). È quella degli archivi una storia segnata da processi di conservazione e perdite, da occultamenti e riemersioni, da distruzioni talvolta volontarie, talaltre accidentali (D’Angiolini e Pavone [1973] 2004, 300-301; Pavone [1961-1970] 2004b, 50-51). Una storia che sfugge in molti casi all’assioma per il quale fare archivistica equivale a conoscere la storia dell’ente presso il quale un dato complesso è andato formandosi. Sono infatti queste le parole della Zanni Rosiello: “la documentazione archivistica, in quanto materiale di base per svolgere determinate attività, è strumento di pratiche di potere perché serve ad esercitarlo; in quanto accumulo-sedimentazione di pratiche di potere è

memoria e sapere del medesimo da conservare, e quindi trasmettere, o da distruggere, e quindi negare alla posterità” (Zanni Rosiello 1987, 54-55).

Si tratta di riflessioni maturate sul campo, formulate sulla scorta di riscontri ottenuti operando direttamente sulle carte. L'Archivio di Stato di Bologna – fra i più antichi e prestigiosi istituti di conservazione documentaria italiani – ha rappresentato infatti per la Zanni Rosiello una sorta di laboratorio ove effettuare sperimentazioni, raccogliere dati, verificare ipotesi (Zanni Rosiello, [1974] 2000a; [1982] 2000c). È lì che ha preso forma quell'idea di metodo storico che si intravede in filigrana in *Archivi e memoria storica*. Parlare di metodo storico significava infatti per l'autrice partire dalla consapevolezza che gli archivi non nascono per diventare “memoria-fonte”, cui attingere per scopi culturali, ma come strumenti di auto-documentazione, legati a ragioni pratiche, giuridiche, amministrative (Valacchi 2018). Gli archivi inoltre non si trasmettono nel tempo, salvo rare eccezioni, per una catena ingovernabile di casualità: la conservazione nei secoli di parte delle scritture prodotte, e la contemporanea distruzione di altre, corrisponde infatti nella maggioranza dei casi a scelte consapevoli, legate ai concetti validità o inutilità di esse: utilità pratica, contingente, per motivi e finalità che scaturiscono dalla contemporaneità e che finiscono per rendere attuali e indispensabili allo svolgimento degli affari correnti anche carte risalenti assai indietro nel tempo. Gli archivi dunque sono una materia plasmabile, scomponibile e ricomponibile (Valacchi 2021; 2023), e gli archivisti di oggi non debbono sconvolgersi dinanzi a palesi operazioni di manipolazione o di contaminazione documentaria operate nel tempo, perché se le scritture del passato sono giunte fino a noi è perché più d'uno nel tempo se ne è servito, riutilizzandole, in ragione di bisogni e necessità che scaturivano dal procedere della storia (Vitali 2014).

### 3. Verso una nuova idea di metodo storico

Una versione dunque aggiornata del metodo storico – quella così proposta dalla Zanni Rosiello – ma a ben guardare perfettamente in linea con quanto era stato già evidenziato – brevemente ma molto efficacemente – da Claudio Pavone ad inizio degli anni Settanta. Sebbene all'apparenza può sembrare “pacifico” che l'archivio “rispecchi” l'ufficio che lo ha generato – aveva fatto notare l'archivista e storico – di fatto invece, studiando le sedimentazioni documentarie con occhio simile a quello con cui si analizzano le relazioni esistenti fra gli strati archeologici, ci si accorge che la disposizione della carte nella maggioranza dei casi non fotografa l'organigramma del soggetto produttore, ma presenta sovente una struttura solo forzatamente sovrapponibile a quella dell'istituto (Pavone [1970] 2004a; [1961-1970] 2004b). Considerazioni, le sue, di forte impatto innovativo, che spingevano nella direzione di una revisione del metodo storico tradizionalmente inteso (Valacchi 2019). Considerazioni formulate fra l'altro sulla scorta di dati concreti; di quanto cioè da lui direttamente verificato partecipando alla redazione della *Guida generale degli archivi di Stato*: un lavoro monumentale di descrizione di centinaia e centinaia di chilometri di carte storiche, patrimonio della nazione italiana, prodotte e tramandate nell'ambito di apparati politico-amministrativi assai diversi fra loro, quali quelli caratterizzanti la Penisola prima del 1861. Descrivere in modo organico e con criteri uniformi una così vasta quantità di fondi, fra loro dissimili per le vicende burocratico-istituzionali che avevano segnato la storia degli antichi Stati italiani, aveva indotto infatti parte della comunità scientifica ad interrogarsi sulla infallibilità del principio cardine dell'archivistica

italiana, ovvero quello dell'uguaglianza fra l'archivio e l'ente, fra l'architettura delle carte e quella dell'istituto (D'Angiolini e Pavone 1981). L'archivistica si apriva in tal modo a nuovi orizzonti problematici e tornava a riflettere – questa volta con fare critico – sulle sue stesse origini, e a dibattere sull'impostazione idealistica ad essa attribuita da Giorgio Cencetti sul finire degli anni Trenta, attraverso la mediazione intellettuale di Benedetto Croce<sup>3</sup>. L'archivistica tornava in tal modo a interrogarsi con Pavone sull'utilità dell'impostazione filosofica in virtù della quale aveva costruito la sua autonomia scientifica rispetto agli altri rami del sapere, e che avrebbe profondamente indirizzato la produzione dottrina, la didattica e la legislazione di settore per gran parte della seconda metà del Novecento (D'Angiolini e Pavone [1973] 2004, 310-311). Nel saggio *Sul fondamento teorico della dottrina archivistica*, Cencetti, portando al massimo grado le teorizzazioni da lui elaborate nel 1937, finì infatti per affermare che più che parlare di specularità fra archivio ed ente occorre insistere sul concetto di identità, perché l'archivio è “in realtà l'ente medesimo o per lo meno è uno degli aspetti di vita di esso” (Cencetti (1935-1963) 1970, 40). Un'affermazione i cui effetti non tardarono a farsi sentire anche rispetto all'idea nuova che andò consolidandosi di metodo storico: “non esiste un problema del metodo di ordinamento”, aveva affermato lo studioso in quelle stesse pagine, dato che “la concretezza del metodo si risolve nella individualità, e ogni archivio ha il suo ordinamento” (Ibidem, 40). Per Cencetti dunque l'archivistica altro non era che una sorta di branca specializzata della storiografia – per dirla con le parole di Valenti – in quanto ordinare un archivio significava per lui “far rivivere in sé compiutamente e minutamente la vita dell'istituto” e trasformare un “archivio morto” in un “archivio vivo” (Valenti (1969) 2000a, 5), al fine di favorirne la fruizione per motivi di studio e di ricerca. Un'archivistica dunque intesa quale storia delle istituzioni, vista come ricostruzione delle vicende più significative che avevano segnato sotto il profilo giuridico la vita dell'ente; ed un metodo storico incentrato sulla convinzione che la struttura originaria di un archivio riproduca la storia, le funzioni e l'organizzazione dell'ente medesimo (Moscadelli 2023). A mettere in discussione questa visione sclerotizzata e rigida dell'archivio, e delle relative metodologie di ordinamento, non era stato solo Pavone, già prima di lui Filippo Valenti aveva gettato le basi per un'archivistica che tenesse conto del mutare delle regolamentazioni, delle prassi burocratiche e delle conoscenze tecnologiche, in quanto fattori determinati nella vita degli archivi sia rispetto alla fase della loro costituzione sia per quanto attiene ai possibili interventi operati nel tempo su di essi (Valenti [1969] 2000a; [1975-1976] 2000b; [1975] 2000c; [1981] 2000d). Ma non solo. Valenti non aveva trascurato altresì di sottolineare come nella storia di un archivio sembramenti, manipolazioni, divisioni e fusioni non costituiscano episodi rari o eccezionali, ma fatti che si ripetono sulla scia dei cambiamenti che inesorabilmente investono le istituzioni, la politica, l'amministrazione, la burocrazia.

È in queste riflessioni che trova forma e sostanza un nuovo modello di metodo storico, che induce a guardare all'archivio senza cedere alla tentazione di considerarlo manifestazione diretta e testi-

---

<sup>3</sup> Sull'azione esercitata dal pensiero di Croce sulla concezione archivistica di Cencetti ha particolarmente insistito Filippo Valenti, che a proposito degli articoli da quest'ultimo pubblicati nel 1937 e nel 1939 (il primo dal titolo *Sull'archivio come “universitas rerum”*; il secondo *Sul fondamento teorico della dottrina archivistica*) così si esprime: “articoli concepiti in chiave della più pura e genuina tradizione crociana che, ripeto, hanno profondamente influenzato buona parte della produzione dottrina e addirittura della stessa legislazione posteriore in fatto di archivi e che, soprattutto, hanno costituito tuttora praticamente la falsariga di tutto quanto l'insegnamento elementare dell'archivistica del nostro paese” (Valenti [1969] 2000, p. 4).



monianza materiale dell'ente produttore, del suo organigramma, delle sue funzioni e competenze. Lungo questa direttrice, l'archivista si fa storico: diventa cioè storico di "qualcosa" – l'archivio – e non semplicemente per il fatto che la sua preparazione culturale è anche di natura storica. Egli diventa tale in quanto analizza e studia l'archivio considerandolo un oggetto storico e utilizzando quegli stessi strumenti critici e quella stessa impostazione filologica che costituiscono i ferri del mestiere dello storico (Pavone 1965; [1970] 2004a; [2000-2001] 2004c; Zanni Rosiello [1982] 2000a). L'archivista-storico è in grado infatti di leggere e interpretare le ragioni dei vari montaggi e rimontaggi cui gli archivi sono andati soggetti nella loro, a volte, plurisecolare esistenza. Manipolazioni mai neutrali, ma risultato di precise scelte di potere, legate a esigenze pratiche di auto-documentazione che inducevano ad estrapolare da un dato *corpus* archivistico talune scritte, per riutilizzarle nell'ambito di altri ambiti e contesti, "secondo i canoni della elaborazione teorica di Giorgio Cencetti, spesso peraltro acriticamente accettati dai suoi epigoni, gli archivisti del passato avrebbero commesso imperdonabili 'errori'" (Zanni Rosiello [1980] 2000c, 254). Come è stato infatti rilevato, formulare giudizi di merito sull'operato di questi antichi burocrati non è la strada giusta da seguire: occorre invece interrogarsi sul perché di quelle scelte e sulle ragioni per le quali, in certi contesti istituzionali e in certe epoche, era stata avvertita l'esigenza di mettere mano sugli archivi del passato, attribuendo loro una nuova forma. Con Valenti e Pavone l'importanza di ripristinare l'ordine originario degli archivi, su cui l'archivistica italiana aveva centrato la sua attenzione per tutta la prima metà del Novecento, diventava una questione di secondaria importanza, se non addirittura un falso problema: il problema centrale per l'archivista che si occupa di documentazione storica non era più quello di stabilire il criterio cui ispirarsi nelle operazioni di ordinamento o di riordinamento, quanto piuttosto comprendere la reale struttura dell'archivio e "rispecchiarla in un inventario che sia, prim'ancora che un amo per pescare, una bussola per orientarsi" (Valenti [1989] 2000f, 125). Ed è appunto questo a rappresentare per Valenti la vera novità: "una novità assai significativa e una coraggiosa presa d'atto di una verità che dovrebbe essere ormai chiara" (Ibidem, 125). Si tratta in buona sostanza di una versione nuova e aggiornata del metodo storico, che continua a conservare la sua posizione di rilievo, come metodo per "eccellenza", nonostante le revisioni e gli aggiustamenti cui esso era stato così sottoposto.

#### 4. Una disciplina "mo' nata"

Nell'articolo dedicato all'edizione italiana dell'opera di Brenneke, Valenti, utilizzando un linguaggio secco e tagliente, parlava espressamente dei "funambolismi" che avevano caratterizzato la letteratura archivistica italiana post cencettiana nel tentativo di affermare "questa benedetta autonomia" della disciplina – sono parole sue – senza però "immiserire l'archivistica di una serie di regole empiriche per segretari e protocollisti" (Valenti [1969] 2000a). Nell'Italia dell'immediato secondo dopoguerra, in cui lo storicismo crociano contendeva il primato allo storicismo gramsciano, si assistette infatti al tentativo di conciliare Casanova con Cencetti, proponendo un'immagine dell'archivistica come terra di mezzo al confine tra la storia istituzionale e la messa a punto di "poco realistiche istruzioni" finalizzate al ripristino dell'ordine originario delle scritte. Elevare al rango di principi sistematici quelle regole e quelle pratiche utilizzate nella gestione ordinaria e materiale degli archivi aveva infatti rappresentato il tratto distintivo dell'archivistica "pura", su cui

aveva centrato l'attenzione Casanova nelle pagine del suo monumentale trattato datato 1928 (Casanova 1928); mentre per Cencetti era il rispetto del vincolo esistente fra i documenti a rappresentare la bussola a cui riferirsi nelle fasi di ordinamento di un complesso documentario: vincolo inteso come fisiologico, indistruttibile, imm modificabile, risultato dalla storia dell'ente stesso. Un progetto ambizioso, quello che così venne a determinarsi, volto da una parte a riconoscere l'attualità dell'insegnamento di Casanova, dall'altro a non distaccarsi dalla linea metodologica tracciata da Cencetti. Ambizioso anche perché si trattava di porre in relazione due visioni dell'archivio che viaggiavano su binari paralleli ma in direzioni opposte. Se per Casanova parlare di metodo storico significava ricostruire all'occorrenza quello che si riteneva che potesse essere stato l'ordine primigenio delle scritture – ridistribuendole gerarchicamente, mettendo al primo posto gli “atti costitutivi”, a seguire gli “atti appartenenti alle categorie consecutive ed esecutive” e in coda le “carte di corredo” (Ibidem, 218) – con Cencetti si era invece giunti a formulare quasi una sorta di catechismo per archivisti, che interpretava i complessi documentari come se si trattasse di una forma di categoria dello spirito, il che induceva a ritenere che ogni archivio ha l'ordinamento che “le competenze, le funzioni e le finalità e la particolare storia dell'ente che lo ha prodotto gli hanno conferito, e non potevano non conferirgli” (Valenti [1975-1976] 2000b, 167). Obiettivo di questa ardita ricucitura dottrinale era quello di attribuire dignità scientifica a una disciplina “mo' nata”, quale appunto era a quel tempo considerata l'archivistica storica. A definirla in tal modo era stato uno dei principali protagonisti di questa vicenda intellettuale, Leopoldo Cassese, che fu tra i primi a tentare di attribuire all'archivistica una solida base epistemologica, coniugando il pensiero di Casanova con quello di Cencetti, appoggiandosi ora agli scritti dell'uno ora dell'altro. La volontà che si scorge in controluce leggendo Cassese è quella di porre l'archivistica sullo stesso piano delle altre discipline aventi dignità accademica, affrancandola dalla condizione di subordinazione scientifica a cui la filosofia dello spirito crociana l'aveva relegata. È per l'appunto la filosofia idealista il bersaglio di un duro *j'accuse* rivolto nel 1949 da Cassese sulle colonne dell'organo d'informazione degli archivisti di Stato (Cassese 1949): sferzanti giudizi indirizzati specificatamente a Benedetto Croce, sebbene mai direttamente citandolo, il quale nel 1915 aveva collocato in una posizione di netta inferiorità rispetto agli storici il lavoro dei filologi – eruditi, bibliotecari, archivisti e archeologi – imputando loro di nutrire immotivatamente la convinzione di “tenere sottochiave la storia” per il solo fatto di custodire negli archivi, nelle biblioteche e nei musei – “le bianche tacite case dei morti” – testimonianze e narrazioni, documenti e monumenti ereditati dal passato. “Un'ingenua credenza”, la loro, aveva altresì affermato Croce, autoalimentata a suo giudizio dalla convinzione di “disserrare a loro libito le fonti [della storia] da cui l'assetata umanità potrà attingerla” senza sapere invece che “la storia e in noi tutti” e le “fonti sono nel nostro petto” e soltanto il nostro petto “è il crogiuolo in cui il certo si converte con il vero, e la filologia, coniugandosi con la filosofia, produce la storia” (Croce 1954, 17). Alle parole irridenti e canzonatorie con cui Croce si era dichiaratamente rivolto a quella “classe di lavoratori che si chiamano filologi”; all'immagine degli archivisti quali “veri animaletti innocui e benefici” per la “fertilità dei campi dello spirito”; al parallelismo tra gli archivisti e gli “innocui e benefici rospi” la cui scomparsa ha rischiato “di recente (...) l'agricoltura francese”; Cassese contrappone una serie di argomentazioni di ordine dottrinale, volte a trasformare dati empirici in pensiero speculativo. È in questo modo che egli intendeva evidenziare come l'archivistica, pur essendo una materia che trae origine da un'attività pratica, non si occupasse esclusivamente di questioni di ordine logistico, legate alla gestione e alla conservazione fisica dei complessi documen-

tari e al rapido rinvenimento delle scritture. L'immagine dell'archivistica proposta da Cassese si basava infatti innanzitutto sulla centralità del concetto di vincolo, così come esso era stato elaborato pochi decenni prima da Cencetti: il vincolo quale caratteristica peculiare e soggettiva dell'archivio e fattore distintivo fra l'archivio e la biblioteca; il vincolo quale punto di riferimento per l'archivista di oggi nell'impostare il proprio lavoro di ordinamento e inventariazione sugli archivi del passato; il vincolo quale elemento da "rianimare" per far rivivere l'archivio e, con esso, il soggetto produttore. Le carte di un archivio a differenza dei libri di una biblioteca – egli sottolinea – sono infatti legate da un vincolo genetico che affonda le sue radici nella storia dell'istituto che lo ha generato (Cassese 1949). Carte prodotte per esigenze pratiche, amministrative, in virtù di bisogni reali e specifici, e dunque testimonianza fervida e diretta delle *res gestae*. È dunque la "fonte documentaria", quella che scaturisce dalla concretezza della quotidianità, prodotta per fini utilitaristici, a costituire per Cassese l'oggetto dell'interesse specifico dell'archivista: l'archivio che non è pensabile senza un ufficio, un istituto o una persona che lo genera e lo alimenta; e il documento, cartaceo o membranaceo che esso sia, pubblico o privato, inteso quale oggetto che racchiudendo in sé un atto di volontà si collega necessariamente ad altre azioni precedenti o susseguenti, che generano a loro volta altre serie di documenti. Scritture le quali si legano necessariamente le une alle altre generando quel vincolo che per Cassese costituisce la materia specifica dell'interesse manifestato dall'archivista nei confronti dell'oggetto archivio.

## 5. Spirito umano e innocui animalletti: storici e archivisti a confronto

Un giudizio sferzante, quello espresso da Croce nei confronti dei "filologi". Tanto più d'impatto se letto attraverso la penna di Cassese. Un giudizio sul quale occorre fermarsi a riflettere, per capirne le reali motivazioni. Quelle che appaiono come parole irriverenti nei riguardi dei conservatori delle fonti storiche, acquistano infatti un significato più ampio se rapportate all'intera opera cui appartengono e, più in generale, al pensiero del filosofo e storico napoletano. È significativo fra l'altro che Cassese compia una forzatura, citando Croce: l'espressione il "nostro petto" – inteso come genere umano – si trasforma nel suo testo nelle parole "petto degli storici" (Cassese 1949). Una svista, probabilmente. Una modifica però, quella così apportata, che finiva per rimarcare il primato della storiografia sull'archivistica, e per collocare in una posizione di inferiorità intellettuale gli archivisti rispetto agli storici. Nell'economia generale dell'opera crociana, il tono irrispettoso con cui l'autore si rivolge ai "filologi" sembra essere più un monito che un atto d'accusa; un avvertimento serio e inequivocabile ad operare in una direzione che non fosse quella dello sterile eruditismo, ad evitare di trasformare archivi, biblioteche e musei in "muti" depositi di oggetti storici, incapaci di dialogare con il presente. È lo "spirito umano", per Croce, il luogo in cui si "raccolgono le tracce della vita passata, gli avanzi, i documenti" per "serbarli il più possibile inalterati o di restaurarli a misura se si alterano": è un "atto di vita, che serve alla vita, quel trascrivere storie vuote e raccogliere documenti morti", e verrà il momento – prosegue ancora – che essi ci "agevoleranno a riprodurre, arricchita, nel nostro spirito la storia passata, rifacendola presente" (Croce 1954, 15). A me sembra infatti che il tono beffardo con cui Croce si era rivolto ai "raccoglitori" delle fonti storiche si fa via via più labile e sfumato se interpretato lungo questa prospettiva; mi sembra cioè che il suo sia più che altro un affondo indirizzato alla storiografia positivista, incentra-



ta sul primato della diplomatica, volta separare la storia dalla filosofia, attenta al particolare piuttosto che al generale. Un affondo rivolto altresì a quella mentalità antiquaria, pedante, sterilmente erudita, puntigliosa, ancora circolante nel contesto intellettuale italiano e europeo dell'epoca. Ed è particolarmente significativo che proprio Cassese finisca, a conclusione del suo articolo, per posizionarsi nello stesso spazio concettuale entro cui Croce aveva formulato le sue riflessioni: i documenti, fa notare Cassese, che siamo soliti suddividere in due distinte classi – le manifestazioni delle *res gestae*, le carte d'archivio, e le *istoria rerum gestarum*, i libri – in realtà perdono ogni loro specificità e finiscono per annullare i loro tratti distintivi se considerati “nell'unità della vita dello spirito, che comprende necessariamente sotto di sé un'infinita varietà e molteplicità di azioni”: quello che appare come “dualismo non è altro che semplice successione di due distinti momenti o atti, i quali traggono sì origine da un'unica fonte, ma sono di diversa natura e si dirigono verso due diverse finalità”: in tal senso “il dualismo tra narrazione e documento o tra materiale bibliografico e materiale archivistico è annullato” (Cassese 1949, 36).

## 6. Da archivista “puro” a “scienziato”

Che l'archivistica dovesse emanciparsi da quella precettistica che finiva per ridimensionarne il portato, riducendola ad una sorta di disciplina del fare più che del pensare; che dovesse cioè ampliare il proprio raggio d'azione oltre quei principi e quelle regole che riguardavano la sola conservazione materiale dei documenti; mi sembra essere un'idea che non apparteneva soltanto a Croce, ma che era patrimonio comune anche di un gruppo non marginale di archivisti suoi contemporanei. Fra questi Armando Saporì, poi professore di storia economica nell'università Bocconi. Saporì archivista di Stato a Firenze è quello che osserva con occhio irriverente e canzonatorio, per alcuni versi ancora più tagliente di quello di Croce, alcuni dei suoi colleghi applicare pedissequamente i principi di quell'archivistica “pura” di casanoviana impostazione. L'archivista “puro” – da lui efficacemente tratteggiato in quella sorta di diario morale e intellettuale che è *Mondo finito* (Saporì [1946] 1971) – è colui cui poco o nulla importa il contenuto delle carte che maneggia, convinto com'è che il suo lavoro consista esclusivamente nell'attribuire un ordine numerico alle scritture: distinguendo i fascicoli pieni da quelli vuoti, ricondizionando le carte in camicie nuove, ricopiando su di esse i vecchi titoli riportati sugli incartamenti, collocando i documenti in scatole da disporre, perfettamente allineate, sugli scaffali dei depositi, in attesa che altri le studino (Saporì 1949; Fabbri 2018). E se poi il contenuto non corrispondeva alla leggenda, questa non era una questione che lo interessava più di tanto: questo genere di archivista, pur vivendo in mezzo alle memorie del passato, si mostrava infatti del tutto estraneo al “fascino” del loro “segreto”.

A questa immagine inconcludente dell'archivista, molto simile a quella di un acculturato magazzino, Saporì contrapponeva quella dello “scienziato”, dotato di “tutt'altra mentalità”. Per l'archivista “scienziato” ordinare e inventariare un complesso documentario significa far “rivivere” le carte che ne facevano parte, traendo da esse storie di sangue e carne. L'archivio per Saporì rappresentava un corpo morto solo agli occhi di chi non era in grado di coglierne il reale significato: “basta un soffio di pensiero perché si muova, torni a soffrire e a godere, e ti riveli l'esperienza di secoli” (Saporì 1949, 130). La vitalità di un archivio dipendeva dalle domande che avrebbe saputo porgli l'archivista: interrogativi volti a conoscere le storie di amore e odio, di pietà e dolore che nelle

pieghe delle carte giacciono in attesa di essere riscoperte: storie che il documento ti svela “come a un confessionale se lo sai interrogare con intelligenza e cuore” (Ibidem, 130). Parole significative quelle utilizzate da Saponi, anche perché evocano molte delle più suggestive pagine dell'*Apologie de l'histoire di Marc Bloch*, e in particolare quella dello storico-orco, che sa “che là dove fiuta carne umana, là è la sua preda” (Bloch 1993, 23). Parole ancor più significative in quanto più che indirizzarci verso campi semantici che si rifanno al concetto di morte – e dunque all'archivio inteso come una materia rigida, immodificabile, granitica – ci guidano al contrario verso l'immagine della vita: “soffio”, “pensiero”, “muoversi”, “tornare”, “soffrire”, “godere”, “amore”, “odio”, e via discorrendo. Parole che ci rimandano ancora una volta a Bloch e agli storici delle *Annales*, verso la loro idea di metodologia della ricerca storica e al loro concetto allargato di documento (Macry 1992). Una convergenza di vedute, quella che si scorge tra Saponi e Bloch, che potrebbe apparire come una sorta di prestito intellettuale da parte dell'archivista nei confronti dello storico, se non fosse che *Mondo finito* venne dato alle stampe nel 1946 e l'*Apologie* uscì invece postuma solo nel 1949, grazie all'intervento di Lucien Febvre. L'idea che l'archivista possa assurgere al ruolo di scienziato prende forma in Saponi principalmente dalla sua esperienza di lavoro presso l'Archivio di Stato di Firenze, al fianco di archivisti di particolare spessore intellettuale, quali Antonio Panella, Antonio Anzillotti e Bernardino Barbadoro; archivisti che non si limitarono a rincorrere l'ordine originario delle carte, riproducendolo in un inventario, ma si interrogarono sulla natura dei complessi documentari, sulle modalità della loro produzione e sul perché solo certe scritture sia sopravvissute giungendo fino a noi. Ma c'è di più. Il loro lavoro si svolgeva “fraternizzando” con gli storici, evitando condizioni di isolamento intellettuali, tenendosi aggiornati con le letture, e soprattutto mostrandosi pronti a modificare i propri convincimenti con l'evolversi degli studi e della ricerca. L'archivista scienziato è “sostanzialmente uno storico che usa i documenti come fonti” (Moscadelli 2018), e la fitta rete delle relazioni epistolari che Saponi intrattenne, già da quando era ancora un giovane archivista, con il *milieu* dell'avanguardia storiografica francese sembrerebbe confermarlo (Franceschini 2018; Moretti 2018; Moscadelli 2018). Intenso è lo scambio epistolare con Pirenne tra il 1928 e il 1935; altrettanto fitto quello con Bloch, che nel 1933 lo invitò a collaborare con la rivista delle *Annales*. Uno scambio di lettere che travalica la drammatica esperienza della seconda guerra mondiale, e che continuò negli anni successivi sia con Febvre che con Braudel e Le Goff. Dunque, più che parlare di un debito di riconoscenza contratto da Saponi verso le *Annales*, bisogna più prudentemente parlare di reciproci prestiti, di vicendevoli cessioni, di scambi reciproci (Poncet 2019). “Votre aimable carte, votre envoi et votre geste amical envers les Annales m'ont beaucoup touché. Nous faisons de notre mieux. Les témoignages de sympathie qui nous arrivent en ce moment nous sont un grand encouragement; et nous sommes particulièrement heureux d'avoir l'appui de savants étrangers, de votre compétence et de votre valeur”, scriveva infatti Bloch in un biglietto spedito a Saponi nel febbraio del 1939.<sup>4</sup> Parole, queste, che non si discostano per contenuto e partecipazione emotiva da quelle pronunciate pubblicamente da Febvre e Braudel a Milano nel 1957, in occasione della presentazione degli studi pubblicati in onore di Saponi, che si tenne nella monumentale cornice della Villa comunale: “Les historiens français ont une immense admiration à l'égard de votre oeuvre. Il vous considèrent (...) comme

---

<sup>4</sup> Una selezione delle lettere inviate a Bloch a Saponi, tra cui quella citata nel testo, è stata pubblicata da Stefano Moscadelli (2018a, 197-225).

le premier médiéviste d'Italie, dirais-je du monde" (Franceschi 2018, 75), ebbe modo infatti di affermare Braudel, alla presenza dei maggiori storici europei del tempo, da Salvemini a Postan, da Deoover a Renouard, a Sestan, a Cantimori. Ed è significativo che proprio alla suggestiva immagine dell'archivista scienziato si rifaccia Febvre, a decenni di distanza dall'uscita di *Mondo finito*, nella sua biografia intellettuale del grande narratore della rivoluzione francese, Michelet: uno storico ma anche archivista, storico della Rivoluzione francese e, contemporaneamente, conservatore degli archivi nazionali di Francia. Michelet ritratto nel suo lavoro d'archivio presenta infatti quegli stessi tratti distintivi che Sapori aveva reputato essere caratteristici dell'archivista scienziato: "Per scoprire gli archivi, rianimarli, metterli a frutto ci volevano poeti e non guardiani", scrive infatti Febvre<sup>5</sup>. Ed aggiunge: "Ma la storia, cosa usciva di lì? Poi è venuto lui Michelet (...), quel poeta, come si dice con disdegno, quel gran poeta, in effetti il più grande lirico del nostro XIX secolo – è lui che fece rientrare gli archivi nella vita, che da carte morte, tirò fuori per primo sangue caldo e vivo" (Mastrogregori 1995, 112-113).

## 7. La quadratura del cerchio

Parole destinate alla longevità, quelle pronunciate da Sapori, il cui eco sembra aver attraversato tutta la seconda metà del Novecento, tanto che se ne coglie ancora il suono in molta parte della letteratura archivistica d'avanguardia prodotta sul finire del secolo. Parole emblematiche di un decisivo cambio di passo: le riflessioni affidate da Sapori a *Mondo finito* si pongono ben al di là da quell'idea casanoviana di un'archivistica pura, intrisa di speculazioni astratte e dibattiti dottrinari, e consentono in pari tempo di superare quell'immagine dell'archivistica consegnataci da Cencetti come storia esclusiva del soggetto produttore. Ma c'è di più: nelle pagine di Sapori l'interrogativo che aveva contrapposto gli archivisti agli storici sembra d'un tratto dissolversi. Se sia l'archivistica una scienza ausiliaria della storia o la storia una scienza ausiliaria dell'archivistica appare come un dilemma ormai superato. Ce lo conferma fra l'altro il volume della Zanni Rosiello *Archivi e memoria*, che rappresenta una sorta di grande *summa* dell'intenso dibattito sviluppatosi intorno agli archivi in quel lungo arco di tempo che va dagli anni Venti agli anni Ottanta del Novecento. Una sorta di sintesi critica della riflessione che aveva caratterizzato l'Italia fra le due guerre e l'Italia della prima fase repubblicana. Una visione della disciplina dove l'idea che l'archivista sia uno storico diventa via via più nitida. Una convinzione questa cui era già giunto anche Ruggero Moscati, prima archivista e poi storico di professione, che nel 1967, in occasione del primo congresso della società degli storici italiani, prendendo la parola non dimenticò di precisare, ad esordio del suo intervento, che un buon archivista è anche storico, in quanto non è "semplice conservatore materiale delle scritture, che fornisce allo storico per l'utilizzazione (...) ma è egli stesso, nel suo lavoro, uno storico" (Moscati 1967). Alle parole di Antonio Pesce, alto dirigente del ministero dell'interno, che nel 1906 invitava gli archivisti di Stato a rifuggire le lusinghe "dell'affascinante Clio" (Pesce 1906), se ne sostituirono altre di opposto significato. Parole nuove che invitavano gli archivisti a confrontarsi con l'evolversi delle metodologie e dei temi della ricerca storica, ma conservando

---

<sup>5</sup> La citazione, tratta dal corso tenuto da Febvre su "Michelet et la Renaissance", è stata ripresa dalla traduzione di Massimo Mastrogregori (Mastrogregori 1995, 112-113).

inalterata la propria identità professionale, dato che “gli strumenti inventariali (...) sono tuttora ritenuti una delle prestazioni più qualificanti del mestiere dell’archivista” (Zanni Rosiello 1987, 154-155). L’archivista che si apre alla storia non è dunque quello che rincorre “affannosamente gli umori peraltro mutevoli di questo o quello studioso, di questa o quella scuola” (Ibidem, 153); si tratta al contrario di un archivista che dialoga con gli storici nella individuazione di nuove forme di memorie-fonti, diverse da quelle di cui si è solitamente occupato, ma senza trascurare la parte più significativa e qualificante del suo mestiere: l’ordinamento e l’inventariazione dei complessi documentari. È la stessa Zanni Rosiello a farcelo presente, precisando che “redigere strumenti inventariali ispirati al ‘metodo storico’, che tengano quindi conto del rapporto sfasatura tra soggetto-istituto produttore e modo in cui è stata organizzata e trasmessa la relativa memoria documentaria, è quanto si richiede a chi è, o vuol diventare, un archivista di tutto rispetto” (Ibidem, 155). Il lavoro dell’archivista non più inteso in una dimensione puramente amministrativa, ma volto a produrre strumenti di ricerca da intendersi come veri e propri “libri di storia”: l’archivista dunque non più considerato solo come un conservatore materiale di antiche carte, loro custode e loro guardiano, ma quale comunicatore e intermediario culturale, così come fu sottolineato dallo stesso Giuliano Catoni, per il quale saper condurre un’“analisi archeologica dell’archivio e dei suoi eventuali ordinamenti precedenti” costituiva una delle competenze imprescindibili del mestiere dell’archivista (Catoni 1983-1984, 155). Significative a tal proposito sono le parole che Federico Valacchi dedica alla stagione di studi archivistici che si colloca tra il primo e il secondo Novecento: feconda di ripensamenti, riflessioni, negoziazioni dottrinarie a cavaliere tra “vecchio” e “nuovo”, tra tradizione e innovazione; una stagione destinata ad aprire la strada a nuove meditazioni, nuove valutazioni e ulteriori cambiamenti. Così infatti scrive Valacchi: “certo la voce viva di Pavone non mancherà di risuonare, qua e là, ma solo per ricordare come lui e altri studiosi di calibro simile, come ad esempio Filippo Valenti, avessero già intravisto molto di quello che sarebbe successo dopo di loro, quasi per illuminazione, e avessero intuito, moltissimi anni prima del manifestarsi concreto di certi fenomeni, che l’archivistica andava dritta verso scenari che potremmo definire di una discontinua continuità, ferma nei suoi assetti etici e metodologici di fondo e, al tempo stesso, decisamente fluida nel modellarsi intorno a inevitabili e talvolta imprevedibili trasformazioni complessive della società” (Valacchi 2019, 161-162).

## Riferimenti bibliografici

- Balestracci, Duccio. 2018. "Comunicare le passioni. Armando Sapori, fra carta e microfono." In *Armando Sapori*, a cura di Stefano Moscadelli e Marzio A. Romani, 125-153. Milano: EGEA.
- Bloch, Marc. 1993. *Apologia della storia o mestiere di storico*. Torino: Einaudi.
- Casanova, Eugenio. 1928. *Archivistica*. Siena: Lazzeri.
- Casanova, Eugenio. 1929. "Archivio e archivistica." In *Enciclopedia Treccani*: [https://www.treccani.it/enciclopedia/archivio-e-archivistica\\_%28Enciclopedia-Italiana%29/](https://www.treccani.it/enciclopedia/archivio-e-archivistica_%28Enciclopedia-Italiana%29/).
- Cassese, Leopoldo. 1949. "Intorno al concetto di materiale archivistico e materiale bibliografico." *Notizie degli Archivi di Stato*, 34-41.
- Catoni, Giuliano. 1983-1984. "L'inventario e la guida dell'archivio: la pubblicazione." *Archiva ecclesiae* 26-27: 151-162.
- Cencetti, Giorgio. (1935-1963) 1970. *Scritti archivistici*. Roma: Il Centro di Ricerca.
- D'Angiolini, Piero, e Claudio Pavone. 1981. "Introduzione." In *Guida generale degli Archivi di Stato italiani I*, 1-31. Roma: Ministero per i beni culturali e ambientali.
- D'Angiolini, Piero, e Claudio Pavone. (1973) 2004. "Archivi e orientamenti storiografici." In *Intorno agli archivi e alle istituzioni. Scritti di Claudio Pavone*, a cura di Isabella Zanni Rosiello, 299-329. Roma: Ministero per i beni e le attività culturali.
- Fabbri, Lorenzo. 2018. "Un 'archivista scienziato' nell'Italia fascista." In *Armando sapori*, a cura di Stefano Moscadelli e Marzio A. Romani, 37-56. Milano: EGEA.
- Franceschini, Franco. 2018. "Armando Sapori e la storiografia internazionale." In *Armando Sapori*, a cura di Stefano Moscadelli e Marzio A. Romani, 73-100. Milano: EGEA.
- Franzese, Paolo (a cura di). 2023. *Filippo Valenti tra filosofia e archivistica*. Roma: Ministero della cultura, Direzione generale archivi.
- Lodolini, Armando. 1957. "Un sessantennio di Archivistica nell'opera di Eugenio Casanova." *Rassegna degli Archivi di Stato* 17: 220-242.
- Macy, Paolo. 1992. "Il passato e gli storici". In *La società contemporanea*, 13-59. Bologna: Il Mulino.
- Mastrogregori, Massimo. 1995. *Il manoscritto interrotto di Marc Bloch*, Pisa-Roma: Istituto editoriali e poligrafici Internazionali.
- Mineo, Leonardo. 2015. "Un nuovo 'manuale' di archivistica: alcune riflessioni." *Archivi* 1: 130-139.
- Moretti, Mauro. 2018. "Attorno ad armando Sapori: reti accademiche e storiografiche." In *Armando sapori*, a cura di Stefano Moscadelli e Marzio A. Romani, 101-124. Milano: EGEA.
- Moscadelli, Stefano. 2018a. "Armando Sapori: la biografia e l'archivio". In *Armando Sapori*, a cura di Stefano Moscadelli e Marzio A. Romani, 3-56. Milano: EGEA.
- Moscadelli, Stefano. 2018b. "Le lettere." In *Armando Sapori*, a cura di Stefano Moscadelli e Marzio A. Romani, 197-225. Milano: EGEA.



- Moscadelli, Stefano. 2023. "Archival science and research tools: methodological aspects from the beginnings of the historical method to standards." *JLIS.it* 14 (3): 15-26. <https://www.jlis.it/index.php/jlis/article/view/531/511>.
- Moscatti, Ruggero. 1967. "L'archivistica." *Clio* 4: 554-565.
- Pavone, Claudio. 1965. "Gli archivi e la ricerca scientifica." *Rassegna degli Archivi di Stato* 2: 299-300
- Pavone, Claudio. (1970) 2004a. "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, 71-75. Roma: Ministero per i beni e le attività culturali.
- Pavone, Claudio. (1961-1970) 2004b. "Mutamenti istituzionali e storie di carte." In *Intorno agli archivi e alle istituzioni. Scritti di Claudio Pavone*, a cura di Isabella Zanni Rosiello, 43-67. Roma: Ministero per i beni e le attività culturali.
- Pavone, Claudio. (2000-2001) 2004c. "Ripensando agli archivi e agli archivisti." In *Intorno agli archivi e alle istituzioni. Scritti di Claudio Pavone*, a cura di Isabella Zanni Rosiello, 377-385. Roma: Ministero per i beni e le attività culturali.
- Pesce, Angelo. 1906. *Notizie sugli Archivi di Stato comunicate alla VII riunione bibliografica italiana*. Roma: Tipografia delle mantellate.
- Petrucchi, Armando. 1978. "Casanova Eugenio". In *Dizionario biografico degli italiani* 21. Istituto della enciclopedia italiana: [https://www.treccani.it/enciclopedia/eugenio-casanova\\_%28Dizionario-Biografico%29/](https://www.treccani.it/enciclopedia/eugenio-casanova_%28Dizionario-Biografico%29/).
- Poncet, Olivier. 2019. "Archives et histoire: dépasser les tournants." *Annales. Histoire, Sciences Sociales* 3-4: 713-743.
- Prosperi, Adriano. 1988. "Storia moderna: il 'privato' e la politica." *L'informazione bibliografica* 2: 237-255.
- Sapori, Armando. (1946) 1971. *Mondo finito*. Milano-Varese: Istituto Editoriale Cisalpino.
- Valacchi, Federico. 2018. "Archivistica, parola plurale." *Archivi* 1: 5-28.
- Valacchi, Federico. 2019. "Ripartire da Pavone. Spunti di archivistica." *Parolechiave* 1-2: 161-175. <https://www.rivisteweb.it/doi/10.7377/96774>.
- Valacchi, Federico. 2021. *Gli archivi tra storia uso e futuro*. Milano: Editrice Bibliografica.
- Valacchi, Federico. 2023. *La verità di carta. A cosa servono gli archivi?*. Perugia: Graphe.it Editori.
- Valenti, Filippo. (1969) 2000a. "A proposito della traduzione italiana dell'«Archivistica» di Adolf Brenneke." In Filippo Valenti, *Scritti e lezioni di Archivistica, Diplomatica e Storia istituzionale*, a cura di Daniela Grana, 3-16. Roma: Ministero per i beni e le attività culturali.
- Valenti, Filippo. (1975-1976) 2000b. "Nozioni di base per un'Archivistica come euristica delle fonti documentarie." In Filippo Valenti, *Scritti e lezioni di Archivistica, Diplomatica e Storia istituzionale*, a cura di Daniela Grana, 135-224. Roma: Ministero per i beni e le attività culturali.

Valenti, Filippo. (1975) 2000c. “Parliamo ancora di Archivistica.” In Filippo Valenti, *Scritti e lezioni di Archivistica, Diplomatica e Storia istituzionale*, a cura di Daniela Grana, 45-81. Roma: Ministero per i beni e le attività culturali.

Valenti, Filippo. (1981) 2000d. “Riflessioni sulla natura e struttura degli archivi.” In Filippo Valenti, *Scritti e lezioni di Archivistica, Diplomatica e Storia istituzionale*, a cura di Daniela Grana, 83-113. Roma: Ministero per i beni e le attività culturali.

Valenti, Filippo. (1981) 2000e. “Riflessioni sulla natura e struttura degli archivi.” In Filippo Valenti, *Scritti e lezioni di Archivistica, Diplomatica e Storia istituzionale*, a cura di Daniela Grana, 83-113. Roma: Ministero per i beni e le attività culturali.

Valenti, Filippo. (1989) 2000f. “Un nuovo libro su archivi e archivisti.” In Filippo Valenti, *Scritti e lezioni di Archivistica, Diplomatica e Storia istituzionale*, a cura di Daniela Grana, 115-132. Roma: Ministero per i beni e le attività culturali.

Vitali, Stefano. 2014. “Archivi, fondi, contesti: una riflessione che continua.” In *L'apporto del pensiero di Filippo Valenti alle discipline archivistiche*, a cura di Euride Fregni, 59-70. Roma: Ministero dei beni e delle attività culturali e del turismo.

Zanni Rosiello, Isabella. 2004. “Un archivista, uno storico.” In *Intorno agli archivi e alle istituzioni. Scritti di Claudio Pavone*, a cura di Isabella Zanni Rosiello, 7-31. Roma: Ministero per i beni e le attività culturali.

Zanni Rosiello, Isabella. (1974) 2000a. “Archivisti e storici: un confronto a distanza.” In *L'archivista sul confine. Scritti di Isabella Zanni Rosiello*, a cura di Carmela Binchi e Tiziana Di Zio, 389-394. Roma: Ministero per i beni e le attività culturali.

Zanni Rosiello, Isabella. (1980) 2000b. “Archivi e potere a Bologna nel Settecento.” In *L'archivista sul confine. Scritti di Isabella Zanni Rosiello*, a cura di Carmela Binchi e Tiziana Di Zio, 389-394. Roma: Ministero per i beni e le attività culturali.

Zanni Rosiello, Isabella. (1982) 2000c. “Sul mestiere di archivista.” In *L'archivista sul confine. Scritti di Isabella Zanni Rosiello*, a cura di Carmela Binchi e Tiziana Di Zio, 371-388. Roma: Ministero per i beni e le attività culturali.

Zanni Rosiello, Isabella. 1987. *Archivi e memoria storica*. Bologna: il Mulino.

# GIS technology as a geographical method in applied library research: interpreting emerging issues through an interdisciplinary perspective

Camilla Quaglieri<sup>(a)</sup>

a) Sapienza University of Rome, <https://orcid.org/0000-0001-6261-4321>

**Contact:** Camilla Quaglieri, [quaglieri.camilla97@gmail.com](mailto:quaglieri.camilla97@gmail.com)

**Received:** 18 January 2024; **Accepted:** 22 March 2024; **First Published:** 15 May 2024

## ABSTRACT

Thanks to current national investigations, research applied to libraries in Italy has an unprecedented amount of data from which are emerging new complex issues and some worrying contradictions. In this framework lies the need to design structured in-depth investigations using innovative interpretative tools. Following an interdisciplinary approach, the article therefore proposes a methodological reflection on the opportunities offered by the integration between geographical knowledge and librarianship. Starting from the comparison between the main application purposes of GIS technology (Geographic Information System) adopted so far, both nationally and internationally, with some new experiments conducted in recent years, the contribution intends to present the potential of a method geographical gis-based for the interpretation of the relationships that link libraries to their context of action.

## KEYWORDS

Methodology; GIS technology; Applied research; Interdisciplinarity.

# La tecnologia GIS come metodo geografico nella ricerca applicata alle biblioteche: interpretare le questioni emergenti attraverso una prospettiva interdisciplinare

## ABSTRACT

Grazie alle attuali indagini a carattere nazionale, la ricerca applicata alle biblioteche in Italia dispone di un'inedita quantità di dati da cui stanno emergendo nuove complesse questioni e alcune preoccupanti contraddizioni. In questo scenario si colloca l'esigenza di progettare indagini di approfondimento strutturate ricorrendo a strumenti interpretativi innovativi. Seguendo un approccio interdisciplinare, l'articolo propone pertanto una riflessione metodologica sulle opportunità offerte dall'integrazione tra il sapere geografico e la biblioteconomia. A partire dal confronto tra i principali scopi applicativi della tecnologia GIS (*Geographic Information System*) adottati finora, sia in ambito nazionale che internazionale, con alcune nuove sperimentazioni condotte negli ultimi anni, il contributo intende presentare le potenzialità di un metodo geografico *gis-based* per l'interpretazione delle relazioni che legano le biblioteche al proprio contesto d'azione.

## PAROLE CHIAVE

Metodologia; Tecnologia GIS; Interdisciplinarietà; Ricerca.

## Premessa

In Italia, per diverso tempo, alla ricerca applicata in ambito biblioteconomico è mancato un sistema informativo di dati che descrivesse le biblioteche a livello nazionale. Nonostante la comunità scientifica abbia percepito già dalla fine degli anni Novanta l'esigenza di misurare la qualità delle biblioteche in termini di efficacia ed efficienza<sup>1</sup>, è solamente negli ultimi anni infatti che questi luoghi della cultura sono diventati protagonisti di un inedito interesse da parte dell'ISTAT (Istituto nazionale di statistica). Si fa riferimento al Censimento delle biblioteche italiane<sup>2</sup> e all'indicatore dedicato alla fruizione delle biblioteche presente nel *Rapporto sul Benessere Equo e Sostenibile* (BES)<sup>3</sup>, ovvero alle due indagini che restituiscono oggi dati strutturati di *input* e *output* a livello nazionale.

Per rendere operative queste conquiste, si dimostra indispensabile interpretare i risultati di tali indagini contestualizzandoli e mettendoli in relazione tra loro (Faggiolani 2021). In particolare, una delle questioni che emerge con forza riguarda un aspetto dalle pericolose implicazioni. Nel corso delle prime tre edizioni del Censimento dell'ISTAT il numero di strutture aperte al pubblico è aumentato di circa 400 unità, passando dalle 7.425 del 2019 alle 7.886 del 2021; l'ultima edizione del Censimento, pubblicata recentemente, non smentisce inoltre questo *trend* per cui il totale delle biblioteche nel 2022 ha raggiunto le 8.131 unità (ISTAT 2021; 2021; 2022; 2023). Parallelamente, tuttavia, i risultati del Rapporto BES mostrano che la percentuale di persone con più di 3 anni che hanno dichiarato di frequentare le biblioteche nel corso del 2021 è stata solamente del 7,4% e il Rapporto 2022, pur registrando un aumento (10,2%), evidenzia che l'abitudine degli italiani a recarsi in biblioteca rimane inferiore al periodo pre-pandemico (15,3%) (ISTAT 2021; 2022; 2023). La relazione tra questi dati rivela un'evidente contraddizione: nonostante le biblioteche siano il luogo della cultura quantitativamente più presente sul territorio nazionale (Arosio e Federici 2022, 51-74), il loro ruolo per la qualità della vita dei cittadini non sembra essere riconosciuto.

Anche a partire da tale contraddizione, deriva l'inevitabile riflessione sulle trasformazioni che dovrebbero investire le biblioteche affinché queste diventino un dispositivo a supporto del sistema del benessere delle persone. Questa urgenza di ripensamento delle biblioteche rappresenta una delle attuali priorità della ricerca applicata e si inserisce nella riflessione più ampia – ormai matura – sul ruolo sociale della cultura<sup>4</sup>.

---

<sup>1</sup> La riflessione sull'evoluzione della ricerca applicata in ambito biblioteconomico è stata oggetto del convegno *Misurazione e valutazione delle biblioteche: ne è valsa la pena? Cambiamenti, impatti e racconti di 30 anni di lavoro comune*, organizzato dal CSBNO (Culture Socialità Biblioteche Network Operativo) in occasione dei suoi 25 anni di attività; il video della giornata è disponibile al link: <https://www.youtube.com/watch?v=XLNerHq219g>; si veda anche (Bertazzoli 2022).

<sup>2</sup> Il censimento non include le biblioteche scolastiche e delle università ed è inserito tra gli obiettivi del *Protocollo d'intesa per lo sviluppo del sistema informativo integrato su istituti e luoghi di cultura*, sottoscritto da Ministero per i beni e le attività culturali e per il turismo (MiBACT, ora Ministero della cultura, MiC), Conferenza delle Regioni, Regioni e Province autonome. La partecipazione delle biblioteche di enti ecclesiastici è stata invece possibile grazie alla firma di una lettera di intenti con la Conferenza Episcopale Italiana (CEI).

<sup>3</sup> Il *Rapporto sul Benessere equo e sostenibile* dell'ISTAT rappresenta un importante strumento di monitoraggio della qualità della vita dei cittadini italiani e del loro benessere non in una prospettiva esclusivamente economica, ma soprattutto sociale. L'indicatore *Fruizione delle biblioteche* è stato introdotto nel dominio *Istruzione e formazione* a partire dal 2020.

<sup>4</sup> Questo processo di ripensamento dell'identità delle biblioteche è stato oggetto di una riflessione comune durante il convegno *Libro città aperta* realizzato a Milano a settembre 2023 per iniziativa della Fondazione Mondadori. Per approfondire al seguente link è disponibile il video dell'intera giornata: <https://www.fondazionemondadori.it/evento/25698/>. A conclusione della giornata sono state formulate cinque tesi per le biblioteche del futuro, cfr. (Faggiolani 2023).

I fattori determinanti che incidono sulla qualità di vita e di salute delle persone non dipendono infatti esclusivamente da condizioni economiche o mediche, ma anche da aspetti sociali (WHO 2009). Tra questi rientrano senza dubbio le opportunità di partecipazione culturale che possono influenzare numerose dimensioni del benessere di una comunità contribuendo alla crescita personale, alla capacità critica di interpretazione della realtà, all'esercizio attivo della democrazia o all'educazione emotiva. Le infrastrutture culturali rappresentano quindi degli importanti strumenti di *welfare* capaci di innescare processi di attivazione sociale nei territori, di aggregazione comunitaria e di capacitazione collettiva (Paltrinieri 2022).

Da questa visione emerge chiaramente che il benessere è un sistema non solo complesso e multidimensionale, ma anche 'radicato' in ogni specifico contesto (territoriale, urbano, sociale ecc.). Secondo questa prospettiva, per risolvere la contraddizione ricordata non sembra possibile osservare la biblioteca senza riflettere anche sulle relazioni che la legano ad altri attori del contesto in cui opera e con i quali interagisce. Nel progettare la possibile trasformazione delle biblioteche è pertanto necessario capire quali condizioni riconducibili al contesto d'azione delle biblioteche siano in grado di incidere sulla loro vitalità, influenzando così la loro capacità di rispondere ai bisogni di chi le frequenta. In altri termini, alla progettazione e valutazione dell'impatto che le biblioteche possono produrre (Faggiolani 2019), si dovrebbe affiancare anche l'interpretazione e il monitoraggio dell'impatto che queste subiscono per effetto dell'interazione con gli altri agenti del contesto in cui operano (Quagliari 2023).

A partire da questo scenario trasformativo, da un punto di vista metodologico non si può evitare di riflettere anche sui cambiamenti che interessano la biblioteconomia in quanto disciplina. Come sarà argomentato meglio più avanti, proprio in questa direzione si colloca l'esigenza di integrare la tecnologia GIS (*Geographic Information System*) nella ricerca applicata alle biblioteche. Questa tecnologia – e il pensiero geografico più in generale – si dimostra infatti particolarmente promettente per la sua trasversalità disciplinare che le consente di connettere fenomeni diversi restituendoli in un quadro interpretativo organico.

## **L'interdisciplinarietà come principio dell'innovazione scientifica**

Nel *framework* appena descritto si possono riconoscere alcuni aspetti considerabili dei veri e propri presupposti per le future ricerche applicate alle biblioteche: la disponibilità di dati; la visione sistemica; l'inter-settorialità degli obiettivi.

Le indagini dell'ISTAT citate nella premessa rappresentano in questo senso un'infrastruttura informativa da cui non si può prescindere per riflettere sulle questioni emergenti che stanno interessando le biblioteche e che appaiono sempre più complesse e sfidanti. Come anticipato, inoltre, sono diverse le sollecitazioni che hanno reso maturo il dibattito sul ruolo sociale della cultura che intende le biblioteche come singole componenti di un macrosistema più complesso – quello del benessere –, costituito da altri luoghi della cultura, istituzioni pubbliche, agenti sociali del terzo settore, nonché dalle relazioni che legano tra loro questi attori (Faggiolani 2022). L'adozione di una visione sistemica per osservare le biblioteche all'interno del proprio contesto d'azione, ovvero la città, e in relazione al proprio territorio appare pertanto indispensabile per progettare interventi mirati e duraturi. Proprio a partire da una prospettiva sistemica di questo genere deriva l'impulso



ad ampliare i confini disciplinari della biblioteconomia ricorrendo a strumenti di indagine inediti e alternativi, ma anche capaci di contribuire a obiettivi intersettoriali.

Una delle principali implicazioni metodologiche rispetto a tali presupposti riguarda l'esigenza di costruire strategie di indagine che consentano di interpretare le connessioni tra fenomeni interni ed esterni alle biblioteche. In qualsiasi campo di studio, infatti, di fronte all'emergere di nuove domande di ricerca, determinate da un aumento estrinseco della complessità, si attiva un processo di innovazione dei metodi e degli strumenti disciplinari che spesso può trovare risposta in altri saperi scientifici. In questo senso, nella relazione tra saperi diversi sembra collocarsi il principio stesso dell'innovazione scientifica che a sua volta si può considerare al contempo la causa e l'effetto di un approccio metodologico interdisciplinare. Questa ricorsività si traduce in una continua ridefinizione dei confini tra le discipline che trova un punto di equilibrio nell'individuazione di obiettivi comuni.

Esiste un documento particolarmente significativo per contestualizzare questa riflessione e che rivela quanto l'innovazione per la ricerca applicata non sia affatto un'esigenza nuova, bensì caratteristica della società contemporanea. In particolare, si tratta di un rapporto pubblicato dal Centro per la ricerca e l'innovazione educativa dell'OCSE (Organizzazione per la cooperazione e lo sviluppo economico), che restituisce gli interventi di un convegno tenutosi a Nizza nel 1970 sul tema *Interdisciplinarity: Problems of Teaching and Research in Universities* (OECD-CERI 1972). L'occasione da cui nacque il seminario fu di condividere i risultati emersi da un'indagine condotta sulle università di 20 dei Paesi membri dell'OCSE con lo scopo di definire per la prima volta da un punto di vista epistemologico il concetto di 'interdisciplinarietà'. Attraverso un corposo questionario venne chiesto a 250 esponenti della comunità accademica internazionale se ritenessero o meno la propria forma di insegnamento interdisciplinare. A questo fine i ricercatori del CERI (Centre for Educational Research and Innovation) formularono la definizione di alcuni termini-chiave, tra cui appunto il concetto di 'inter-disciplinarietà'. Leggendo la loro proposta, i criteri che caratterizzano questo approccio di insegnamento e di ricerca sono da ricondurre alla diversità tra le discipline interessate, alla natura del rapporto che le lega e al perseguimento di obiettivi comuni. Se anche in altri approcci (ad esempio la pluri-disciplinarietà) lo scambio di saperi avviene tra discipline appartenenti ad ambiti di studio differenti, la connessione tra i diversi saperi nel caso dell'interdisciplinarietà non consiste in una giustapposizione, ma in un'interazione profonda e reciproca di concetti, metodologie, strumenti e dati; questa precondizione rende l'interdisciplinarietà una modalità attraverso cui le singole discipline offrono un contributo, sia teorico che metodologico, al raggiungimento di obiettivi di indagine comuni.

Nel rapporto l'interdisciplinarietà viene infatti definita come:

An adjective describing the interaction among two or more different disciplines. This interaction may range from simple communication of ideas to the mutual integration of organising concepts, methodology, procedures, epistemology, terminology, data, and organisation of research and education in a fairly large field. An interdisciplinary group consists of persons trained in different fields of knowledge (disciplines) with different concepts, methods, and data and terms organised into a common effort on a common problem with continuous intercommunication among the participants from the different disciplines (Ibidem, 24).

Alla base della ricerca dell'OCSE, e poi del convegno di Nizza, c'è stata l'intuizione dei limiti delle singole discipline scientifiche rispetto all'aumento della complessità dei compiti a cui il sapere accademico era chiamato a trovare risposte, che già negli anni Sessanta era percepito come l'inizio di una radicale trasformazione della società (tecnologica, economica, istituzionale, culturale ecc.). L'intento perseguito in quella occasione è stato quindi di proporre nuovi modelli di insegnamento al fine di promuovere un'innovazione della ricerca applicata che permettesse di rispondere ai bisogni futuri. Per questo motivo, nell'introduzione del rapporto OCSE-CERI si può leggere che “the ‘inter-discipline’ of today is the ‘discipline’ of tomorrow” (Ibidem, 8).

Questo sguardo di rinnovamento rivolto al futuro che, come si è detto nella premessa, sta caratterizzando anche l'attuale dibattito biblioteconomico rappresenta l'orizzonte interdisciplinare verso cui indirizzare la ricerca applicata alle biblioteche.

## La tecnologia GIS nella ricerca applicata alle biblioteche: da strumento a metodo

La propensione a rinnovare le strategie di indagine proprie di una disciplina di fronte a esigenze conoscitive sempre più articolate e interconnesse caratterizza molti campi di ricerca. Il contributo che le singole discipline possono offrire in uno scenario interdisciplinare non appare però uguale per tutte. Si può riconoscere, infatti, come alcune discipline per loro natura svolgano un ruolo di ‘connettore’. Questo sembra essere il caso della geografia che in un quadro interdisciplinare può rappresentare, come scrive Spada, una “scienza delle relazioni orizzontali, ovvero che studia le relazioni tra i singoli fenomeni indagati da altre discipline” (Spada 2007, 52).

Questa interpretazione esplicita il collegamento tra la disciplina geografica e gli attuali obiettivi intersettoriali della biblioteconomia. La scienza geografica può offrire, infatti, oltre che strumenti adatti, anche una prospettiva innovativa per mettere in connessione l'azione delle biblioteche con il contesto in cui operano e in relazione ai fenomeni che le interessano nella dimensione macro-sistemica del benessere più volte richiamata. In particolare, tra gli strumenti geocartografici la tecnologia GIS, acronimo inglese traducibile come ‘Sistema Informativo Geografico’, rappresenta una promettente ‘lente di ingrandimento’ con cui poter meglio osservare le caratteristiche delle biblioteche italiane, le relazioni che le lega ad altri attori, progettarne e monitorarne i cambiamenti. La tecnologia GIS rappresenta, infatti, un potente strumento informatico che consente di gestire, analizzare e interpretare grandi quantità di dati georeferenziati<sup>5</sup> restituendoli attraverso rappresentazioni cartografiche (le tradizionali mappe). Questa potenzialità tecnologica è resa possibile grazie allo sviluppo di *software* capaci di garantire contemporaneamente alcune funzioni fondamentali, quali la redazione automatica di mappe (*automated mapping*), la gestione dei dati e la loro analisi spaziale. La componente *software*, tuttavia, non basterebbe da sola a rendere un GIS un sofisticato strumento di interpretazione di fenomeni complessi e sono pertanto indispensabili anche altre componenti tra cui un *database* strutturato di dati e il personale, ovvero l'insieme delle competenze teoriche e tecniche necessarie per il suo utilizzo.

---

<sup>5</sup> La geolocalizzazione, o georeferenziazione, di un oggetto nello spazio avviene attraverso l'attribuzione di una posizione univoca rispetto a un sistema di riferimento comune e la successiva trasformazione in una coppia di coordinate geografiche.

Per questa ragione, la tecnologia GIS non rappresenta solo un *software* per la creazione di mappe o un'evoluzione digitale delle attività cartografiche, bensì un insieme di strumenti che permette di gestire dati spaziali secondo specifiche metodologie e con lo scopo di rispondere a esigenze concrete rappresentando informazioni descrittive di elementi che esistono o di eventi che accadono nello spazio (Serravalli 2001, 23).

In queste caratteristiche risiede la trasversalità disciplinare della tecnologia GIS e permettono di comprendere quanto:

Queste capacità distinguono il GIS da altri sistemi informativi e lo rendono appetibile ad una gran quantità di soggetti come supporto alla spiegazione di avvenimenti, alla previsione di risultati ed alla pianificazione di strategie. Le possibilità di applicazione dei GIS ai vari settori della pianificazione, programmazione e gestione di territori e/o ambienti (esterni o confinanti) ne fanno uno strumento globale (Poletti 2001, 7).

Nonostante sia stato limitato all'area americana, si può riscontrare un interesse della biblioteconomia nei confronti di questa tecnologia già a partire dagli anni Settanta, nato grazie alla pubblicazione nel 1973 della prima raccolta di mappe tematiche dell'Illinois e del Missouri da parte dell'U.S. Bureau of census. Le biblioteche americane sono state, infatti, tra i primi soggetti a sfruttare gli imponenti progetti geocartografici realizzati dal governo degli Stati Uniti per monitorare e descrivere le caratteristiche demografiche della popolazione.

Per comprendere come la tecnologia GIS sia stata impiegata nella ricerca applicata alle biblioteche nei decenni successivi sono utili i risultati di una revisione della letteratura scientifica LIS di ambito internazionale (Bishop and Mandel 2010). La *review* è stata condotta su 34 articoli<sup>6</sup> e l'analisi complessiva ha restituito due differenti modalità di implementazione di questo strumento: "analyzing service area population" e "managing facilities and collections" (Ibidem, 539).

Nel dettaglio, le indagini descritte negli articoli della prima categoria riguardano studi di tipo territoriale e demografico utili da una parte a ricostruire quella che viene definita la *market area* di una biblioteca, dall'altra all'analisi dei pubblici attraverso la profilazione dell'utenza; l'obiettivo perseguito in questi casi consiste nel migliorare i servizi bibliotecari a partire dalle caratteristiche del territorio in cui la biblioteca opera (prettamente in relazione ad aspetti urbanistici quali la viabilità, la presenza di barriere naturali, la vicinanza con fermate dei trasporti pubblici ecc.) e sulla base della segmentazione e distribuzione della popolazione di riferimento. La seconda categoria di articoli, invece, comprende progetti in cui la tecnologia GIS è stata adottata per migliorare la fruibilità delle biblioteche grazie all'osservazione dell'uso degli spazi e dello scaffale aperto da parte degli utenti; il monitoraggio delle strutture architettoniche e della distribuzione delle collezioni con il supporto della tecnologia GIS permetterebbe in questo modo di ottimizzare gli interventi migliorativi di sistemazione interna delle sedi.

Contrariamente a quanto accaduto negli Stati Uniti, la riflessione italiana sul potenziale contributo dei GIS alle attività di misurazione e progettazione dei servizi è stata invece tardiva e non sembra essersi pienamente sviluppata. Sebbene nella letteratura di settore in Italia la pubblicazione del

---

<sup>6</sup> La ricerca dei contributi è stata effettuata su due database, ovvero il *Library Literature and Information Full Text* e il *Library, Information Science and Technology Abstracts (LISTA)*.

primo contributo su questo tema risalgia al 2005 (Venuda 2005a; 2005b), i risultati di una sperimentazione italiana della tecnologia GIS in biblioteca sono stati invece pubblicati a 10 anni di distanza (Dapiaggi 2014). L'obiettivo di tale indagine, condotta da Daniele Dapiaggi, consisteva nell'ottimizzare i percorsi di prestito interbibliotecario attraverso la creazione di mappe tematiche delle 29 strutture del Sistema Brianza Biblioteche (Dapiaggi 2017). La finalità applicativa più generale della ricerca non era solo di offrire uno strumento di gestione del singolo servizio studiato, ma di promuovere anche processi autovalutativi a partire dalla realizzazione di mappe per monitorare gli indicatori degli standard proposti dall'Associazione italiana biblioteche (AIB 2000).

A prescindere dal numero e dall'area geografica, queste testimonianze sui principali scopi applicativi dei GIS sperimentati nella ricerca sulle biblioteche confermano quanto questa tecnologia rappresenti uno strumento di gestione dei servizi bibliotecari nell'ottica di supporto ai *processi di decision making*.

Seguendo l'evoluzione della riflessione biblioteconomica sulle pratiche di valutazione, sembra tuttavia possibile immaginare un'applicazione più ambiziosa della tecnologia GIS nelle indagini dedicate alle biblioteche. Con la valutazione dell'impatto, che rappresenta l'effetto generato dall'interazione tra la biblioteca e il suo ambiente, l'interesse della ricerca applicata si è infatti inevitabilmente aperto verso l'esterno delle biblioteche attribuendo al loro contesto d'azione un rinnovato interesse. A partire da questa visione valutativa appare ineludibile ampliare la prospettiva di indagine e chiedersi anche quali effetti produca il contesto sulla biblioteca e con quali modalità tale cambiamento avvenga.

L'integrazione del pensiero geografico all'interno delle pratiche di valutazione dell'impatto inverso, quindi degli effetti subiti e non solo generati dalle biblioteche, permetterebbe di valorizzare la capacità insita nei GIS di connettere un qualsiasi sistema interagente con il proprio ambiente, come la biblioteca, ai fenomeni che si manifestano nel contesto in cui questo opera.

Lo sviluppo interno al dibattito geografico sulla tecnologia GIS rivela come tale integrazione sia possibile. Nel periodo compreso tra gli anni Sessanta – quando venne coniato il termine GIS – e i primi decenni del XXI secolo, geografi ed esperti di cartografia si sono interrogati sulla natura di questa tecnologia arrivando a riconoscerla alla fine degli anni Novanta come una scienza autonoma. La fondazione di centri di ricerca americani, infatti, contribuì notevolmente a sviluppare teorie e metodologie caratteristiche e distintive di una 'scienza informativa geografica'. Come sostiene Goodchild, uno dei principali protagonisti della teorizzazione del concetto di GiScience è stato David Mark (Goodchild 2010, 6), definendola come “the basic research field that seeks to redefine geographic concepts and their use in the context of geographic information systems” (Mark 2000, 47).

L'evoluzione epistemologica che ha portato alla GiScience, senza addentrarsi nelle specificità di questo dibattito, assume una certa importanza per la biblioteconomia perché offre l'opportunità di considerare la tecnologia GIS nella ricerca applicata alle biblioteche non più 'solo' come uno strumento, ma anche come un possibile metodo di indagine. Questa transizione metodologica sarebbe possibile proprio riconoscendo piena autonomia alla GiScience in quanto disciplina e alla tecnologia GIS in quanto insieme di procedure interpretative con cui l'annessa scienza si esprime. Da un punto di vista metodologico, un metodo rappresenta un “sistema definito, creato dal ricercatore, per intraprendere una ricerca empirica di tipo investigativo: è l'approccio complessivo, spesso definito ‘strategia di ricerca’” (Pickard 2010, 45). L'adozione di un particolare metodo in-

fluenza l'insieme delle decisioni metodologiche che un ricercatore prende nella costruzione del disegno complessivo del progetto di ricerca e dovrebbe dipendere sempre dalle domande conoscitive a cui si intende rispondere. Considerare la GiScience un metodo di ricerca richiede pertanto non solo di padroneggiare le competenze tecniche necessarie per l'uso della tecnologia GIS – già indispensabile per il suo utilizzo come strumento di ricerca –, bensì di assumere uno 'sguardo' geografico durante l'intero processo di indagine, ovvero nelle diverse fasi di raccolta, analisi e interpretazione dei dati.

In questo modo, il focus di indagine tende a spostarsi dallo studio di singoli oggetti e concetti (le strutture, i servizi, le collezioni, l'accessibilità, la fruizione ecc.), all'interpretazione del significato delle relazioni spaziali tra fenomeni che interessano 'anche' le biblioteche. Lo slittamento da strumento a metodo non riguarda quindi tanto le procedure di applicazione della tecnologia GIS, quanto le finalità conoscitive perseguite. Si passerebbe così da una giustapposizione pluridisciplinare di strumenti, a un'integrazione profonda di saperi all'interno di un più ampio quadro interdisciplinare.

Le sperimentazioni descritte nel prossimo paragrafo possono aiutare a rendere evidente come la scelta di un 'metodo geografico' *gis-based* abbia permesso di rispondere all'esigenza di ricostruire in una dimensione territoriale le relazioni tra le biblioteche e gli altri attori del loro contesto d'azione, nonché contribuire all'interpretazione dei complessi fenomeni innescati da tali relazioni. L'opportunità di visualizzare i dati in una dimensione spaziale e su più livelli territoriali ha rivelato, infatti, delle connessioni che sarebbero rimaste altrimenti nascoste.

## Sperimentazioni a confronto

La proposta di prevedere l'evoluzione della tecnologia GIS da strumento a metodo strutturato di indagine rappresenta l'esito di una riflessione scaturita dall'esperienza maturata nel corso di alcune sperimentazioni realizzate nell'ambito dell'attività di ricerca del Laboratorio di biblioteconomia sociale e ricerca applicata alle biblioteche (BIBLAB) presso La Sapienza Università di Roma. Si tratta, in particolare, di uno studio condotto sulla chiusura di 330 biblioteche avvenuta nel corso del 2020 per motivi non legati alla gestione dell'emergenza pandemica (Quaglieri 2023) e dell'indagine sulla polifunzionalità di 292 biblioteche riferita invece ai dati ISTAT 2019<sup>7</sup> (Faggiolani e Quaglieri 2023; Faggiolani, Federici e Quaglieri 2023). Sebbene i fenomeni indagati siano stati differenti, entrambe le ricerche hanno previsto l'adozione di un approccio interdisciplinare e di una metodologia mista (Amaturo e Punziano 2006; Massara 2019). Questi studi condividono, inoltre, un disegno di ricerca comune e articolato in una prima fase geografica e in un successivo approfondimento di tipo narrativo. La fase geografica, che interessa maggiormente per questa riflessione metodologica, è schematizzabile come segue:

---

<sup>7</sup> Attraverso la costruzione di un indicatore dedicato per rilevare la capacità di offerta delle biblioteche italiane, l'ISTAT ha realizzato una mappatura *data driven* della loro polifunzionalità. Le 6 attività che denotano la polifunzionalità di queste biblioteche sono: 1) la promozione della lettura, laboratori del libro, gruppi lettura con esclusione delle attività dedicate ai bambini; 2) assistenza o supporto ai cittadini nello scrivere curriculum, compilare moduli, fare compiti ecc.; 3) animazione e/o laboratori per bambini (0-13 anni); 4) alfabetizzazione informativa (information literacy); 5) facilitazione digitale (digital literacy); 6) altre attività didattiche e/o di formazione.



- Analisi territoriale a livello macro della distribuzione nazionale delle biblioteche interessate dal fenomeno indagato utile per il campionamento ragionato;
- Analisi territoriale a livello micro del contesto urbano in cui le biblioteche del campione individuato operano attraverso la costruzione di profili di comunità;
- Interpretazione degli scenari emergenti dall'analisi complessiva dei profili di comunità costruiti con lo scopo di comprendere le condizioni contestuali in cui il fenomeno indagato si manifesta.

Mettendo a confronto questa articolazione con le sperimentazioni descritte in precedenza sull'uso della tecnologia GIS come strumento di indagine, l'elemento innovativo riguarda l'adozione di una prospettiva geografica *gis-based* al fine di osservare come i fenomeni indagati si manifestavano a livello macro e in quali condizioni le biblioteche studiate operavano a livello micro. In queste due indagini la finalità applicativa della tecnologia GIS non è stata quindi di tipo gestionale con lo scopo di ottimizzare uno o più servizi. Al contrario, l'intento perseguito è stato di natura interpretativa del significato delle chiusure e della polifunzionalità così da supportare la successiva fase narrativa all'interno di un più ampio progetto di ricerca interdisciplinare.

Nel dettaglio, la costruzione dei profili di comunità ha rappresentato un momento del processo di indagine particolarmente significativo per la riflessione che si sta conducendo perché ha modificato la percezione dell'uso della tecnologia GIS per la ricerca applicata alle biblioteche.

Le analisi di comunità rappresentano una strategia di indagine ampiamente discussa e sperimentata in ambito biblioteconomico e si possono definire come:

attività che consiste nella raccolta sistematica, nell'organizzazione e nell'analisi di informazioni relative alla comunità di riferimento della biblioteca che si attua principalmente attraverso l'utilizzo di fonti secondarie (statistiche, anagrafiche ecc.) ma anche attraverso fonti primarie (ricerca diretta) per restituire un quadro chiaro della comunità in cui la biblioteca si trova ad operare sotto il profilo socio-demografico, economico, produttivo ecc. (Faggiolani 2012, 181).

Il loro impiego viene comunemente previsto per desumere le esigenze dell'utenza potenziale di una biblioteca, a partire dalle caratteristiche del contesto in cui questa opera, confrontandole con i servizi offerti e al fine di trarne indicazione per valutare e conseguentemente migliorare i servizi stessi (Galluzzi 2001, 176).

Per le ricerche sulle chiusure e sulla polifunzionalità gli obiettivi conoscitivi che hanno guidato le due fasi geografiche, e più nel dettaglio la costruzione dei profili di comunità, sono stati tuttavia diversi. In entrambi gli studi, infatti, i profili di comunità non sono stati incentrati sulle caratteristiche del contesto d'azione della biblioteca che influenzano l'utenza, bensì sui tratti che contraddistinguono le relazioni sistemiche che all'interno del contesto influenzano invece la stessa biblioteca. In questo senso, l'oggetto studiato è cambiato e sarebbe pertanto più corretto intendere i profili realizzati come 'profili urbani'.

Proprio su tale cambio di prospettiva ha agito il metodo geografico supportato dalla tecnologia GIS. Nelle due indagini, infatti, questo tipo di studio è avvenuto attraverso un'attenta attività di osservazione a distanza raccogliendo numerose informazioni sul contesto urbano di un campione ragionato di biblioteche, con lo scopo di descrivere il sistema culturale di cui queste erano parte e in cui il fenomeno delle chiusure e della polifunzionalità si manifestava. In particolare, per ogni

Comune è stata mappata la presenza di luoghi o spazi dedicati all'offerta di opportunità di partecipazione culturale alle proprie comunità di riferimento (altre biblioteche, teatri, archivi, librerie, associazioni del terzo settore, circoli ecc.). Le cartografie che seguono sono un esempio dell'esito di questa operazione.

Alla fine di questa lunga attività di raccolta, georeferenziazione, descrizione e analisi delle informazioni l'interpretazione finale non ha riguardato il singolo contesto urbano, ma gli scenari emergenti che rispecchiavano quali condizioni sistemiche avevano potuto da una parte contribuire alla chiusura e dall'altra favorire lo sviluppo polifunzionale della biblioteca. In questo senso, la tecnologia GIS ha consentito di ricostruire in parallelo un elevato numero di profili urbani per interpretare le condizioni scatenanti un fenomeno che accomuna biblioteche presenti in contesti urbani e territoriali anche molto diversi tra loro. Ad esempio, nel caso dell'indagine sulle chiusure l'emergere di uno scenario di isolamento ha mostrato l'assenza di un sistema culturale che potesse supportare l'attività della biblioteca (Figura 1); lo scenario definito di sinergica interazione, emerso invece nel corso dello studio sulla polifunzionalità, ha rivelato come intorno ad alcune biblioteche polifunzionali esista un ecosistema di luoghi della cultura che amplificano la rete del sistema culturale dei contesti urbani in cui queste operano (Figura 2).

A partire dall'esperienza acquisita grazie a queste indagini la transizione metodologica della tecnologia GIS da strumento a metodo nelle ricerche applicate alle biblioteche sembra plausibile e offre l'opportunità di interpretare la grande quantità di dati di cui il settore dispone; di ricostruire

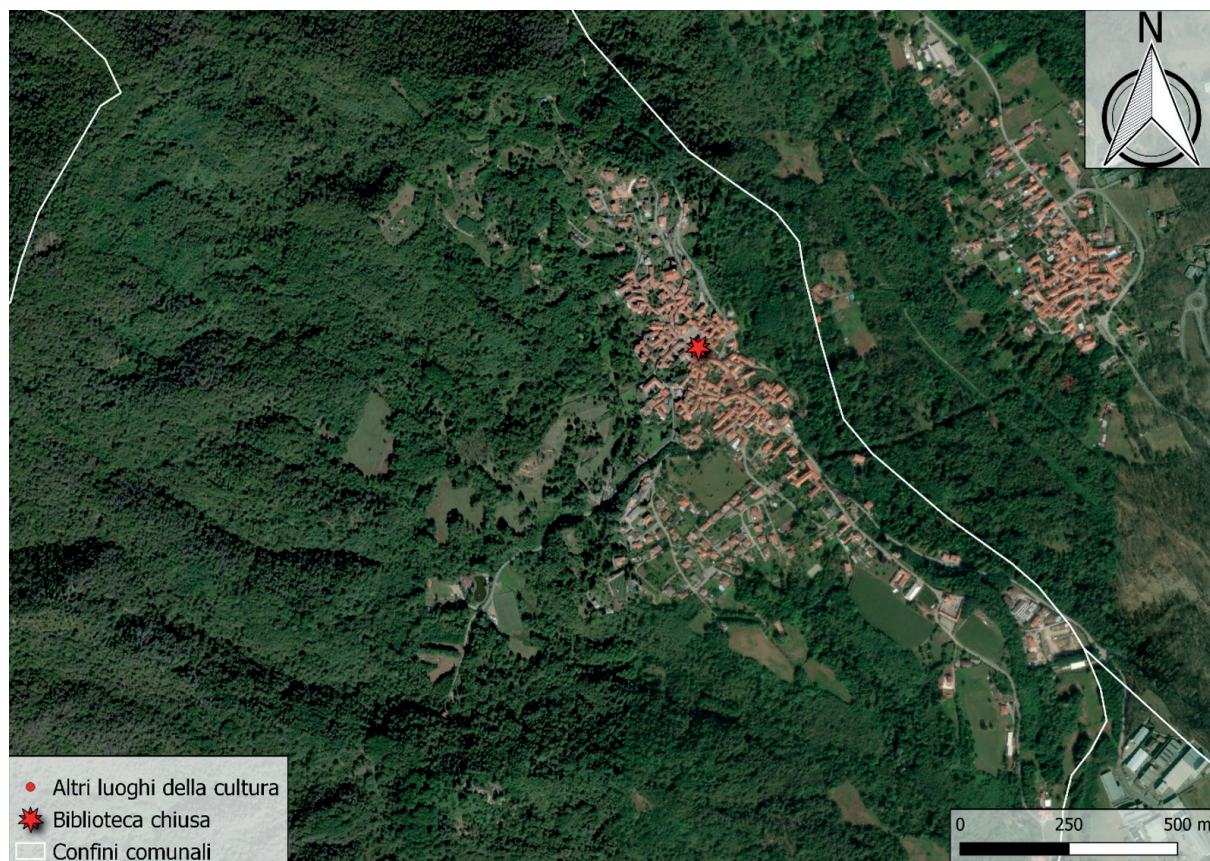


Figura 1. Esempio di scenario di isolamento.





Figura 2. Esempio di scenario di sinergica interazione.

i profili urbani delle biblioteche in una dimensione macro-sistemica; di fare emergere le relazioni intersettoriali che legano questa infrastruttura al contesto in cui opera. Per questa ragione il metodo geografico *gis-based* appare strettamente connesso ai presupposti che caratterizzano oggi la ricerca applicata alle biblioteche.

### **Verso una svolta interdisciplinare della biblioteconomia?**

Nel corso dell'argomentazione sono stati introdotti alcuni presupposti per la ricerca applicata alle biblioteche oggi che la potrebbero condizionare in modo decisivo nei prossimi anni e che rappresentano anche il risultato di un'evoluzione profonda della riflessione biblioteconomica, sia teorica che metodologica, sulle pratiche di valutazione.

Il percorso che ha portato alla consapevolezza del valore sociale delle biblioteche si inserisce, infatti, nel cambio di paradigma della biblioteconomia determinato dall'esigenza di conoscere il cambiamento generato nella vita delle persone dal contatto con i servizi della biblioteca, quindi di valutare l'impatto sociale di questa infrastruttura culturale (Faggiolani e Solimine 2013). Tale esigenza ha richiesto uno sforzo interpretativo di evidenze nuove – non misurabili e qualitative – per il quale sono state necessarie metodologie e strumenti di ricerca innovativi. Andando a ritroso, anche il precedente passaggio alla biblioteconomia gestionale è stato determinato dall'emergere di nuovi interessi conoscitivi (l'efficacia, l'efficienza e la qualità dei servizi ad esempio) che sono

stati soddisfatti ricorrendo a saperi e culture scientifiche diverse dalla biblioteconomia. Questi due momenti hanno segnato il percorso evolutivo della disciplina ed entrambi hanno ampliato profondamente la prospettiva per valutare le questioni emergenti nel dibattito scientifico.

Richiamare questa evoluzione mostra l'identità interdisciplinare della biblioteconomia. In uno scenario interdisciplinare esistono, infatti, saperi 'connettori', come si è visto per la geografia, e saperi 'catalizzatori' e sembra essere proprio questo il caso della scienza biblioteconomica che ha dimostrato di possedere la capacità di accogliere prospettive, concetti, metodologie e strumenti di indagine da altri saperi scientifici.

Per immaginare una trasformazione delle biblioteche compatibile con il ruolo sociale riconosciuto alle infrastrutture culturali per il benessere individuale e collettivo, si dimostra pertanto necessario sperimentare metodi di ricerca inediti e alternativi senza alterare l'identità disciplinare della biblioteconomia, ma assecondando la sua naturale propensione all'interdisciplinarietà.

## Riferimenti bibliografici

AIB (Associazione italiana biblioteche. Gruppo di lavoro gestione e valutazione). 2000. *Linee guida per la valutazione delle biblioteche pubbliche italiane. Misure, indicatori, valori di riferimento*. Roma: Associazione italiana biblioteche.

Amaturo Enrica, e Gabriella Punziano. 2006. *I mixed methods nella ricerca sociale*. Roma: Carocci.

Arosio, Fabrizio Maria, e Alessandra Federici. 2022. “Il profilo delle biblioteche attraverso i dati. Piattaforme d’informazione e socialità.” In *Le biblioteche nel sistema del benessere. Uno sguardo nuovo*, a cura di Chiara Faggiolani, 51-74. Milano: Editrice Bibliografica.

Bertazzoli, Agnese. 2022. “Bilanci e nuove prospettive della ricerca applicata per le biblioteche: resoconto del convegno ‘Misurazione e valutazione delle biblioteche: ne è valsa la pena? Cambiamenti, impatti e racconti di 30 anni di lavoro comune.’” *AIB studi* 62 (3): 653-666. <https://doi.org/10.2426/aibstudi-10257>.

Bishop, Bradley W., e Lauren H. Mandel. 2010. “Utilizing geographic information systems (GIS) in library research.” *Emerald* 28 (4): 536-547. <https://doi.org/10.1108/07378831011096213>.

Dapiaggi, Daniele. 2014. “Il GIS in biblioteca a supporto del *decision making*.” *Biblioteche oggi* 32 (7): 21-28. <http://dx.doi.org/10.3302/0392-8586-201407-021-1>.

Dapiaggi, Daniele. 2017. *Geolocalizzare per decidere. Connettere la biblioteca al territorio*. Milano: Mimesis.

Faggiolani, Chiara. 2012. *La ricerca qualitativa per le biblioteche. Verso la biblioteconomia sociale*. Milano: Editrice Bibliografica.

Faggiolani, Chiara. 2017. “‘Ciò che misuriamo influenza ciò che facciamo’: una nuova stagione per la valutazione in biblioteca pensando al BES.” In *La biblioteca pubblica nelle politiche di welfare. Nuovi criteri di valutazione dell’impatto sociale*, 21-34, allegato a *La biblioteca aperta. Tecniche e strategie di condivisione*, Milano 16-17 marzo 2017, a cura di Editrice Bibliografica. Milano: Editrice Bibliografica.

Faggiolani, Chiara. 2019. *Conoscere gli utenti per comunicare la biblioteca. Il potere delle parole per misurare l’impatto*. Milano: Editrice Bibliografica.

Faggiolani, Chiara. 2021. “Un indicatore dedicato alle biblioteche nel Rapporto BES dell’ISTAT. Una grande conquista per il nostro settore.” *AIB studi* 61 (1): 7-10. <https://doi.org/10.2426/aibstudi-13248>.

Faggiolani, Chiara. 2022. *Le biblioteche nel sistema del benessere. Uno sguardo nuovo*. Milano: Editrice Bibliografica.

Faggiolani, Chiara. 2023. “Le cinque tesi per le biblioteche che verranno.” *Agenzia di stampa CULT*. <https://www.agenziacult.it/notiziario/le-cinque-tesi-per-le-biblioteche-che-verranno/>.

Faggiolani Chiara, Alessandra Federici, e Camilla Quaglieri. 2023. “Biblioteche, infrastrutture culturali e polifunzionalità. Una mappatura *data driven*.” *AIB studi* 63 (2): 245-262. <https://doi.org/10.2426/aibstudi-13883>.



- Faggiolani Chiara, e Giovanni Solimine. 2013. "Biblioteche moltiplicatrici di welfare. Verso la biblioteconomia sociale." *Biblioteche oggi* 31 (3): 15-19. <http://dx.doi.org/10.3302/0392-8586-201303-015-1>.
- Faggiolani, Chiara, e Camilla Quaglieri. 2023. "Infrastrutture urbane di prossimità. Uno studio sulla polifunzionalità degli spazi pubblici dedicati alla cultura." *Biblioteche Oggi Trends* 9 (2): 78-90. DOI: 10.3302/2421-3810-202302-078-1.
- Galluzzi, Anna. 2001. "Analisi di comunità. Uno strumento per la pianificazione dei servizi." *Bollettino AIB*, 41 (2): 175-186.
- Goodchild, Michael F. 2010. "Twenty years of progress. GIScience in 2010." *Journal of Spatial Information Science* 1: 3-20. DOI: 10.5311/JOSIS.2010.1.2.
- ISTAT (Istituto nazionale di statistica). 2021. *Rapporto BES 2020. Il benessere equo e sostenibile in Italia*. Roma: ISTAT. <https://www.istat.it/it/archivio/254761>.
- ISTAT (Istituto nazionale di statistica). 2021. *Statistiche culturali*. Roma: ISTAT. <https://www.istat.it/it/archivio/264586>.
- ISTAT (Istituto nazionale di statistica). 2022. *Rapporto BES 2021. Il benessere equo e sostenibile in Italia*. Roma: ISTAT. <https://www.istat.it/it/archivio/269316>.
- ISTAT (Istituto nazionale di statistica). 2022. *Statistiche culturali*. Roma: ISTAT. <https://www.istat.it/it/archivio/279105>.
- ISTAT (Istituto nazionale di statistica). 2023. *Rapporto BES 2022. Il benessere equo e sostenibile in Italia*. Roma: ISTAT. <https://www.istat.it/it/archivio/282920>.
- ISTAT (Istituto nazionale di statistica). 2023. *Statistiche culturali*. Roma: ISTAT. <https://www.istat.it/it/archivio/292298>.
- Mark, David M. 2000. "Geographic information science. Critical issues in an emerging crossdisciplinary research domain." *Journal of the Urban and Regional Information Systems Association* 12 (1): 45-54.
- Massara, Flavia. 2019. "Mixed methods. Come integrare ricerca qualitativa e quantitativa in biblioteca." In *Conoscere gli utenti per comunicare la biblioteca. Il potere delle parole per misurare l'impatto*, a cura di Chiara Faggiolani, 139-169. Milano: Editrice Bibliografica.
- OECD CERi (Organisation for Economic Cooperation and Development Centre for Educational Research and Innovation). 1972. *Interdisciplinarity. Problems of Teaching and Research in Universities*. Washington: OECD Publications Center.
- Paltrinieri, Roberta. 2022. *Il valore sociale della cultura*. Milano: Franco Angeli.
- Pickard, Alison J. 2010. *La ricerca in biblioteca. Come migliorare i servizi attraverso gli studi sull'utenza*. Milano: Editrice Bibliografica.
- Poletti, Angela. 2011. *GIS. Metodi e strumenti per un nuovo governo della città e del territorio*. Rimini: Maggioli editore.
- Quaglieri, Camilla. 2022. *L'impatto delle politiche culturali sulla vitalità delle biblioteche. Una prospettiva transdisciplinare*. Roma: Associazione Italiana Biblioteche.

Serravalli, Alessandro. 2001. *GIS. Teorie e applicazioni*. Imola: La Mandragola.

Spada, Alessandra. 2007. *Che cos'è una carta geografica*. Roma: Carocci.

Venuda, Fabio. 2005a. "Il GIS (Geographic Information System) in biblioteca. Prima parte". *Bollettino AIB* 45 (3): 327-346.

Venuda, Fabio. 2005b. "Il GIS (Geographic Information System) in biblioteca. Prima parte". *Bollettino AIB* 45 (4): 477-488.

WHO (World Health Organization). 2009. *Commission on social determinants of health*. [https://apps.who.int/gb/ebwha/pdf\\_files/A62/A62\\_9-en.pdf](https://apps.who.int/gb/ebwha/pdf_files/A62/A62_9-en.pdf).

## Teaching cataloguing after RDA 3R project: Lessons learned

Konstantinos Kyprianos<sup>(a)</sup>, Foteini Efthymiou<sup>(b)</sup>, Georgia Katsira<sup>(c)</sup>

a) University of West Attica, <https://orcid.org/0000-0003-2948-1269>

b) University of West Attica, <https://orcid.org/0000-0002-7970-0856>

c) University of West Attica, <https://orcid.org/0009-0000-7488-7837>

**Contact:** Konstantinos Kyprianos, [kyprian@uniwa.gr](mailto:kyprian@uniwa.gr); Foteini Efthymiou, [feuthim@uniwa.gr](mailto:feuthim@uniwa.gr);  
Georgia Katsira, [gkatsira@uniwa.gr](mailto:gkatsira@uniwa.gr)

**Received:** 12 January 2024; **Accepted:** 22 February 2024; **First Published:** 15 May 2024

### ABSTRACT

This paper describes how RDA 3R was incorporated into a two-semester cataloguing course syllabus for first-year students and investigates students' perceptions and feelings about the courses taught. The information gathered will guide the further development and improvement of the course to fully meet the needs of students and the requirements of the market, especially with the emergence of new resource description standards. Following the course teacher's approval, the researchers distributed questionnaires during class hours. This method ensured that all students participated. The questionnaire was completely anonymous. It included two demographic questions and the CEQ 23 instrument with a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Besides that, a self-assessment questionnaire was employed to better understand their confidence in the knowledge gained. Quantitative analysis was carried out on the data related to 63 individual students. The quantitative analysis employed the statistical package SPSS. The findings revealed that students are reasonably confident in their new skills but prefer more practice over theory. Finally, they mentioned that they would like more precise goals and an extension of the courses to three semesters, as the workload is quite heavy. Overall, students feel relatively confident about the knowledge gained and believe they can catalogue in a library setting. The CEQ 23 instrument can be used in various knowledge organisation courses. It can provide a better understanding, with the ultimate goal of changing the curriculum or repurposing information organisation courses, allowing any Library and Information Science (LIS) school to judge the future of studies and their direction.

### KEYWORDS

Cataloguing courses; RDA Toolkit; CEQ23; Teaching and Evaluation; Students' performance.

## Introduction

Cataloguing and classification are considered the two pillars of librarianship and they have always been at the very heart of library and information work (Sibiya & Shongwe, 2018). This could be attributed to the fact that a quality catalogue will help patrons search, retrieve, identify, locate, and effectively use library resources. Cataloguing has arguably become increasingly vital as technology has become omnipresent and libraries have evolved into technological hubs, yet job postings for cataloguers do not appear to be diminishing (Turner, 2020).

In this direction, several Library and Information Science (LIS) schools are offering cataloguing courses into their curricula and it is considered a core course in the discipline (Chen & Joyce, 2019). It is a complex and challenging course in which students must comprehend and use various tools and instructions. More specifically, future cataloguers must use a variety of cataloguing standards to create descriptive records (the well-known surrogates) for items in the library collection (e.g. Resource Description and Access – RDA, MACHine Readable Cataloguing – MARC, Library of Congress Subject Headings – LCSH, etc.). An additional challenge that adds to the complexity of the course is the fact that there is an extra difficulty in understanding, using and studying these tools and guides, especially for people whose first language is not English. For example, in Greece, there is no official translation of the basic tools for cataloguing.

Cataloguing is not a new course in library and information studies curricula (Snow & Hoffman, 2015). Some principles or rules for the description and categorisation of library items and collections appeared from the beginning of their development. This became even more pronounced with the emergence of the discipline in the late 19th century. Over the years, the cataloguing rules and the tools used have evolved to better meet the needs of describing items in library collections and thus facilitate users in searching and retrieving information.

Nowadays, with the advancement of technology and the creation of new forms of resources (e.g., electronic publications, video, websites, and diverse continuing resources, etc.) and new ways of describing knowledge (e.g., Resource Description and Framework – RDF, etc.), cataloguing standards and rules had to be updated and amended to satisfy the changing needs of users. In this context, the cataloguing guidelines and the RDA Toolkit were redesigned entirely in 2020 in both the software and the content (the so-called Resource Description and Access Restructure and Redesign (RDA 3R) project). The project resulted in the substitution of the previous edition of the RDA Toolkit (called from now on original RDA Toolkit) and the new RDA Toolkit took the name official RDA Toolkit. The new RDA guidelines (official RDA) are based on the most recent bibliographic conceptual model, the IFLA Library Reference Model (LRM) standard, and its structure differs from that of its predecessors, which included many numbered rules and explicit instructions. The instructions are now more abstract, giving the cataloguer more flexibility in representing the resource information. Hence, cataloguing guidelines have been restructured and redesigned in such a way that cataloguing courses must be updated to reflect these changes. A report on approaches to teaching RDA in the LIS classroom (Sze, 2022) identified this problem in introducing RDA after 3R and the new official RDA Toolkit, as well as what cataloguing teachers should keep in mind and what approaches educators who have already used the new RDA 3R Toolkit have followed.

Considering the aforementioned, the new RDA Toolkit was introduced in cataloguing courses at the University of West Attica's Department of Archival, Library, and Information Studies (ALIS)

for the academic year 2022-23. The course was offered to the department's undergraduate students over two semesters so that they could understand all of the cataloguing concepts and tools. Along these lines, this study seeks to elicit information about the course taught and students' thoughts and feelings about cataloguing after completing the two-semester course. The information gathered will guide the further development and improvement of the course to fully meet the needs of students and the requirements of the market, especially with the emergence of new resource description standards.

To achieve this, students were given a questionnaire at the end of the second semester to evaluate the courses in terms of difficulty, learning outcomes, comprehension of curricular concepts, and recommendations for enhancing the courses. The survey was based on the Course Experience Questionnaire (CEQ) (Ramsden, 1991) and it was expanded with additional questions (self-assessment questionnaire) mainly focusing on students' perceptions of the course and suggestions for improvement (Kyprianos et al., 2022).

## Related work

This section will focus on previous works about the evaluation of cataloguing courses in general with the original RDA, what constitutes an efficient cataloguing course, and what are the core competencies that a future cataloguer should have, since no similar surveys relating to the application of the new official RDA Toolkit into the curriculum have been conducted. When it comes to teaching cataloguing, it appears that several surveys have shown that students prefer more exercises and practical examples than theory (Kyprianos et al., 2022; Chen & Joyce, 2019; Engelson, 2019; Snow & Hoffman, 2015; Al Hijji & Fadlallah, 2013; Mugridge, 2008).

Moreover, instructor effectiveness is one of the key topics identified in cataloguing course surveys. Snow and Hoffman (2015) state that the instructor's effectiveness involves attitude, knowledge, enthusiasm, teaching skills, and engagement. Furthermore, Engelson (2019) stated that educators must find the right balance between theory and practice when teaching cataloguing, which is a key concern. Another consideration for instructors is that the cataloguing course needs analytical thinking and questioning learning abilities rather than simple memorisation of factual knowledge (Chen & Joyce, 2019). Having this in mind, clear explanations, approachability, and responses to student questions all contribute to favourable assessments of instructor effectiveness.

Additionally, the course content also impacts the positive evaluation of cataloguing courses. Students appear to like courses that cover a wide range of cataloguing topics, such as principles, standards (e.g., MARC, RDA), metadata, and classification systems. The applicability of the course content to real-world scenarios and practical applications is also highly valued in this regard (Engelson, 2019; Sibiyi & Shongwe, 2018; Snow & Hoffman, 2015). Hence, course organisation and structure also appear to impact the cataloguing course evaluation. In particular, students prefer clear instructions and access to supplemental resources that enhance their knowledge, well-organised course materials, and logical progression of topics taught (Chen and Joyce, 2019; Kyprianos et al., 2022).

Furthermore, the placement of cataloguing in the actual environment of a library appears to alter students' attitudes regarding cataloguing courses. According to Snow and Hoffman (2015),



putting cataloguing in the real world of a library means demonstrating how cataloguing benefits users, providing authentic items and examples for cataloguing practice, giving students access to cataloguing tools used in practice, analysing local practices, and gaining experience with an Integrated Library System (ILS). Chen and Joyce (2019) support this assumption by tying classroom activities to themes students see impacting their future roles as librarians.

To continue, various studies have been conducted in this regard over the years about cataloguing courses offered in the curriculum and job requirements for cataloguing and metadata experts (Al Hiji & Fadlallah, 2013; Sibiya & Shongwe, 2018; Chen & Joyce, 2019; Monyela, 2021; Sibiya & Chuma, 2021; Zhang, 2023). More specifically, it seems that traditional cataloguing standards are still in great demand when posting a librarian job and many schools around the world are still offering information organisation courses, particularly library cataloguing, and classification in their curriculum. Moreover, Turner (2020), who studied the cataloguing job market, concluded that cataloguing jobs are still in great demand.

Here, it should also be mentioned that through the years several core competencies that a future cataloguer should conquer have been proposed. According to the Cataloguing Competencies Task Force (2017), Snow et al. (2023) and Frederick (2018) students of Library and Information Schools should have specific core competencies to become cataloguing and metadata professional librarians. The three necessary competencies are a) knowledge competencies, b) skill and ability competencies, and c) behavioural competencies. The first competence is related to the ability of students to be aware of the foundational cataloguing and metadata principles, the systems and the technology related to these systems, and the trends in the cataloguing and metadata profession. Following that, the second competency is related to the ability of students to synthesize and use all the guidelines, standards, and systems to create bibliographic data. Finally, the last competency focuses on the ability of students and future librarians to achieve interpersonal communication, be able to solve problems and be user-centred and user-oriented.

## Course description

The two-semester cataloguing course is available to first-year students in the Department of Archival, Libraries and Information Studies at the University of West Attica. These courses are offered in the school's first and second semesters, and students can acquire all the skills and knowledge required to work as cataloguers in libraries. In this context, students should obtain the essential basic (theoretical) knowledge (i.e., applicable terminology, the theoretical framework for cataloguing, historical overview.), as well as the specialised knowledge (i.e., tools and standards) necessary to perform cataloguing. Therefore, when considering the course structure, it was decided that the first semester should cover the theoretical background students should gain. In contrast, the second semester will introduce students to the tools required for cataloguing and address them with more practical exercises and real-life scenarios. The decision was made having in mind initially Dobreski's (2019) webinar 'Teaching RDA after 3R', in which he pinpointed that some teaching materials, examples, and labs will need to change, especially in relation to the Toolkit's new interface, LRM, and the new Relationships. Secondly, Sze's (2022) report on approaches to teaching RDA in the LIS classroom gave the educators interesting insights for creating the structure of the courses. Finally,

the third factor that influenced the course curriculum modification was the results from previous research by Kyprianos et al. (2022) where an evaluation of the cataloguing course was performed and the results revealed that the students prefer more cataloguing practice (creation of bibliographic records), balancing theory and practice, and transferring cataloguing in the real world.

In this line of thought, the students are offered a mixed course of 5 hours (3 hours of theory and 2 hours of practice) per week for 13 weeks each semester, meaning that the students have 26 weeks of cataloguing. In general, the course covers the theories, principles, and methods of bibliographic description and the application of international standards to create library catalogues. Furthermore, it covers the fundamental concepts of descriptive cataloguing, such as the elements of bibliographic description, the description of all types of library resources, the selection of access points, the creation of authorised access points, the principles and practices of authority work, and the application of encoding standards. The structure of the two courses is outlined in Table 1.

<b>Introductory cataloguing course</b>	
Week 1	Introduction to the course
Week 2	Library catalogues and terminology related to cataloguing
Week 3	Historical overview of cataloguing, libraries, and tools related to cataloguing
Week 4	Introduction to Entity-Relationship Models
Week 5-6	Introduction to IFLA LRM
Week 7	Introduction to Linked Data
Week 8	Introduction to the functionality of RDA Toolkit
Week 9	General instructions and recording methods
Week 10	RDA implementation scenarios and RDA controlled vocabularies
Week 11	RDA Toolkit – Describing manifestations
Week 12	RDA Toolkit – Describing carriers
Week 13	Final examination
<b>Advanced cataloguing course</b>	
Week 1	Introduction to the course and revision of the topics covered in the previous semester
Week 2	RDA Toolkit – Identifying works and expressions
Week 3-5	RDA Toolkit – Identifying persons, families, and corporate bodies and recording relations
Week 6	Creating application profiles for bibliographic descriptions of any resource
Week 7	Introduction to Bibliographic MARC21
Week 8	Introduction to MARC21 Authorities
Week 9	Introduction to ILS Koha
Week 10	Introduction to the subject description
Week 11	RDA Toolkit – Diachronic works
Week 12	RDA Toolkit – Cases of other works (movies, music, maps, etc.)
Week 13	Final examination

Table 1. Course syllabus.

According to the curriculum, students must complete the first-semester course with at least a passing mark to attend the second-semester course. Such a decision was made since the course difficulty increases, and if students do not understand the fundamental concepts taught in the first semester, they would be unable to cope with the expectations and needs of the second-semester course.

More specifically, the objectives of the course can be summarised as follows:

- Recognise the significance of descriptive cataloguing in libraries.
- Learn the fundamentals of cataloguing terminology.
- Use descriptive cataloguing's core concepts, principles, and objectives.
- Learn the fundamentals of creating bibliographic metadata for information resources.
- Learn the fundamentals of creating authority metadata.
- Understand the connection between descriptive cataloguing and information discovery.
- Use standards and tools in descriptive cataloguing.
- Employ controlled vocabularies commonly associated with descriptive cataloguing.

Here, it should be noted that the course heavily relies on descriptive cataloguing with only one week out of 26 that refers to subject cataloguing because the latter is being taught thoroughly in the third semester of the curriculum. Upon completion of that course, students are able to use subject indexing systems (e.g., Library of Congress Subject Headings (LCSH), Dewey Decimal Classification (DDC), Universal Decimal Classification (UDC), Library of Congress Classification System (LCC), etc.), to understand the process of assigning subjects and organizing content by subject, and associate the subject with the classification number. Thus, for the needs of the cataloguing course, students are given pre-prepared subject headings and classification numbers to insert into the corresponding MARC fields. This way, students at the end of the course can create a full bibliographic record based on cataloguing standards and guidelines.

## Methodology

As mentioned earlier in this paper, this quantitative survey was based on the CEQ instrument since it is a useful tool to get insights about teaching quality and can help teachers understand the efficiency of the curriculum and how they can improve it (Byrne & Flood, 2003). Additionally, the instrument was expanded with the employment of a self-assessment questionnaire (Kyprianos et al., 2022) allowing the teachers to identify exactly the difficulties of the course taught and how they can address these problems. The CEQ is an instrument developed in Australia where it is extensively used as a performance indicator of teaching quality (Byrne & Flood, 2003). More specifically, it was designed in the context of a teaching and learning theory in which students' perceptions of curriculum, teaching, and assessment are recognized as influencing their learning approaches and the quality of their learning outcomes, implying that it is an excellent tool for evaluating the cataloguing course. The original CEQ contains 30 items, but the most widely used version is the short form, which includes 23 questions (Asonitou et al., 2018).

Having the above in mind, the questionnaire was distributed electronically to the students enrolled in the two-semester cataloguing course and it was optional (only students who wished to complete the questionnaire could do so). Since this research was considered a course evaluation,

it did not require the approval of the university's ethics committee. According to the results, all students who took the course completed the questionnaire (63 students). Data were collected at the end of the second semester (June 2023), when students had finished their courses and had a firm grasp on the content delivered. Following the teacher's approval, the researchers distributed questionnaires during class hours. This method ensured that all students had the opportunity to participate in the survey. The questionnaire was completely anonymous. It included two demographic questions and the CEQ 23 instrument with a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

More specifically, according to the CEQ 23 instrument, five scales were included: (1) the Good Teaching Scale (GTS), containing six items measuring teacher efforts to increase student interest and provide feedback to students for motivating and guiding them toward success, (2) the Clear Goals and Standards Scale (CGSS), with four items on the students' perceived degree of clarity concerning graduation requirements, (3) the Appropriate Workload Scale (AWS) including four items for assessing the perceptions of sustainability of the overall academic workload, (4) the Appropriate Assessment Scale (AAS) having three items capturing student perceptions of the assessment methods' adequacy, and (5) the Generic Skills Scale (GSS) containing six items measuring the level of development of student analytic, problem-solving and communication skills (Byrne & Flood, 2003).

After completing the CEQ 23 questionnaire, students were given a self-assessment questionnaire using an agreement/disagreement 5-point Likert scale to better understand the knowledge acquired by the courses and the level of confidence that students have in their cataloguing skills. The 19-question self-assessment questionnaire was based on prior work by Kyprianos et al. (2022) and modified accordingly to meet the needs of evaluating the new official RDA Toolkit. Finally, an open-ended question was included in the questionnaire, asking participants what suggestions they had for improving course instruction.

## Results

### Sample profile

The majority of the sample were females (71.4%), while the ratio for males was 1:4 (25.4%), and 2 respondents (3.2%) did not want to answer this question. Additionally, the participants were asked to indicate their year of study. More specifically, as shown in Figure 1, most of the participants are in their first year of studies (35%), while the fewest participants are in their fourth year of study (8%). This is justified by the fact that the course is addressed to freshmen students, but the course's difficulty forces them to repeat it for the next few years of their academic careers. As previously stated, the course's complexity (many tools and guidelines are used in conjunction to achieve the creation of bibliographic descriptions and data), the fact that this course is offered to freshmen who have no prior knowledge of terminology or experience with libraries, and the practices they follow, and the lack of Greek translations of the tools used for cataloguing are all factors that interfere with students' performance. Finally, according to the correlation analysis, it seems that the year of study and the gender of the respondents do not affect the participants' responses.

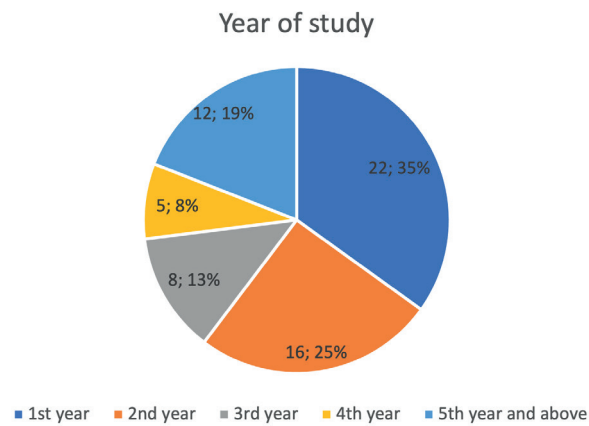


Figure 1. Year of study

### Reliability and validity of CEQ

The internal consistency of each instrument's scale was measured using Cronbach's coefficient alpha. The alpha can range from 0 to 1; the higher the value, the stronger the internal consistency. However, because reliability varies with the number of items on the scale, the alphas for different scale lengths are not directly comparable (Ainley, 2001). More specifically, the overall Cronbach's coefficient alpha is 0.729, while the separate values of each scale are GTS 0.818, CGSS 0.754, AAS 0.340, AWS 0.715, and GSS 0.796 (see Table 2). The low result of the AAS scale comes as no surprise since it contains a few items, and it is generally known that the number of items influences Cronbach's coefficient value (Cronbach, 1951). The other values are adequate for getting reliable results, according to Nunnally (1978).

	Cronbach's coefficient
<b>GTS</b>	0.818
<b>CGSS</b>	0.754
<b>AAS</b>	0.340
<b>AWS</b>	0.715
<b>GSS</b>	0.796
Overall	0.729

Table 2. Cronbach's coefficient

Additionally, exploratory factor analysis was applied to the CEQ following the approach in previous validation studies (Ramsden, 1991; Wilson et al., 1997; Byrne & Flood, 2003). Kaiser-Meyer-Olkin (KMO) and Barlett Sphericity tests were employed to determine data compatibility for factor analysis. More specifically, the sampling appropriateness KMO value was determined to be 0.713. According to Field (2009), this value is deemed sufficient when it is greater than 0.50 and



is considered “good” when it is between 0.70-0.80. Moreover, the Bartlett test of sphericity ( $\chi^2 = 722.706$ ,  $p < .001$ ) also showed the adequacy of the sample.

Principal Components Analysis and Oblique Rotated Component Matrix were chosen as the factor analysis to reveal the scale’s factor design (Field, 2009; Asonitou et al., 2018). According to the findings, six factors were identified explaining 67.3% of the overall variance. The first factor accounts for 30%, the second for 12%, the third for 9.5%, the fourth for 6.2%, the fifth for 5%, and the sixth for 4.6%. Table 3 shows the distribution of the items according to the factors and their factor loads. As shown in Table 3, all the items have factor loads above 0.40. According to Field (2009), values above 0.40 are accepted as ideal. Consequently, the items contributed significantly to the factors.

Items	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
<i>GTS</i>						
7					-0.837	
16				0.512		
15	0.614					
19	0.663					
3	0.611					
17	0.593					
<i>CGSS</i>						
1	0.578					
23	0.857					
13			0.458			
6	0.633					
<i>AWS</i>						
22			0.655			
20		-0.623				
4		-0.805				
14						0.473
<i>AAS</i>						
8				-0.422		
12			0.799			
18				-0.878		
<i>GSS</i>						
2						0.685
5						0.854
10						0.704

9				0.452		
11						0.432
21						0.653
<b>Eigenvalues</b>	<b>6.934</b>	<b>2.748</b>	<b>2.174</b>	<b>1.414</b>	<b>1.151</b>	<b>1.060</b>
<b>Variance explained</b>	<b>30.149</b>	<b>11.947</b>	<b>9.451</b>	<b>6.147</b>	<b>5.004</b>	<b>4.608</b>
<b>Total variance explained</b>	<b>67.307</b>					

Table 3. Factor structure of CEQ item scores

A closer look at the data shows that factor 1 has 7 items from two scales. Items 15, 19, 3, and 17 from the Good Teaching Scale (GTS) and items 1, 13, and 6 from the Clear Goals and Standards Scale (CGSS) relate this way to the teachers' motivation and explain things to the students with the students' performance. Moving on, factor 2 has two items from the Appropriate Workload Scale (AWS), items 20 and 4. These two reverse-coded items explain the negative value, indicating that the workload was not too heavy and there was no pressure on the students. Next, factor 3 contains three items (22, 12, 13) from different scales, connecting the expectations of the cataloguing course with the comprehension of the curriculum taught. Additionally, factor 4 has four items (16, 8, 18, 9) from three scales. According to factor 4, students have related the teachers' feedback with the assessment scale and the memorisation of the curriculum taught. Factor 5 has only one item (15) related to the time the teachers spent commenting on the exercises performed by the students. Finally, factor 6 contains 5 items (2, 5, 10, 11, 21) from the Generic Skills Scale (GSS) and one item (14) from the Appropriate Workload Scale (AWS), meaning that the time given during the semester was enough for the students to develop their general skills.

From the above analysis, it is understood that items of scales are not fitted correctly based on the instructions of the CEQ 23 instrument. However, Cronbach's alpha of the overall questionnaire met the criterion of 0.729, and only the Appropriate Assessment Scale (AAS) was lower (0.340). Consequently, our tool is adequate for drawing results.

## Descriptive data

After studying the validity of CEQ 23, we move on to the descriptive analysis of the data collected. More specifically, according to the first scale of CEQ 23 (i.e., GTS), students appear to be rather satisfied with the teaching staff and the effort they made to understand the difficulties that the students had with the course (Q15, mean 4.11) and the helpful feedback that was given to them regarding their development (Q16, mean 4.11). Additionally, similar results have been recorded in Q3 (mean 4.00) regarding student motivation. Despite their satisfaction with the feedback, the respondents need more comments and directions regarding their work (Q7, mean 3.27). Moreover, the students want their teachers to explain things more clearly (Q17, mean 3.65) and to make the lesson more interesting (Q19, mean 3.81) (see Table 4).

<i>GTS</i>	<b>N</b>	<b>Mean</b>	<b>Median</b>	<b>Mode</b>	<b>Std. Deviation</b>	<b>Percentage</b>
3. The staff of this course motivated me to do my best work	63	4.00	4.00	5	1.032	63.49%
7. The staff put much time into commenting on my work	63	3.27	3.00	3	0.954	51.90%
15. The staff made a real effort to understand the difficulties I might be having with my work	63	4.11	4.00	4	0.918	65.24%
16. The staff normally gave me helpful feedback on how I was going	63	4.11	4.00	4	.863	65.24%
17. The staff was extremely good at explaining things	63	3.65	4.00	3	1.050	57.94%
19. The staff worked hard to make their subjects interesting	63	3.81	4.00	4	0.981	60.48%

Table 4. Good Teaching Scale

Moving on to the next scale of CEQ 23 (i.e., CGSS), the participants responded cautiously to questions 1, 6, 13, and 23 (mean 3.02, 3.56, 3.03, and 3.67, respectively), indicating that the courses' goals and standards were unclear (see Table 5). This could be attributed to the fact that students had to learn many new terms and theories in a short period to be able to learn cataloguing in two semesters. This is even more challenging when cataloguing courses are taught during the first semesters of school when students have yet to grasp the fundamental notions and concepts of the school to which they have been admitted. Especially when cataloguing requires synthetic thinking while considering many different tools and guidelines (e.g., IFLA LRM, RDA Toolkit, Controlled Vocabularies, MARC21, and Koha).

<i>CGSS</i>	<b>N</b>	<b>Mean</b>	<b>Median</b>	<b>Mode</b>	<b>Std. Deviation</b>	<b>Percentage</b>
1. It was always easy to know the standard of work expected	63	3.02	3.00	4	1.143	47.94%
6. I usually had a clear idea of where I was going and what was expected of me in this course	63	3.56	4.00	4	1.089	56.51%
13. It was often hard to discover what was expected of me in that course	63	3.03	3.00	3	1.107	48.10%
23. The staff made it clear right from the start what they expected from students	63	3.67	4.00	4	1.136	58.25%

Table 5. Clear Goals and Standards Scale

The third scale of CEQ 23 is AWS (see Table 6). According to the participants' responses, they felt that they needed more time to comprehend the notions and topics covered in this course (Q14 mean 3.14, and Q22 mean 3.11). This is justified by the fact that students must absorb a large amount of knowledge quickly to perform cataloguing. On the contrary, the students stated that the workload and the pressure to perform well in this course was moderate (Q4 mean 3.05, and Q20 mean 2.92).

<i>AWS</i>	<b>N</b>	<b>Mean</b>	<b>Median</b>	<b>Mode</b>	<b>Std. Deviation</b>	<b>Percentage</b>
4. The workload was too heavy	63	3.05	3.00	3	1.054	48.41%
14. I was generally given enough time to understand things I had to learn	63	3.16	3.00	4	1.110	50.16%
20. There was a lot of pressure on me to do well in this course	63	2.92	3.00	3	1.126	46.35%
22. The sheer volume of work to be got through in this course meant it couldn't all be thoroughly comprehended	63	3.11	3.00	3	1.166	49.37%

Table 6. Appropriate Workload Scale

Next, the fourth scale of CEQ 23 is AAS (see Table 7). The responses reveal that the participants recognise that learning to catalogue requires critical thinking since they stated that to do well in this course, they do not need a good memory (Q8, mean 2.49), and they believe that the teachers were focusing more on what they had understood rather than on what they had memorised (Q12, mean 2.40).

<i>AAS</i>	<b>N</b>	<b>Mean</b>	<b>Median</b>	<b>Mode</b>	<b>Std. Deviation</b>	<b>Percentage</b>
8. To do well in this course, all you really needed was a good memory	63	2.49	2.00	2	1.076	39.52%
12. The staff seemed more interested in testing what I had memorised than what I had understood	63	2.40	2.00	3	1.171	38.10%
18. Too many staff asked me questions just about facts	63	3.29	3.00	3	0.923	52.22%

Table 7. Appropriate Assessment Scale

Finally, the fifth scale of CEQ 23 is GSS (see Table 8). More specifically, the students stated that the course improved their analytical skills (Q5, mean 3.70). This could be attributed to the fact that the students were confronted with cataloguing problems in real-life examples and had to think about how to deal with the problems to describe the resources. On the other hand, a

low score was received for the question related to the ability to work as a team (Q9, mean 2.87). This can be justified because the course exercises were done individually. Furthermore, the participants need to be more confident about their problem-solving skills (Q2, mean 3.17 and Q10, mean 3.19) or their ability to plan their work (Q21, mean 3.25). Finally, the participants do not feel that the course particularly enhanced their writing communication skills (Q11, mean 2.90).

GSS	N	Mean	Median	Mode	Std. Deviation	Percentage
2. The course developed my problem-solving skills	63	3.17	3.00	3	0.871	50.32%
5. The course sharpened my analytical skills	63	3.70	4.00	4	0.891	58.73%
9. The course helped me develop my ability to work as a team member	63	2.87	3.00	3	1.157	45.56%
10. As a result of my course, I feel confident about tackling unfamiliar problems	63	3.19	3.00	3	0.877	50.63%
11. The course improved my skills in written communication	63	2.90	3.00	3	1.088	46.03%
21. My course helped me to develop the ability to plan my own work	63	3.25	3.00	3	1.150	51.59%

Table 8. Generic Skills Scale

### Self-assessment questions

The following part of the questionnaire contains 19 questions that assess the knowledge gained by the students during the two-semester course (see Table 9). Most of the concepts taught in the course appear to have been grasped by the participants, based on their responses, which are mostly above average (mean 3.40-3.94). The sole question with the lowest mean score (3.27) was SA16, which assessed students' understanding of the relationship between IFLA LRM and RDA. This result is noteworthy because the training began with the IFLA LRM presentation so that students could comprehend the logic and structure behind the guidelines in the official RDA Toolkit. However, such a result contradicts participants' responses to question SA9, which refers to students' ability to use the RDA Toolkit (mean 3.94). Teaching the IFLA LRM helped students understand the RDA's structure even though they may not have comprehended the entire connection.



	N	Mean	Median	Mode	Std. Deviation	Percentage
SA1. I understand the notion of “information organisation”	63	3.76	4.00	4	0.962	59.68%
SA2. I understand the cataloguing guidelines and tools	63	3.71	4.00	4	0.941	58.89%
SA3. I understand the operation of the Electronic Catalogue (OPAC)	63	3.70	4.00	4	1.042	58.73%
SA4. I understand the types of resources	63	3.87	4.00	4	0.924	61.43%
SA5. I understand the methods of describing resources	63	3.70	4.00	4	0.978	58.73%
SA6. I can identify the authorised access points	63	3.75	4.00	4	0.933	59.52%
SA7. I understand the cataloguing rules	63	3.51	3.00	3	0.948	55.71%
SA8. I can apply the cataloguing rules to create a bibliographic record	63	3.56	4.00	4	0.929	56.51%
SA9. I understand how to use the RDA Toolkit	63	3.94	4.00	4	0.982	62.54%
SA10. I understand the entities in the RDA Toolkit	63	3.90	4.00	4	0.946	61.90%
SA11. I understand the relationships in the RDA Toolkit	63	3.60	4.00	4	1.025	57.14%
SA12. I understand the MARC21 standard so I can create a bibliographic record	63	3.73	4.00	4	0.937	59.21%
SA13. I understand the relationship between subject headings and classification number	63	3.40	3.00	3	1.009	53.97%
SA14. I understand the necessity of standardisation of Persons and Corporate Bodies	63	3.75	4.00	4	0.950	59.52%
SA15. I understand the ways of representing information in the RDA Toolkit (structured, unstructured information, etc.)	63	3.48	3.00	3	1.148	55.24%
SA16. I understand the relationship between the IFLA LRM and the RDA	63	3.27	3.00	4	1.167	51.90%
SA17. I understand the relationship between bibliographic MARC21 and authorities MARC21	63	3.48	4.00	4	1.075	55.24%
SA18. Using Koha helped me understand how to create bibliographic records using rules and guidelines	63	3.73	4.00	4	1.081	59.21%
SA19. I understand the MARC21 standard so I can create an authorised record	63	3.70	4.00	4	1.010	58.73%

Table 9. Self-assessment questions

## Open-ended question

At the end of the questionnaire, an open-ended question allowed students to share their ideas and recommendations for making the course more understandable and easier for them. The majority of participants said that they preferred more practice over theory, which is consistent with earlier studies on cataloguing courses (Engelson, 2019; Kyprianos et al., 2022; Snow & Hoffman, 2015; Snow et al., 2018; Chen & Joyce, 2019).

Furthermore, many participants stated that they would like more cataloguing courses because they had to cover a lot of information in a short time. Results that are in compliance with research performed by Snow et al. (2018).

Finally, some participants stated that they would like to perform cataloguing in Koha earlier in the second semester because that way, they could understand different concepts and instructions better and faster. Such a statement contradicts the course syllabus's rationale: teachers believe that theoretical background should be taught first so that students can later apply what they have learned.

## Discussion

This study examines the learning results of a two-semester cataloguing course at the University of West Attica's Department of Archival, Library, and Information Studies. After teaching the new RDA guidelines for the first time, the CEQ 23 provided valuable insights into students' learning outcomes and perceptions.

According to the findings, students believe their teachers motivate them to do their best work. Such a result complies with the statement of Chen and Joyce (2019), who believe that the motivation of students is essential to achieve better student learning outcomes (see Table 4, Q3, mean 4.00). Moreover, the students rated favourably that they received feedback on the laboratory exercises while applying the theory (see Table 4, Q16, mean 4.11). This result is also confirmed by the research performed by Veitch et al. (2013).

Moving on to instructor effectiveness, according to Snow and Hoffman (2015), the most crucial characteristic of instructor effectiveness is his/her ability to provide explicit and practical instruction. Students notably stated that they were generally satisfied with the course syllabus and what was expected of them in this course (see Table 5, Q6, mean 3.56). Such a result could be justified because students are in their first year of study and, in most cases, are unfamiliar with the terminology, tools, and tasks performed in a library.

In line with the results from similar surveys, which revealed that students need to have critical and reflective thinking to perform cataloguing (Kyprianos et al., 2022; Chen & Joyce, 2019), the present study confirmed the belief that the course of cataloguing does not require pure memorisation (see Table 7, Q8 mean 2.49, and Q9 mean 2.40). The students have to learn how to use the specific tools that are needed to perform cataloguing rather than memorise specific rules and guidelines. Students perceive such a process after the first lessons when they are asked to apply what they have learned in theory to real-life scenarios with varying degrees of difficulty and characteristics (e.g., works in various forms (books, journals, online resources.), works with multiple or no creators).

Additionally, comparing the self-assessment questions to the previous work of Kyprianos et al. (2022), the course has enhanced students' confidence in the knowledge gained. As seen from Table 10, the students in the present study seem to understand better information organisation in general and the tools needed to perform cataloguing. A significant improvement has been performed in question SA12, which refers to understanding the MARC21 standard for creating bibliographic records (i.e., previous study mean 2.94, present study mean 3.73).

	Previous study (mean)	Previous study (%)	Present study (mean)	Present study (%)
SA1. I understand the notion of "information organisation"	3.56	56.51%	3.76	59.68%
SA2. I understand the cataloguing guidelines and tools	3.44	54.60%	3.71	58.89%
SA3. I understand the function of the Electronic Catalogue (OPAC)	3.1	49.21%	3.7	58.73%
SA4. I understand the types of resources	3.29	52.22%	3.87	61.43%
SA5. I understand the methods of describing resources	3.29	52.22%	3.7	58.73%
SA6. I can identify the authorised access points	3.07	48.73%	3.75	59.52%
SA8. I can apply the cataloguing rules to create a bibliographic record	3.22	51.11%	3.56	56.51%
SA12. I understand the MARC21 standard so I can create a bibliographic record	2.94	46.67%	3.73	59.21%
SA14. I understand the necessity of standardisation of Persons and Corporate Bodies	3.1	49.21%	3.75	59.52%

Table 10. Comparison of self-assessment question to the previous study from Kyprianos et al. (2022)

Therefore, it seems that the modification and restructuring of the course curriculum benefited the students and helped them to understand many concepts that may have been challenging for them in previous years, meaning that the implementation of RDA after 3R has been rather successful. Such a result can also be justified by the fact that students performed better at the cataloguing course in 2023 as compared to the findings of the previous study by Kyprianos et al. (2022). According to Figure 2, more students passed the course in 2023 with a grade of 50 or above, whereas approximately 40 students failed the course in 2021.

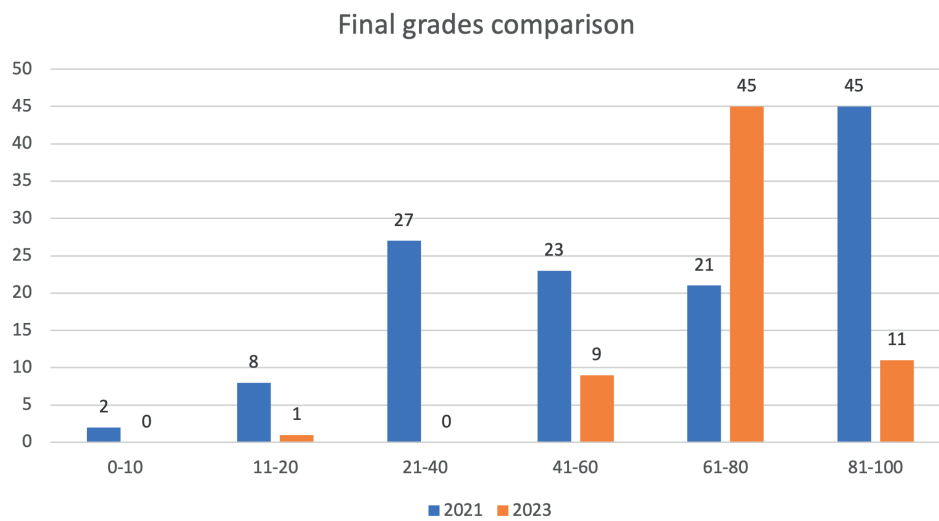


Figure 2. Final grades comparison

Finally, the students prefer practice over theory based on their responses to the open-ended question. This topic has sparked much debate (Chen & Joyce, 2019; Engelson, 2019; Snow & Hoffman, 2015; Moulaison, 2012; Intner, 2002). Furthermore, the students said they would like to take further cataloguing courses. In this context, teachers should consider adding more laboratory exercises and an extra cataloguing course to the department's curriculum in the future.

## Conclusions

This study used the CEQ 23 instrument, which measures student satisfaction with teaching outcomes in their course of study (Ramsden, 1991). CEQ 23 has been widely used by various institutions and disciplines worldwide (Asonitou et al., 2018) and can provide valuable information regarding the courses under consideration. The CEQ is designed to evaluate entire degree programs rather than specific courses, units of study, or teachers (Asonitou et al., 2018), but in our case, it can serve as a prototype for the evaluation and validation of similar cataloguing courses, or it can be expanded to evaluate the curriculum of related departments and schools throughout Greece or abroad. In addition, the presentation of the content of the cataloguing course and its successful implementation can serve as a model or basis for the creation of similar courses in other library and information science departments that wish to integrate a cataloguing course using the new RDA Toolkit into their curriculum. Given the recent introduction of the new RDA Toolkit, this research could serve as a springboard for other related research.

The survey's main findings are as follows: (1) clearer goals should be stated at the beginning of the courses, primarily because the courses are aimed at non-cataloguing freshman students with no prior knowledge of library science; (2) the creation of an extra cataloguing course to reduce the workload during the two semesters; (3) finding the right balance between theory and practice, as well as introducing more real-world scenarios; and (4) improving the collaboration between students.



The self-assessment questionnaire revealed that the students feel relatively confident about the knowledge gained and believe they can catalogue in a library setting. The findings are significantly better than previous research (Kyprianos et al., 2022), indicating that curriculum improvements have enhanced students' skills and competencies.

Future work should focus on developing new approaches to teaching cataloguing to first-year students and finding the correct balance of theory and practice, mainly when teaching RDA after the 3R project, where a robust theoretical background is required to perform cataloguing.

Finally, the CEQ 23 instrument can be used in various knowledge organisation courses such as metadata, digital libraries, and subject access systems. This will provide a better understanding, with the ultimate goal of changing the curriculum or repurposing information organisation courses in general. The CEQ 23 can also be applied to all courses in the school to provide a comprehensive curriculum assessment. This will allow the department to judge the future of studies and their direction.

## References

- Al Hijji, Khalfan Zahran, and Omar Sulaiman Fadlallah. 2013. 'Theory versus Practice in Cataloging Education in Oman: Students' Perspectives'. *Cataloging & Classification Quarterly*, 51 (8): 929-944. DOI: 10.1080/01639374.2013.832456.
- Asonitou, Sofia, Athanasios Mandilas, Evangelos Chytis, and Dimitra Latsou. 2018. 'A Greek Evaluation of the Course Experience Questionnaire: Students' Conceptions of the Teaching Quality of Higher Education Accounting Studies'. *International Journal of Business and Economic Sciences Applied Research*, 11 (2): 51-62. DOI: 10.25103/ijbesar.112.06
- Ainley, John. 2001. 'Course Experience Questionnaire 2000: an interim report'. Graduate Careers Council of Australia Ltd. <https://www.nagcas.org.au/documents/item/488>.
- Byrne, Marann, and Barbara Flood. 2003. 'Assessing the Teaching Quality of Accounting Programmes: An evaluation of the Course Experience Questionnaire'. *Assessment & Evaluation in Higher Education*, 28 (2): 135-145. DOI: 10.1080/02602930301668
- Cataloging Competencies Task Force. 2017. 'Core competencies for cataloging and metadata professional librarians'. ALA ALCTS CaMMS Competencies and Education for a Career in Cataloging Interest Group. <http://hdl.handle.net/11213/7853>
- Cronbach, Lee J. 1951. 'Coefficient alpha and the internal structure of tests'. *Psychometrika*, 16 (3): 297-334. DOI: 10.1007/bf02310555
- Chen, Suzhen, and Margaret Joyce. 2019. 'Teaching a cataloging/metadata course in a changing world: Experience and reflection'. *International Journal of Librarianship*, 4 (2): 111-122. DOI: 10.23974/ijol.2019.vol4.2.132
- Dobreski, Brian. 2019. 'Teaching RDA after 3R' [PowerPoint slides]. <https://www.slideshare.net/ALAELearningSolutions/teaching-rda-after-3r>
- Engelson, Leslie A. 2019. 'Sufficiency of Cataloging Education: School Librarians Respond'. *Journal of Education for Library and Information Science*, 60 (4): 285-311. DOI: 10.3138/jelis.2018-0072
- Field, Andy. 2009. 'Discovering Statistics Using SPSS'. London: SAGE.
- Frederick, Donna Ellen. 2018. 'Core competencies for cataloging and metadata professional librarians – the data deluge column'. *Library Hi Tech News*, 35 (8): 15-20. DOI: 10.1108/LHTN-08-2018-0047
- Intner, Sheila S. 2002. 'Persistent issues in cataloging education: Considering the past and looking toward the future'. *Cataloging & Classification Quarterly*, 34 (1/2): 15-28. DOI: 10.1300/J104v34n01\_02
- Kyprianos, Konstantinos, Foteini Efthymiou, and Dimitrios Kouis. 2022. 'Students' Perceptions on Cataloging Course'. *Libri*, 72 (2): 171-182. DOI: 10.1515/libri-2021-0054
- Monyela, Madireng. 2021. 'Cataloguing Education in the Era of 4IR: The Way Forward'. *Library Philosophy and Practice (e-journal)*, 5390. <https://digitalcommons.unl.edu/libphilprac/5390>
- Mugridge, Rebecca L. 2008. 'Experiences of newly-graduated cataloging librarians'. *Cataloging & Classification Quarterly*, 45(3): 61-79. DOI: 10.1300/J104v45n03\_06

- Moulaison, Heather Lea. 2012. 'A new cataloging curriculum in a time of innovation: Exploring a modular approach to online delivery'. *Cataloging & Classification Quarterly*, 50 (2/3): 94-109. DOI: 10.1080/01639374.2011.653096
- Nunnally, Jum C. 1978. 'Psychometric theory' (2nd ed.). New York: McGraw-Hill.
- Ramsden, Paul. 1991. 'A performance indicator of teaching quality in higher education: the course experience questionnaire'. *Studies in Higher Education*, 16 (2): 129-150. DOI: 10.1080/03075079112331382944
- Sibiya, Philangani Thembinkosi, and Mzwandile Muzi Shongwe. 2018. 'A comparison of the cataloging and classification curriculum and job requirements'. *Library Management*, 39 (6-7): 474-487. DOI: 10.1108/LM-09-2017-0089
- Sibiya, Philangani, and Kabelo Given Chuma. 2021. 'Are Cataloguers' Skills Still Relevant? A Critical Reflection on South African Cataloguing'. *Mousaion*, 39 (3). DOI: 10.25159/2663-659X/8669
- Snow, Karen, and Gertchen L. Hoffman. 2015. 'What Makes an Effective Cataloging Course? A Study of the Factors that Promote Learning'. *Library Resources & Technical Services*, 59 (4): 187-199. DOI: 10.5860/lrts.59n4.187
- Snow, Karen, Gertchen L. Hoffman, Maurine McCourry, and Heather Sandy Moulaison. 2018. 'Phoenix or Dodo? Re-envisioning Cataloging Education'. In: *Re-envisioning the MLS: Perspectives on the Future of Library and Information Science Education Advances in Librarianship*, Volume 44B, 227-239. DOI: 10.1108/S0065-28302018000044B013
- Snow, Karen, Bobby Bothmann, Staci Ross, Elizabeth Russey Roke, Pam Swaidner, and 2017 ALCTS Cataloging Competencies Task Force. 2023. 'Core Competencies for Cataloging and Metadata Professional Librarians'. ALA ALCTS CaMMS Competencies and Education for a Career in Cataloging Interest Group. <http://hdl.handle.net/11213/20799>
- Sze, Elisa. 2022. 'Report on approaches to teaching RDA in the LIS classroom'. <http://www.rda-rsc.org/sites/all/files/RSC-Papers-2022-1-ReportOnApproachesToTeachingRDAinTheLISclassroom.pdf>
- Turner, Rachel. 2020. 'Analyzing Cataloging Job Descriptions: Are Cataloging Jobs Disappearing, Changing, or Merging?'. *Cataloging & Classification Quarterly*, 58(6): 591-602. DOI: 10.1080/01639374.2020.1795768
- Veitch, Madeline, Jane Greenberg, Caroline Keizer, and Wanda Gunther. 2013. 'The UNC-Chapel Hill RDA boot camp: Preparing LIS students for emerging topics in cataloging and metadata'. *Cataloging & Classification Quarterly*, 51 (4): 343-364. DOI: 10.1080/01639374.2012.736124
- Wilson, Keithia L., Alf Lizzio, and Paul Ramsden. 1997. 'The development, validation and application of the course experience questionnaire'. *Studies in Higher Education*, 22 (1): 33-53. DOI: 10.1080/03075079712331381121
- Zhang, Lei. 2023. 'The Knowledge Organization Education Within and Beyond the Master of Library and Information Science'. *Knowledge Organization*, 50 (3): pp. 202-213. DOI: 10.5771/0943-7444-2023-3-202

## The Season of Living Archives: a Generative Provenance

Federico Valacchi<sup>(a)</sup>

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

**Contact:** Federico Valacchi, [federico.valacchi@unimc.it](mailto:federico.valacchi@unimc.it)

**Received:** 01 January 2024; **Accepted:** 01 February 2024; **First Published:** 15 May 2024

### ABSTRACT

The contribution, starting from consolidated archival concepts, introduces the theme of living archives by evaluating their potential political role and the impact they can have on cultural and social dynamics. It also evaluates what the tools for contextualized use of these resources could be, also in light of a cross-action of the principles of provenance and relevance.

### KEYWORDS

Archives; Provenance; Living Archives.

## La stagione degli archivi viventi: una provenienza generativa

### ABSTRACT

Il contributo, partendo da concetti archivistici consolidati, introduce il tema degli archivi viventi valutando il loro potenziale ruolo politico e l'impatto che possono avere sulle dinamiche culturali e sociali. Valuta inoltre quali potrebbero essere gli strumenti per un uso contestualizzato di queste risorse, anche alla luce di un'azione incrociata dei principi di provenienza e rilevanza.

### PAROLE CHIAVE

Archivi; Provenienza; Archivi viventi.

Gabriel Garcia Marquez nella sua autobiografia ha scritto che «la vida no es la que uno vivió, sino la que recuerda y cómo la recuerda para contarla» (Marquez 2002). Il grande colombiano, con la grazia della sua arte, risolve in una manciata di parole l'annosa questione della reale affidabilità dell'archivio in rapporto alle ragioni della produzione e, soprattutto, della conservazione.

L'archivio, infatti, è in ultima analisi un ricordo narrativo. Per quanto lo si costruisca con tecniche raffinate e lo si valuti con il dovuto rigore scientifico, resta intriso di inevitabile soggettività. L'archivio, e soprattutto l'archivio storico, più o meno volontariamente e più o meno consapevolmente, è una rappresentazione formalizzata. È frutto di una spontaneità sub iudice, molto spesso vigilata dagli interessi di chi lo genera.

La costruzione dell'archivio, sospesa tra il rigore metodologico della descrizione e la soggettività inevitabile dell'interpretazione, è del resto essa stessa una chiave di lettura/scrittura della memoria, una possibilità tra le molte di dare corpo alla realtà per poterla raccontare. Si aggiunga a questo che per sua natura l'archivio è polifunzionale e cambia più volte destinazione d'uso nel corso del ciclo vitale. Di conseguenza, un'innata trasversalità informativa ne determina vari stadi di esistenza e ne rimodula ogni volta i contenuti, senza reale controllo da parte dei mediatori esterni, destinati a subire l'archivio mentre tentano di governarlo.

Gli archivi non sono veri. Possono al massimo risultare autentici, ma non sono contenitori di inossidabili verità astratte, perché, appunto, sono il risultato di una sedimentazione controllata. La mediazione archivistica, allora, ha prima di tutto il compito di proteggere gli utenti da inganni documentari sempre in agguato, attingendo a tutte le sue tecniche e a tutti i suoi strumenti.

Il concetto di contesto, portato alle estreme conseguenze, in questo senso non è più semplice ricostruzione di alcune entità informative (produttore, conservatore etc). Il contesto si spinge verso la psicologia (potremmo dire l'antropologia?) della descrizione e incrocia le tortuose vicende della conservazione reale.

Gli archivi alla nascita sono immanenze documentarie giustificate solo e soltanto dall'uso che se ne può fare. Nel passaggio dalle ragioni della produzione a quelle della conservazione a tempo indeterminato un fondo archivistico matura però qualità che in origine erano solo abbozzate e diventa un tassello di più ampi mosaici di memoria individuale e collettiva. Se la storia della produzione è comprensione della realtà intesa come un presente dato in cui l'archivio si è formato, quella della conservazione è un percorso molto più immaginifico e faticoso. Dal fatto alla sua cristallizzazione storica succedono molte cose, non tutte costruttive, per la verità.

Il compito della funzione archivistica, almeno in teoria, è quello di tenere sotto controllo la vitalità inesauribile dell'archivio, riportandolo a schemi che, per quanto artificiali, ci aiutino a usarlo e a capirlo. Ci occorre un metodo, ma il metodo può anche tradirci. Gli archivi non sono specchi di chi li produce e cambiano volto nel tempo. Per tentare di governarli ci vogliono tecnica e fantasia, rigore e intuizioni, umiltà ed eterodossia.

Il metodo ci insegna che siamo figli della provenienza. L'archivistica speciale dura e pura, intesa come storia delle istituzioni, è il trionfo di un approccio condannato a una contestualizzazione determinata dalla ricerca spasmodica del profilo del soggetto produttore.

La fiducia nel rispecchiamento totale a suo tempo non ha però nulla di ingenuo. Negli scenari in cui matura è l'inevitabile punto di partenza, la naturale manifestazione di un metodo che nasce dal disagio informativo e dal faticoso rapporto tra quantità delle informazioni e qualità possibile della ricerca. Il rispecchiamento è una soluzione emergenziale, una scorciatoia cognitiva che cerca



di coniugare assunti etici e teorici con la realtà e con le ragioni più profonde della conservazione. Gli archivi sono figli di presenti e le circonvoluzioni dell'archivistica speciale pura ci restituiscono una collezione di presenti possibili, ma non ci fanno vedere fino in fondo le ragioni e le conseguenze di processi conservativi che agiscono sulla sedimentazione finale quanto e più degli assetti dei produttori.

Il processo di mediazione, anche a prescindere da queste considerazioni, si incardina comunque nell'ordine possibile che si riesce a conferire all'archivio. Buona parte dell'archivistica italiana, anzi, nasce proprio intorno all'ordinamento e all'illusionismo di un vincolo destinato esso stesso a modificarsi nel tempo e a perdere il proprio potere logico e connettivo.

L'ordine è il mito fondante di un'archivistica a posteriori, persuasa di un disordine inevitabile e rassegnata al fatto che il tempo, nella maggioranza dei casi, scombini le tassonomie funzionali introdotte ad hoc lungo il ciclo vitale. La fiducia nelle origini e il mito del soggetto produttore, per quanto temperati dalle vicende conservative, restano le linee guida di un procedimento tecnico destinato a ricostruire (e talvolta a forzare) il passato a tavolino. Nelle oscillazioni tra tecnica e rappresentazione l'ordine archivistico diventa un fatto politico e culturale, specchio, questa volta sì, della sensibilità della società che in un momento dato riordina con criteri che derivano dagli orientamenti e dalle esigenze di quella stessa società.

Se l'ordine è espressione di una visione del mondo prima che dell'archivio, le origini e gli sviluppi dei metodi di ordinamento e costruzione dell'archivio, con le conseguenti ricadute d'uso, ci aprono prospettive interessanti sulle comunità che su quegli archivi agiscono e in quegli archivi si riconoscono.

Stando alla dimensione strettamente archivistica l'evoluzione del metodo ci parla innanzitutto di una prima fondamentale discrasia tra il principio di pertinenza e quello di provenienza, cioè tra un mondo che privilegia i dati e uno che si prende cura anche e soprattutto delle ragioni e delle conseguenze di quei dati. Come vedremo, malgrado molta letteratura abbia insistito sulla polarizzazione dei due principi, pertinenza e provenienza non costituiscono aprioristicamente un ossimoro e possono talvolta coabitare in maniera relativamente pacifica o, più semplicemente, necessaria. Bene o male, comunque, continuiamo a fare professione di fede nella provenienza. Da Cencetti a Pavone e da Pavone all'interoperabilità, però, molte cose sono cambiate. L'ordine, intanto, non è più soltanto un monumento postumo. Negli archivi informatici, punto estremo di un range peraltro molto ampio, l'idea e la possibilità di ordine come sappiamo bene si manifestano anzi ancora prima dell'archivio stesso.

In linea generale, il soggetto produttore non è più l'affidabile galantuomo che abbiamo conosciuto e la sua credibilità di esegeta dell'archivio è messa in discussione dagli stessi meccanismi della produzione. Dal rispecchiamento all'interoperabilità l'archivio sembra più che mai solo lo specchio di sé stesso. Ci si mostra al massimo come una rifrazione indotta dalle peculiarità di processi di produzione e uso distribuiti nello spazio e nel tempo. Lo sbriciolamento del soggetto produttore dentro all'interoperabilità mette infatti in discussione l'univocità del concetto stesso di soggetto produttore.

Chi sono oggi i *creators*? La frammentazione della produzione e della conservazione nelle logiche di una crescente interoperabilità, istituzionale prima e tecnologica poi, ci fa parlare di una concorrenza delocalizzata di produttori, che modifica le gerarchie dell'azione documentaria e della conseguente sedimentazione.

La nostra società produce archivi polverizzati, genera informazione disseminata e, alla fine, è una società paradossalmente disordinata, le cui testimonianze documentarie si comprendono male alla luce del solo principio di provenienza e di un'idea troppo rigida di archivio.

Sembra utile ragionare sugli archivi partendo da nuovi elementi di valutazione, decidendo se nella concretezza del fare la linearità della trama giuridica e istituzionale, con la conseguente lettura filologica dell'archivio, possa ancora in qualche modo orientare i processi di sedimentazione prima e di interpretazione e uso poi.

Conviene infatti fare i conti con nuove evidenze documentarie, eccentriche rispetto alle categorie che abbiamo utilizzato fin qui per concettualizzare l'archivio. La realtà della sedimentazione e della conservazione non ci mette di fronte soltanto ad archivi *propri*, per quanto con le peculiarità del supporto binario. Reclamano la nostra attenzione anche aggregazioni documentarie che si muovono sui confini incerti di costruzioni sempre più distanti dall'ideale di provenienza. In genere questi "archivi" sono il risultato della facilità con cui si possono realizzare le costruzioni documentarie digitali e beneficiano della potenza di trasmissione delle reti, cui sono fisiologicamente destinati.

Sono archivi artificiali, risultato di un'aggregazione tematica (la pertinenza...) piuttosto che di una produzione necessaria in senso stretto (Valacchi 2023). Il produttore lascia il posto all'aggregatore e apre nuovi orizzonti informativi, oltre l'idea delle sedimentazioni apparentemente "naturali e spontanee" che abbiamo sempre chiamato archivi.

La stessa parola totemica, da cui tutto deriva, non basta più a darci il polso degli accadimenti documentari (Pezzica 2020). Servono sempre nuovi aggettivi per connotarla, nel tentativo di cogliere tutte le sfumature di un panorama documentario in perenne ridefinizione. Abbiamo ormai metabolizzato il concetto di archivi inventati o partecipativi, e sempre più di frequente siamo chiamati a confrontarci con raccolte digitali di documenti selezionati a partire dall'integrità di fondi analogici. Nel frattempo, dati e informazioni rilevanti circolano nei canali insubordinati e sfuggenti dei social.

Poco conta che questi complessi sprizzino eterodossia da tutti i pori, perché a poco serve in archivistica un'ortodossia astratta dalla realtà. Molte di queste aggregazioni sono comunque *archivi*, semplicemente perché espressione di un bisogno informativo che la società contemporanea manifesta e consolida in questi termini. Il solo fatto che esistano è sufficiente a consigliarci di considerarli, soprattutto per le conseguenze che in alcuni casi essi possono avere sulla società nel suo insieme.

Conviene riflettere in profondità, andando oltre il dogma dell'irrevocabile univocità sancito dalla provenienza, sul rapporto che c'è tra la granularità della sedimentazione e un malessere che sembra attraversare la nostra contemporaneità. In questo senso la moltiplicazione delle fenomenologie documentarie non è più una semplice espansione del concetto di archivio, o una sua incontrollabile deriva, ma rappresenta un indicatore di sofferenza informativa.

L'idea che determinate comunità vogliano rappresentarsi fuori dagli schemi del "mainstream documentario" – e di una storia raccontata a tavolino dai presunti vincitori di un lungo conflitto sociale ed economico – ci riporta alle radici politiche dell'arma documentaria. Certe aggregazioni possono alimentare e sostenere una vera e propria controcultura, un antagonismo (anche) archivistico capace di testimoniare l'esistenza e la persistenza di idee e persone altre dalla narrazione che il pensiero unico occidentale continua a proporre ad ogni livello.

Oltre i confini dell'archivio partecipativo, si apre così il mondo dei *living archives*, che la letteratura anglosassone colloca nella dimensione *post custodial* (Cook 1992), individuando in quel *post* una discontinuità rispetto a una lunga tradizione conservativa e interpretativa (Almeida e Hoyer 2019). Il modello *post custodial*, sia detto per inciso, non stupisce più di tanto la tradizione archivistica italiana. In definitiva, infatti, può essere ricondotto al nostro policentrismo, inteso come delega della conservazione agli stessi soggetti produttori. Questa delocalizzazione della custodia è però in certi frangenti soprattutto il segnale di un ribaltamento dei meccanismi di produzione, uso e sedimentazione e quindi del concetto di provenienza. Sembrerebbe insomma opportuno parlare di una *post provenienza*, più che di una *post custodia*. La *post custodia* nel nostro caso è infatti solo l'inevitabile riflesso meccanico della generazione di un archivio *not politically correct*, in piena sintonia con gli obiettivi di soggetti produttori moltiplicati, orientati a costruire evidenze documentarie alternative alle memorie istituzionalizzate.

Il fatto nuovo, metodologicamente rilevante e, per certi versi, davvero rivoluzionario, sta quindi nel superamento di un consolidato *ius archivi* e nel ridimensionamento di una provenienza monocratica, non più garantita da un impianto giuridico e istituzionale a senso unico. Il problema sembra risiedere nella fisionomia della produzione, oltretutto negli assetti della conservazione. La moltiplicazione della provenienza è il segnale di un ribaltamento di prospettive che colloca l'archivio (o ciò che ne resta) al centro di dinamiche sociali di portata ben più ampia di quella della loro ricaduta documentaria secca.

Il concetto e le pratiche del *living archive* rispondono a una serie di esigenze di rappresentazione dinamica della memoria che vanno dall'arte contemporanea<sup>1</sup> al design e a più generali e diffuse pratiche partecipative (Gaetarelli 2021), spingendosi perfino ad ipotesi di "archiviazione biologica".<sup>2</sup>

I *living archives* sono aggregazioni frutto di una relativa spontaneità dal basso che testimonia la vivacità informativa di ambienti anche molto diversi tra loro. Nel momento in cui ribaltano la piramide e i flussi della produzione, creano i presupposti per una riflessione ad ampio raggio sul ruolo pubblico e "comunitario" dell'archivio.

Il *living archive*, con molta e audace libertà, si potrebbe disambiguare in italiano come "archivio dal vivo", visto che nella nostra percezione l'archivio è comunque in ogni caso un'entità vitale. È, cioè, un archivio che viene formandosi non solo e non tanto come sedimentazione di norme e funzioni che generano fatti giuridicamente rilevanti. Deriva piuttosto da un'esigenza vitale di autodocumentazione che emerge spesso da contesti sociali alternativi al "potere costituito" o da ambienti nei quali la memoria è un'ipotesi in costante divenire, come nel caso degli artisti contemporanei, solo per fare un esempio.

La funzione archivistica in tutto ciò si arricchisce di senso politico e alimenta un attivismo che non è semplice collazione informativa, ma diventa motivata partecipazione alla costruzione di identità plurali e tendenzialmente "rivoluzionarie".

In questa accezione gli archivi sono intesi «as a site that is inclusive, is never complete and in which the archivist is an "active participant" in constructing the history that is archive» (Howard, Jarvie, e Wright 2021, 211).

<sup>1</sup> Curating Living Archives: <https://curatinglivingarchives.network/>.

<sup>2</sup> Istituto Europeo di Design (IED): <https://www.ied.edu/projects/biobits-exploring-living-archives/>.

Questo archivio in divenire è segnale di una dimensione attiva e partecipativa del fare documentario e del bisogno diffuso di condivisione dell'informazione. Investe in pieno la deontologia professionale, ridefinendone confini e obiettivi, ed è sorretto da una marcata dimensione collaborativa. Tutto ciò conferisce orientamento archivistico a una documentazione apparentemente disorganica che diviene così un potenziale strumento di costruzione del dissenso. In questa lettura si porta a nudo la funzione radicale, quasi ancestrale, dell'oggetto informativo. L'archivio è il luogo dove istanze sociali altrove trascurate incontrano una produzione documentaria che dà loro voce e ne garantisce la diffusione (Rhodes 2014).

Progetti come *Interference Archive* muovono proprio da qui: «the mission of Interference Archive is to explore the relationship between cultural production and social movements»<sup>3</sup>.

Interessante, tra i molti, anche il caso dei *Community-Driven Archives Overview* della Wilson Library: «the project supports historically underrepresented history keepers in telling, sharing, and preserving their stories»<sup>4</sup>.

Un attento e tempestivo interesse nei confronti del “movimentismo archivistico” si era peraltro manifestato in Italia in tempi non sospetti per merito di Marco Grispigni e Leonardo Musci. La loro *Guida alle fonti per la storia dei movimenti in Italia*, è uno strumento dalla condivisibile impostazione descrittiva e catalografica, ma non disdegna ampie aperture alla dimensione storiografica. La Guida entra tra l'altro nel merito delle problematiche degli archivi dei movimenti in una precisa stagione storica, assumendo che il movimento «ha nella sua forma tipica di quegli anni (...) una forte caratterizzazione antistituzionale e generazionale» (Grispigni e Musci 2003).

Sempre nel quadro italiano da segnalare anche il caso dell'Associazione per un archivio dei movimenti (AIMO) che «raccolge fondi documentari e materiale grafico prodotti dai movimenti politici e sociali a Genova e in Liguria, dagli anni '60 in poi»<sup>5</sup>.

In casi simili siamo sulla soglia di una percezione dell'archivio inteso come strumento di partecipazione attiva al dibattito politico e sociale, all'archivio che si fa possibile strumento di lotta informativa. Nel momento in cui si accetta questa dimensione relativamente nuova della dimensione documentaria, si superano i limiti fisiologici della narrazione a senso unico che segna in profondità l'efficacia dei sistemi archivistici tradizionali, almeno nella loro dimensione istituzionale. Il *living archive*, per sua natura, dà voce ad attori sociali senza i quali è problematico restituire davvero la complessità di una realtà ferita da divaricazioni sociali e culturali sempre più ampie.

Almeida e Hoyer sintetizzano il tema nel cosiddetto decalogo dei *living archives*:

The living archive cultivates human agency through collaboration and embodied action.

The living archive is generative.

The living archive is performative

The living archive must be a specific, culturally situated place.

The living archive emerges as an activist strategy that (re)constructs reality through a cross section of information, community organizing, and radical history.

<sup>3</sup> Interference Archive: <https://interferencearchive.org/who-we-are/about/>.

<sup>4</sup> University Libraries. Community-Driven Archives: <https://library.unc.edu/wilson/shc/community-driven-archives/about/>.

<sup>5</sup> Associazione per un archivio dei movimenti: <https://www.archiviomovimenti.org/>.

The living archive is both a method for interrogating the past and “an irreducible experience of the future.”

The living archive is itself a politics.

The living archive is not neutral.

The living archive enables political pluralism and strives for emancipation.

The living archive is an alternative formulation that presents a threat to those who seek to consolidate power and oppress others. In the living archive, everyone is an archivist.

The living archive relies on funding models that don't compromise its core values.

The living archive is social and takes up space in the world.

The living archive creates community and aspires to be a nexus between communities. In the living archive conceptions of permanence and preservation are determined by community ideas and intentions.

The living archive strives to preserve and understand rather than collapse difference.

The living archive changes how we think about relationships between the past, present, and future and makes space for non-human perspectives (Almeida e Hoyer 2019, 29).

Il manifesto evoca una serie di tematiche, magari non tutte e non del tutto originali, che rilanciano con forza il ruolo politico attivo dell'archivio. Di un archivio però che non è più soltanto espressione di una concezione gerarchica di sé stesso e del mondo che lo circonda, ma che rovescia la piramide per poi distribuirsi in una rappresentazione orizzontale multidimensionale.

Nella loro interpretazione politica aggregazioni di questo tipo sono sicuramente antidoti efficaci alla normalizzazione documentaria di archivi pubblici irrigiditi in una contemplazione narcisistica e in molti passaggi etnocentrica.

Non più “il luogo dove l'ordine è dato” ma piuttosto il luogo del disordine costituito, in cerca di nuove possibili letture delle relazioni sociali. L'archivio non è più semplice *mediazione* dei fatti ma diventa un luogo *immediato* dove fare politica, semplicemente accumulando testimonianze documentarie. La stessa descrizione archivistica finisce con l'essere coinvolta, perché “il conflitto è la dimensione permanente che lega momento produttivo e momento illustrativo e di studio di queste carte” (Musci 2021, 82).

I *living archives* veicolano una controinformazione diversa anche da quella edulcorata che si coglie ad esempio negli archivi partecipativi resi disponibili dal Roy Rosenzweig Center<sup>6</sup>. In quel caso, infatti, l'aggregatore rimane comunque un soggetto istituzionale ben connotato e l'ipotesi di democratizzare la storia deve fare i conti con un'idea molto circoscritta e non troppo inclusiva di democrazia.

I nostri archivi dal vivo, almeno in alcune loro manifestazioni, sono invece il collettore della voce nascosta dell'occidente e in questo senso hanno uno straordinario potere “eversivo”. Non servono solo ad accumulare testimonianze documentarie, ma possono anche determinare i presupposti per una profonda rivoluzione del pensiero e dell'azione politica. Li sostiene infatti l'idea tangibile di un attivismo basato sull'informazione e, appunto, sulla controinformazione. A ben guardare sono garanzia documentaria del diritto di esprimere opinioni diverse da quella che Leonardo Musci ha felicemente definito “la religione della memoria condivisa” (Musci 2021, 83).

---

<sup>6</sup> Si veda: <https://rrchnm.org/our-story/> e la dichiarazione, esplicita ma discutibile sotto molti punti di vista, “We use digital media to democratize history”.



Dal punto di vista strettamente archivistico in questo tipo di risorse la apparente discrasia tra provenienza e pertinenza è definitivamente superata. L'antitesi, infatti, non serve più a perimetrare per sottrazione un universo nel quale tende a scomparire l'autorevolezza del soggetto produttore. Ognuno è potenziale produttore e sono le comunità e le loro idee e determinare la fisionomia e la forza delle aggregazioni, dentro a un'utopia documentaria sotto molti punti di vista davvero affascinante.

Non è difficile intravedere in questi fenomeni grandi spazi di manovra anche per la disciplina degli archivi. Sulla scia di simili suggestioni l'archivistica può aiutarci a sognare un mondo nuovo. E lo può fare nel momento esatto in cui sceglie di dedicare le sue attenzioni non solo allo studio delle carte in quanto tali, ma anche all'impatto reale e propositivo di documenti multiformi sui possibili processi di trasformazione della società. Un'archivistica capace di andare oltre alla nebulosità di un'idea vaga di certificazione giuridica e alla retorica della memoria e del passato ad ogni costo può contribuire effettivamente a una rivoluzione culturale di cui si sente una forte esigenza.

Questo non significa naturalmente rinnegare consolidati valori pregressi e rifiutare una continuità disciplinare senza la quale ogni trasformazione ed ogni suggestione diventano letali salti nel vuoto. Dal punto di vista del metodo, e nell'interesse di un possibile ampliamento del ruolo disciplinare, occorre quindi analizzare attentamente le trasformazioni degli assetti documentari, misurandole alla luce dell'unico ordine di grandezza che conosciamo, quel metodo da cui siamo partiti, con i suoi principi, le sue potenzialità e i suoi limiti.

Cosa cambia rispetto a quei principi? Cosa significa parlare di provenienza in contesti tanto sfuggenti rispetto alla auspicata canonizzazione della produzione? E ancora, come è possibile cercare di contenere ed utilizzare la marea informativa che si sprigiona da nuove tendenze di produzione e sedimentazione?

La moltiplicazione e l'alterazione del principio di provenienza generano doppi, quando non tripli, documentari che risulta difficile governare e ricondurre a un qualsivoglia schema di modello conservativo. Il web è esso stesso un gigantesco e inestricabile modello conservativo, l'incarnazione di un policentrismo finale e privo di coordinate umane, geografiche e istituzionali.

Tutto ciò impone una riflessione di ordine epistemologico rispetto al ruolo che le discipline documentarie – e non solo l'archivistica – possono continuare a svolgere nella gestione di una conoscenza ormai pressoché ingovernabile con i consueti strumenti tassonomici.

In particolare, spostare il ragionamento dalle logiche di approssimazione all'informazione proprie del metodo storico al terreno molto più vasto della conoscenza impone una rinnovata attenzione ai contenuti, vale a dire alle informazioni. In linea di massima si deve tentare di invertire il senso di marcia di un'euristica che di necessità muove dal contesto in cerca di contenuti contestualizzati. In archivi dove il legame con il contesto (e cioè con il produttore) si sfuma se non si dissolve, il secolare approccio all'informazione basato sulla provenienza perde efficacia. Anche in questo caso, allora, può esserci utile tornare a riflettere sulla pertinenza e sulle materie, non più per sottolinearne le storture archivistiche ma per cercare di capire se in qualche modo ci possono orientare dentro ad archivi che per loro natura nascono già fortemente tematizzati. È un percorso di descrizione aumentata ancora da sviluppare compiutamente ma, al di là delle etichette e di certa retorica tecnologica, soggettazione, indicizzazione, text mining, AI possono essere gli strumenti che puntando al contenuto ci consentono di governare anche i contesti.

In questa sede, però, non ci interessa entrare nel merito delle soluzioni tecnologiche, che pure sono la sostanza applicativa del nostro ragionamento. Si vuole tentare piuttosto di segnalare in conclusione le conseguenze di un approccio basato sulla pertinenza o, per meglio dire, sulla provenienza integrata. Ciò che potremmo fare è in realtà incrociare due principi in apparenza antitetici, quali provenienza e pertinenza, senza rinunciare ai vantaggi che da ognuno possono derivare. Il concetto di provenienza rimane centrale. Serve a difendere, fosse solo in maniera banalmente comparativa, l'idea di un contesto necessario, qualunque sia la natura e la forma dell'archivio. La pertinenza, per parte sua, consente entrare nel merito dei contenuti che a questo punto sono elementi essenziali alla comprensione dell'archivio, e dei *living archives* e assimilati in particolare. È infatti a partire dai contenuti che in un'aggregazione artificiale si può risalire alla molteplicità dei soggetti produttori e si può tentare la sintesi di un contesto che si auto implementa senza sosta, infiltrandosi dal basso e risorgendo ogni volta dalle sue ceneri.

L'archivistica che fu speciale in questi scenari trova nuovo vigore e nuovi spazi ed è probabilmente chiamata ad un'ulteriore specializzazione, necessaria a dare risposte alle forti complicazioni documentarie che segnano il nuovo millennio.

Nella consapevolezza, antica ma nuovissima, che comprendere gli archivi significa conoscere la realtà e conoscere la realtà è il solo modo che abbiamo per sperare di cambiarla.

## Riferimenti bibliografici

- Almeida, Nora, e Jen Hoyer. 2019. "Living Archive in the Anthropocene." *Journal of Critical Library and Information Studies* 3 (1). <https://doi.org/10.24242/jclis.v3i1.96>.
- Cook, Terry. 1992. "The Concept of the Archival Fonds in the Post-Custodial Era: Theory, Problems and Solutions." *Archivaria* 35: 24–37.
- Gaetarelli, Cecilia. "Living archives: designing digital archives as spaces for creative and imaginative practices." Tesi, Politecnico di Milano, 2021.
- Grispigni, Marco, e Leonardo Musci. 2003. *Guida alle fonti per la storia dei movimenti in Italia 1966-1978*. Roma: Istituto Poligrafico e Zecca dello Stato – Archivi di Stato.
- Howard, Mark, Katherine Jarvie, e Steve Wright. 2021. "Rancière, political theory and activist community appraisal." *Archives and Manuscripts* 49 (3): 208–27. <https://doi.org/10.1080/01576895.2021.1987938>.
- Marquez, Gabriel Garcia. 2002. *Vivir para contarla*. Barcelona: DeBolsillo.
- Musci, Leonardo. 2021. "Movimenti e archivi. Punti fermi e questioni aperte." In *Tramandare la memoria sociale del Novecento. L'archivio di Gino Cerrito presso la Biblioteca di scienze sociali dell'Università di Firenze. Atti della giornata di studio (Firenze, 21 novembre 2019)*, a cura di Enrica Boldrini e Lucilla Conigliello, 81–85. Firenze: Firenze University Press. <https://doi.org/10.36253/978-88-5518-289-8.08>.
- Pezzica, Lorenzo. 2020. *L'archivio liberato. Guida teorica-pratica ai fondi storici del Novecento*. Milano: Editrice Bibliografica.
- Rhodes, Tamara. 2014. "A living, breathing revolution: How libraries can use 'living archives' to support, engage, and document social movements." *IFLA Journal* 40 (1): 5–11. <https://doi.org/10.1177/0340035214526536>.
- Valacchi, Federico. 2023. "Se l'archivio è artificiale. Verso uno ius archivi partecipativo?." *AIDA Informazioni* 1–2: 153–70. <https://doi.org/10.57574/596529288>.

## The Nightingale of Keats. Documents, archives, and information objects

Camilla Domenella<sup>(a)</sup>, Riccardo Fedriga<sup>(b)</sup>, Margherita Mattioni<sup>(c)</sup>

a) University of Macerata, <https://orcid.org/0000-0002-4700-811X>

b) Alma Mater Studiorum – University of Bologna, <https://orcid.org/0000-0002-2291-7800>

c) Alma Mater Studiorum – University of Bologna, <https://orcid.org/0009-0009-8472-579X>

**Contact:** Camilla Domenella, [c.domenella@unimc.it](mailto:c.domenella@unimc.it); Riccardo Fedriga, [riccardo.fedriga@unibo.it](mailto:riccardo.fedriga@unibo.it);  
Margherita Mattioni, [margherita.mattioni2@unibo.it](mailto:margherita.mattioni2@unibo.it)

**Received:** 27 December 2023; **Accepted:** 15 March 2024; **First Published:** 15 May 2024

### ABSTRACT

The paper first analyzes the complexity of the notion of the archival document, in its fundamental definition and its updates prompted by the shift to the digital dimension. It then sets out to consider – from a digital hermeneutic point of view – the notion of ‘document’ from the perspective of the information object and therefore of the intentional object. Finally, the paper discusses the ways in which the archive is a technological enhancement for the interrogability and ‘consultability’ of knowledge.

### KEYWORDS

Archival Records; Information Object; Intentionality; Social Object; Information Selection.

## L’usignolo di Keats. Documenti, archivi e oggetti informativi

### ABSTRACT

Il contributo analizza in prima battuta la complessità della nozione di documento archivistico, nella sua definizione fondamentale e nei suoi aggiornamenti sollecitati dal passaggio alla dimensione digitale. Si pone poi l’obiettivo di considerare – in chiave ermeneutica – la nozione di documento sotto il profilo dell’oggetto informativo e quindi dell’oggetto intenzionale, per inquadrare infine le modalità secondo le quali l’archivio è protesi tecnologica per l’interrogabilità e la consultabilità di sapere e conoscenza.

### PAROLE CHIAVE

Documento archivistico; Record; Oggetto informativo; Oggetto intenzionale; Archivi.

## Che cos'è un documento

Chiedersi oggi che cos'è un documento significa avviare una riflessione sul sapere e la conoscenza: sui modi, cioè, in cui questi sono intesi, prodotti, registrati, archiviati, condivisi e infine diffusi. Questa prospettiva – ambiziosa ma a nostro avviso necessaria – intende dunque chiedersi in che senso il documento rappresenti un elemento di congiunzione fondativa tra le condizioni che rendono possibile la relazione tra atti di conoscenza, oggetti del sapere e società.

In questo quadro, la definizione di documento è assai più complessa di quel che sembra ad un primo sguardo: come tutte le cose il cui utilizzo ci appare scontato, infatti, la loro definizione risulta assai più difficile. Per cercare di venire a capo della difficoltà, invece di avviare la nostra indagine da una definizione astratta e generica di documento, scegliamo di partire da una sua applicazione in atto – forse la più evidente al senso comune: il documento archivistico. Ma anche qui i problemi non sono pochi. Infatti, il documento archivistico – ovvero quell'oggetto linguistico depositario della memoria individuale e collettiva, redatto o composto in maniera tale da dover essere conservato e consultato – è così dipendente dalle proprie caratteristiche da esserne indistinguibile. Non solo. Nel passaggio dall'archivio cartaceo all'archivio digitale, la nozione di documento ha infatti subito una varietà di aggiornamenti, i quali, se da un lato mirano a descriverne e giustificarne la natura che da analogica si fa digitale, dall'altro lato tendono a mantenerne le finalità e le caratteristiche essenziali che ne permettono l'identificazione e la diagnostica<sup>1</sup>.

Questo scivolamento definitorio è particolarmente evidente nel passaggio dalla nozione di documento come oggetto cartaceo, materiale e fisico, a quella di *record* archivistico di natura informatica, la quale predilige letteralmente la funzione di registrazione dell'informazione<sup>2</sup>.

Tuttavia, tanto nel caso dell'analogico quanto in quello del digitale, bisogna ricordare come il documento sia anzitutto un *contenuto informativo* che è tale perché corredato da una serie di descrizioni che ne restituiscono il contesto di individuazione, formazione e classificazione. In questo quadro, il documento, sia esso fisico oppure elettronico, rappresenta l'*oggetto reale* (rispettivamente oggetto-documentale nel caso del documento fisico, e oggetto-dati nel caso del documento elet-

---

<sup>1</sup> Dalle prime definizioni offerte dalla diplomazia alle ultime dedicate al documento elettronico e a quello informatico stabilite dal regolamento europeo eIDAS (electronic IDentification Authentication and Signature) e da quello italiano CAD (Codice Amministrazione Digitale) la nozione di documento archivistico è andata via via modificandosi, aderendo anzitutto al quadro giuridico di riferimento e ricomprendendo le possibilità tecnologiche emerse nei processi di digitalizzazione. Paola Carucci scrive che un documento archivistico è una “*rappresentazione* in forma libera o secondo determinati requisiti di un fatto o di un atto relativo allo svolgimento dell'attività istituzionale, statutaria o professionale di un ente o di una persona” (Carucci e Messina 1998, 29). Tale definizione fa eco alla nozione di *record*, sottolineata da Maria Guercio, secondo cui “a record is the recorded *representation* of an act, produced in a specific form – the form prescribed by the legal system – by a creator in the course of its activity” (Guercio 1997, 221). Con Pigliapoco, ricordiamo inoltre come il documento nasca “in modo naturale, come necessità pratica per lo svolgimento delle attività e una volta formato entra a far parte dell'archivio per volontà del soggetto produttore, che può decidere di conservarlo per obbligo di legge o semplicemente per avere memoria dell'atto rappresentato” (Pigliapoco 2016, 14). In questo senso, sia nel caso del documento analogico che nel caso del documento elettronico, parliamo dunque di una *rappresentazione di fatti o atti*, che – come vedremo fra un attimo – si manifesta come contenuto informativo. Sempre sull'OAIS in relazione alla nozione di documento digitale, si veda (Sebastiani 2008:14)

<sup>2</sup> L'accezione di *record* è certamente più ampia e non si estende esclusivamente alla natura informatica o elettronica del documento. Ciò che qui si intende sottolineare è come la nozione di *record* sembri porre maggiormente l'accento sulla funzione di *registrazione* dell'informazione, mentre la nozione di documento appare piuttosto legata al risultato di tale registrazione. La sfumatura – puramente semantica – ci sembra tuttavia interessante sul piano concettuale.



tronico) cui le descrizioni e le informazioni si riferiscono. In tal senso, l'oggetto reale è dunque il punto di ancoraggio del corredo informativo e rappresenta l'unità minima, di tipo non strutturato e non semantico, per la rappresentazione del contenuto informativo. L'insieme dell'oggetto reale e delle sue descrizioni è invece ciò che chiameremo *oggetto informativo*.

## Documenti e oggetti informativi

Per chiarire la relazione tra oggetto reale e oggetto informativo, ci muoveremo dunque a partire dalle riflessioni compiute nel campo dell'archivistica digitale. Pur non esaurendo la nostra trattazione all'interno di questa prospettiva, rileviamo come la 'traduzione' digitale della nozione di documento archivistico offra un tracciato categoriale utile alla comprensione della nostra proposta. Dal punto di vista strettamente archivistico, un oggetto informativo si presenta dunque come l'insieme di due elementi: l'oggetto in senso proprio – che può essere digitale oppure fisico, e che d'ora in avanti chiameremo *documento* –, e l'insieme di informazioni (*corredo informativo*) che ne descrivono il contesto e ne permettono la rappresentazione.

Questa impostazione tiene conto del modello concettuale OAIS (Open Archival Information System) descritto nello standard ISO 14721: 2012, su cui si fonda la maggioranza dei progetti di *long-term digital preservation*. Il riferimento al modello OAIS ci consente di mettere in luce la 'doppia natura' del documento, digitale oppure fisico, e contemporaneamente sottolineare la sua unitarietà come oggetto informativo. OAIS pone infatti l'accento sulla dimensione digitale dell'*archivio* (e non del documento), che si attesta quale *processo* di conservazione, senza tuttavia limitare l'applicazione di tale processo ai soli documenti elettronici. La scelta di attenerci al modello OAIS ha inoltre il pregio di proporre una modellizzazione concettuale e non direttamente tecnologica, offrendo dunque il *framework* di riferimento per una tematizzazione filosofica dell'oggetto informativo. Il nostro punto di partenza è quindi quello di muoverci in ambiente digitale (quello dell'archivio OAIS) per evidenziarne la specificità senza tuttavia trascurare la possibile duplice natura del documento.

Nel contesto digitale OAIS le singole parti che riguardano l'oggetto informativo e il loro schema di relazioni è così riassunto:

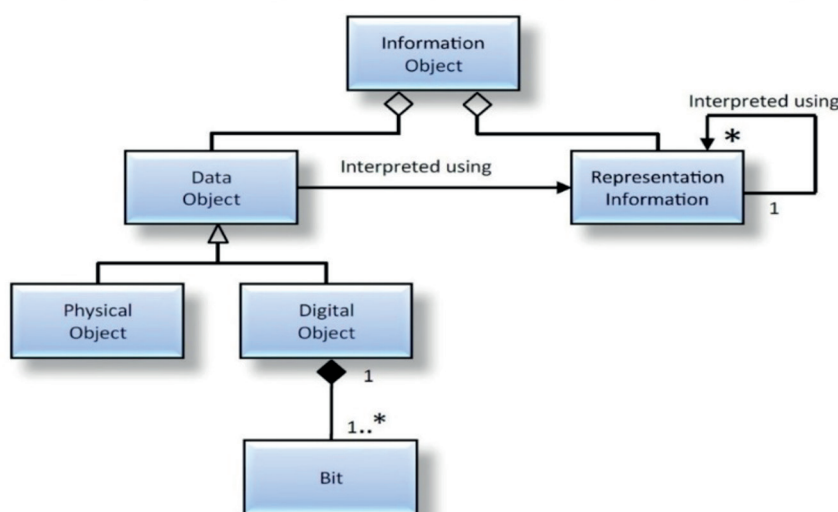


Figura 1. Struttura dell'oggetto informativo secondo il modello concettuale OAIS

L'oggetto informativo – insieme di dati e metadati – non è dunque esclusivamente corrispondente al documento (oggetto-dati) che si intende archiviare ma non è neppure, semplicemente, il corredo delle informazioni a esso collegato. In quanto insieme di entrambi, l'oggetto informativo si presenta come l'intreccio delle reciproche relazioni che insistono fra il documento e i metadati e, dunque, le informazioni che lo descrivono.

È importante notare che questo insieme di relazioni non interviene soltanto fra dati e metadati, ma anche fra un metadato e un altro, agendo dunque su più livelli. In questo senso, l'oggetto informativo non è semplicemente la somma degli elementi 'oggetto-dati' e 'metadati', quanto piuttosto il sistema di mutue relazioni di codifica e descrizione che ne consentono – infine – l'interpretabilità<sup>3</sup>.

L'implementazione delle relazioni attesta non soltanto l'affidabilità archivistica dell'oggetto informativo – la possibilità di ricostruire il contesto di produzione di un *record* è un esempio della *affidabilità* archivistica<sup>4</sup> – ma anche l'affidabilità rispetto al flusso di credenze che si riversano sul documento<sup>5</sup>. L'interpretazione dell'oggetto informativo *dipende* infatti dall'intreccio di relazioni delle informazioni che intercorrono fra dati e metadati. Sotto questo punto di vista, l'oggetto informativo è quindi un ente informazionale<sup>6</sup>, che manifesta la propria presenza nell'ordine di una *mediazione relazionale* la cui determinabilità è scissa dalla presenza reale.

---

<sup>3</sup> L'apertura semiotica del documento è stata riconosciuta, almeno in parte, come una delle sue cifre costitutive da diversi studiosi nei settori della biblioteconomia, della documentazione e della scienza dell'informazione. Per quanto riguarda il primo ambito, rimandiamo, fra gli altri, ad Alfredo Serrai, che pensa ai documenti come a 'oggetti che portano segni' la cui decodifica richiede un sistema condiviso di interpretazione/decodifica (Serrai 1973), ma anche a Piero Cavaleri, che li concepisce come "complessi semiotici capaci di produrre significato nel momento in cui vengono interpretati" (Cavaleri 2013, 27). Per quanto concerne, invece, l'*information science* ricordiamo, all'interno del panorama italiano, l'*Introduzione alla scienza dell'informazione* di Alberto Salarelli (2012), nella quale, riprendendo in certa misura l'Umberto Eco del *Trattato di semiotica generale* (1975), l'autore assegna un ruolo essenziale agli schemi di (de)codifica nelle procedure di individuazione e comprensione dei dati informativi contenuti nei documenti. Tuttavia, è bene fin da subito sottolineare come, al contrario di queste interpretazioni, così come della famosa teoria del documento stipulata da Suzanne Briet (Briet 1951), la nostra prospettiva d'analisi si discosta da una concettualizzazione precipuamente semiotica della realtà documentale, preferendo indagarla in termini di legami reciproci tra a) il mondo dei documenti, b) l'ontologia sociale (intesa, come si vedrà, in termini di "intelligenza sociale") e c) la memoria condivisa. Per una breve ma efficace panoramica sulle teorie del documento del XX secolo rinviamo, infine, a Buckland 1997 e *Id.* 2018.

<sup>4</sup> Come ricordiamo, a titolo emblematico, con Nicola Barbuti, i *record* sono "entità dinamiche e diacroniche che registrano e conservano nelle descrizioni i processi di digitalizzazione che li hanno creati e quelli che hanno caratterizzato il loro successivo ciclo di vita". In quanto rappresentazioni persistenti costituite da dati e metadati, gli oggetti informativi necessitano delle descrizioni contestuali. Senza il corredo informativo, infatti, un dato è soltanto un dato, e per quanto esso sia "ricercabile, accessibile e interoperabile con altri non fornisce alcuna garanzia di qualità, sufficienza e affidabilità dei contenuti informativi che contiene" (Barbuti 2020, 20).

<sup>5</sup> Il tema del flusso di credenza che si riversa sul documento è ben sottolineato da Ferraris nella sua definizione di oggetto sociale. Secondo Ferraris, nel documento in quanto oggetto sociale "la credenza risulta determinante dell'essere", vale a dire che è indispensabile il *riconoscimento* degli oggetti sociali da parte di soggetti, altrimenti gli oggetti sociali (e i documenti che sono la forma più elevata degli oggetti sociali) non esisterebbero. (Ferraris 2014, 359-360). Degli oggetti sociali tratteremo più avanti. Quello che qui ci preme sottolineare è come l'affidabilità del documento – corredato delle informazioni che lo descrivono – sia fondamentale non soltanto per la dimensione costitutiva dell'oggetto informativo ma anche per la dimensione formativa dell'oggetto intenzionale e per quella performativa dell'oggetto sociale. Questi oggetti – come vedremo – non sono oggetti ontologicamente separati, ma gradi e modo di uno stesso 'percorso' aperto di interpretabilità. Si veda, Eco 2011a, 81 sgg.

<sup>6</sup> La nozione di "ente informazionale" fa qui riferimento alla prospettiva offerta da Luciano Floridi, per il quale l'"ente informazionale" sembra essere l'"elemento minimo di significato da cui ha origine qualsivoglia postulazione". (Manna 2021,

Ciò che è importante notare, dunque, è come l'oggetto informativo non esaurisca la propria dimensione epistemologica ed ermeneutica nel contesto delle relazioni di cui pure esso è costituito. Se così fosse, infatti,

1. il documento (in questo caso l'oggetto-dati) verrebbe per così dire reso fluido, assimilato e 'riassorbito' nel sistema di relazioni atte invece a descriverlo;
2. i metadati non troverebbero alcun ancoraggio all'oggetto reale, venendo dunque a sostituirsi all'oggetto stesso in una inesorabile deriva di descrizioni.

I contesti di codifica degli oggetti informativi producono infatti significato e tale significato costituisce un elemento di senso irrinunciabile per la loro stessa costituzione: i tracciati di metadati equivalgono, sotto questo rispetto, a specifici set di regole interpretative che funzionalizzano i segni esibiti dagli oggetti reali, in quanto attivano alcuni di essi a scapito di altri, agendo, quindi, ad un livello di significazione degli oggetti. D'altro canto, pur essendo *in sé* in modo indipendente dal campo di relazioni in cui sono inseriti, gli oggetti reali si trasformano in oggetti informativi solo se inseriti all'interno di uno o più livelli di codifica che ne esplicitino alcune proprietà rilevanti ponendone, per forza di cose, in latenza delle altre.

Questa osservazione mette in luce una tensione che è bene illustrata dal caso del trattamento digitale dei *marginalia* d'autore. Cosa diremmo, infatti, di una traccia manoscritta posta in modo ricorrente a commento dei testi che viene sussunta sotto un'unica categoria analitica affinché quest'ultima possa essere universalmente richiamata, senza ulteriori distinzioni di contenuto o contesto, nella descrizione di tutte le sue occorrenze all'interno di una data unità documentaria? L'univocità e l'omogeneità sono due cifre imprescindibili per la costruzione di modelli coerenti di descrizione degli oggetti, certo, ma la ricorrenza di una stessa notazione a margine apposta su documenti diversi e in momenti di lettura differenti può assumere valori e funzioni qualitativamente distanti fra loro e veicolare, perciò, *significati* differenti. E proprio tale aspetto qualitativo è ciò che l'adozione di uno schema di *mark up* standardizzato corre il rischio di trascurare. Questo esempio ci rimanda in realtà ad una posizione di buon senso realista secondo cui l'esigenza descrittiva non può essere assiologicamente subordinata ad alcuna, per quanto importante, esigenza di interoperabilità, né dovrebbe essere trascesa in funzione della stessa: ciò che descrive non può in nessun caso sostituirsi a ciò che è descritto *salva veritate* (Sokolov 2009, 60-61). Questo tipo di descrizioni meta-testuali o meta-informative devono di conseguenza essere assunte come generalizzazioni incomplete degli oggetti-dati che, all'interno di *protesi di memoria* come gli archivi, *simulano* ma non sostituiscono né tantomeno restituiscono in alcun modo gli oggetti stessi, limitandosi a permetterci di identificarli quando interroghiamo tali protesi (o sistemi).

Restando, dunque, nel campo della digitalizzazione del patrimonio culturale, ogni oggetto informativo presuppone che il relativo oggetto reale sia anzitutto traslato e 'tradotto' in sequenze di

---

164-74). A nostro avviso, tuttavia, la minima unità completa per la produzione di significato non è l'ente informazionale, bensì l'oggetto informativo. È anzi l'oggetto informativo a fondare – per così dire – l'ente informazionale. Proponiamo quindi un'inversione di priorità tra ente informazionale e oggetto informativo: quest'ultimo garantisce infatti l'ancoraggio all'oggetto reale aprendo al contempo ad una mediazione relazionale – più specificamente intenzionale – che gli è *in nuce*, e dunque in potenza.

bit, cioè di atomi informativi di tipo *strutturale* (e non semantico)<sup>7</sup> e quindi in ente informazionale digitale, e che a questa prima fase di acquisizione segua l'attribuzione di una qualche descrizione standardizzata, che dovrebbe variare in base al contenuto alla quale va applicata, come nel caso dell'archivio d'autore. Diversamente dagli scarni metadati generati di *default* da qualsiasi operazione digitale (pensiamo alla data e all'ora in cui è scattata un'immagine o in cui è stato inviato un messaggio), la metadazione descrittiva, amministrativa e strutturale che accompagna l'implementazione di biblioteche e archivi digitali *produce* il corredo informativo dei rispettivi oggetti digitali, eppure lo fa rimanendo in un ordine meramente rappresentazionale. Ciò significa che tale corredo non si equivale a una struttura che *dice*, o riflette, l'essenza stessa dei contenuti espressi nei documenti o, men che meno, degli oggetti reali – materiali o astratti – rappresentati dagli oggetti informativi, bensì consiste in un set di *segni*, che non solo restano tali ma “non è necessario divengano altro” (Cavaleri 2013, 270).

## Oggetti informativi e oggetti intenzionali

Questo scenario favorisce una lettura degli oggetti informativi come oggetti intenzionali. Espressione che risale al filosofo austriaco Franz Brentano – poi a lungo sviluppata nelle tradizioni fenomenologica e analitica – quella di oggetto intenzionale riguarda, in origine, una caratteristica peculiare degli atti mentali: quella di riferirsi a certi oggetti che non necessariamente esistono, o possono esistere in modo diverso da come li percepiamo. Che cosa sono, dunque, questi oggetti? Come ci rivolgiamo e riferiamo a essi? Il modo in cui ne parliamo nei nostri resoconti linguistici può essere rivelatore di schemi mentali che ne giustificano l'esistenza? E ancora: nel caso che tali oggetti non esistano nella realtà, come nel caso di oggetti fittizi o illusori, dobbiamo postulare differenti livelli di esistenza? Qualora essi esistano, nel nostro dirigerci verso di essi non vi è alcun motivo di postulare l'esistenza di due oggetti, uno reale e uno nella mente che lo rappresenta. Se penso, per esempio, alla Donazione di Costantino – intesa come testimone documentale e non come atto giuridico – non sto pensando alla sua rappresentazione, come se fosse una riproduzione più piccola nella mente grazie alla quale posso avervi accesso, ma proprio all'oggetto fisico. Ma non necessariamente quella Donazione esiste, o esiste in quel momento, in quello spazio etc. Sia che esista sia che non esista sia che, inoltre, esista diversamente da come si percepisce, si pensa, si crede l'oggetto, l'idea che sta alla base di questa concezione è che vi sia una molteplicità di riferimenti e di piani di discorso che riportano tali riferimenti. Il nostro punto di vista è che tali piani, lungi dall'escludersi o dal costituirsi di moltiplicarsi di livelli ontologici, uno dentro l'altro, siano piuttosto costitutivamente complementari.

La nozione di oggetto intenzionale, inteso come nucleo unitario di relazioni e proprietà date nelle varie rappresentazioni, presuppone che ci sia un complementare riferimento a un oggetto esterno, reale (qualora questo ci sia). Il che non comporta che le due nozioni stiano una *dentro* l'altra in una sorta di precedenza di tipo metafisico-esistenziale, né che l'una sia *accanto* all'altra in una

---

<sup>7</sup> Attingiamo qui alla distinzione fra informazione strutturale e informazione semantica proposta da David Chalmers perché consente di disambiguare, quantomeno nelle applicazioni digitali, l'accezione, rispettivamente, strutturale/materiale di 'informazione', per la quale quest'ultima è un insieme di sequenze di *bit* allocate su supporti fisici, dalla sua accezione – appunto – semantica, in virtù della quale l'informazione si riferisce a fatti o a proposizioni (Chalmers 2022, 152).

stessa ontologia né, infine, che siano *una* immagine e *una* sua rappresentazione. Diversamente, la risultante è che vi sia una possibile ambiguità di resoconti, attraverso i quali si rendono i piani di riferimento, e che tale ambiguità sia inemendabile. Gli oggetti intenzionali, ed è per questo che li utilizziamo in questo contesto, non solo ci assicurano la possibilità di distinguere tra le cose in quanto tali, il loro modo di darsi nelle singole rappresentazioni (per esempio le singole e idiosincratice interrogazioni di un archivio), e la possibilità di darne un aggregato unitario di dati o descrizioni; essi possono anzi essere concettualizzati a partire dalla loro proprietà di essere portatori di informazione e quindi produttori, in potenza, di significato. Infine, possono essere funzioni identificative di oggetti fisici, complementari e ambigui da un punto di vista epistemologico, ma mai sovrapponibili.

Il passaggio da oggetto informativo a oggetto intenzionale è quindi regolato dalle relazioni che intercorrono fra l'oggetto reale, l'oggetto informativo e lo stesso oggetto intenzionale. L'oggetto intenzionale si presenta a questo punto come entità di riferimento, unitaria e condivisa, che permette di identificare gli oggetti reali sotto il *modo* degli oggetti informativi.<sup>8</sup> In questo senso, l'oggetto intenzionale non si sostituisce all'oggetto fisicamente esistente – non ne è né il simulacro né la simulazione – e neppure lo crea. Lo schema rappresentativo dell'oggetto intenzionale e dell'oggetto reale riflette, a un livello superiore, quello dell'oggetto informativo.

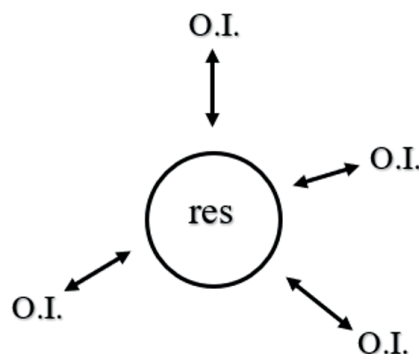


Figura 2. Oggetti intenzionali e oggetto reale (*res*)

Qualora vi sia, e nell'esperienza comune tale modalità di esistenza è un *habitus* talmente forte da farci comunque percepire oggetti unitari a partire da frammenti o completi da punti di vista parziali, l'oggetto fisicamente esistente *precede* – sia sul piano ontologico sia su quello epistemologico – l'insieme degli oggetti intenzionali. Ma gli oggetti intenzionali non necessariamente esistono: possiamo pensare, credere, desiderare e persino amare gli unicorni.

Torniamo alla nostra Donazione di Costantino. L'oggetto reale della Donazione di Costantino corrisponde al suo essere un documento: la *res* è il documento, io sono di fronte ad un documento. Questo è tutto ciò che 'sappiamo'. Detto in altri termini: io non so se esso sia un *documento* vero o falso, non so se esso è la 'Donazione di Costantino', non so se il suo *contenuto* sia vero o falso, non so collocarlo storicamente e neppure, poniamo, archivisticamente. È qui che vengono in aiuto

<sup>8</sup> Per un panorama circa identificatori persistenti per oggetti digitali, cfr. (Sebastiani, 2005), in part. 70-71.



gli oggetti intenzionali. Un oggetto intenzionale potrebbe indicarmi che il documento è archivisticamente valido, un ‘secondo’ oggetto intenzionale potrebbe indicarmi che il documento è la ‘Donazione di Costantino’, un ‘terzo’ oggetto intenzionale potrebbe farmi credere, tanto a livello individuale quanto come credenza collettiva, che questo documento è filologicamente incoerente e un ultimo oggetto intenzionale potrebbe infine rivelarmi che, sino a quando non ne sarà scoperta un’altra vera e l’asserzione sottoposta a un processo di revisione, ‘questa Donazione di Costantino’ è storicamente falsa. Nessuno di questi oggetti intenzionali ha cancellato o ‘creato’ l’oggetto reale (il documento), né tantomeno si è ad esso sostituito<sup>9</sup>.

Nella relazione intenzionale, vi è dunque una riserva apparentemente inestinguibile di virtualità e di ambiguità che non sono in alcun modo legate all’aspetto simulativo: in linea di principio, infatti, gli oggetti reali ammettono una serie potenzialmente indefinita di atti e quindi di opzioni – circostanza che, tradotta nel linguaggio della tradizione filosofica peirceana, implica che gli oggetti reali sono, di per sé, inevitabilmente interessati da un meccanismo virtuale di *semiosi* illimitata (Peirce 1980). L’oggetto informativo, invece, sembra per contro darsi nel momento in cui tale deriva interpretativa, ineluttabilmente scatenata dall’oggetto reale nella sua veste di ente informazionale, s’interrompe o, meglio, viene interrotta, cioè ogni qualvolta l’oggetto viene ‘catturato’ all’interno della trama di relazioni, la quale viene attivata da uno specifico sistema di decodifica (come abbiamo chiarito nell’esempio di dati e metadati) ma anche e soprattutto da un contesto di significazione.

L’ente informazionale si fa così oggetto informativo per opera del suo essere all’interno di un processo di senso che assolve al compito di dare rappresentazione, e quindi forma, all’insieme di dati grezzi esibiti dall’oggetto medesimo sotto forma di segni o di resoconti linguistici nel linguaggio naturale (Eco 2023). È quindi in forza di questo processo se il documento della Donazione di Costantino assume su di sé una vera e propria struttura informativa aperta, a seguito dell’imposizione di una griglia interpretativa che ponga in latenza ogni altro potenziale schema di intelligibilità. Gli oggetti intenzionali, allora, intesi come galassie generiche, servono per l’identificazione degli oggetti presenti nei nostri depositi culturali. Alla luce di tale interpretazione, la dimensione dell’archivio si manifesta non tanto come struttura o come schema classificatorio, quanto piuttosto come processo di senso che si basa sempre su una soglia, per quanto minima, di realtà ed è volto a darle rappresentazione generica. Sotto questo aspetto, l’archivio, è dunque il concreto strumento che, sotto il modo degli oggetti informativi, agisce per l’identificazione degli oggetti reali. Parimenti, in quanto sistema aperto di intenzionalità esso svolge, in ultima analisi, una funzione dialettica fra evidenza e latenza, potenzialità e attualità.

In questo contesto, dunque, l’archivio manifesta il proprio concetto nel suo essere *per tutti*: esso custodisce la memoria e la conoscenza universali, laddove per universale non si intende qui la

---

<sup>9</sup> L’enucleazione degli oggetti intenzionali è qui puramente indicativa ed esemplificativa. Sul tema, facciamo nostra la metafora dello specchio evocata da Elena Esposito per mettere in luce la dimensione in questo senso essenzialmente *operativa* o *performativa* della realtà virtuale: similmente a ciò che avviene con gli oggetti riflessi da uno specchio, i prodotti della virtualizzazione non sono infatti da intendersi come veri e propri duplicati degli oggetti reali, poiché ad essere ‘duplicata’ è soltanto l’osservazione degli oggetti stessi. Per quanto le possibilità di manipolazione dei dati della virtualità digitale siano, com’è evidente, di gran lunga maggiori a quelle dell’azione riflettente di uno specchio, la virtualizzazione degli oggetti non produce, in ogni caso, alcuna duplicazione sul piano della realtà oggettuale. Detto altrimenti, la virtualizzazione non comporta alcuna sostituzione in quanto l’oggetto virtuale, pur rappresentando l’oggetto reale, di fatto “non fa le veci di nulla” (Esposito 2001, 236-237).

totalità e l'universalità del sapere, quanto piuttosto la trasversalità unitaria di cui l'archivio stesso si compone. In questo senso, il concetto di archivio può essere linguisticamente reso attraverso un'immagine: esso è la *metafora* di un concetto *agente* della conoscenza collettiva e individuale. Tale metafora si mantiene *viva*, peraltro, anche nel passaggio dal cartaceo al digitale in quanto consente di rintracciare al suo interno una dimensione poetica. Essa è legata, in primo luogo all'azione intenzionale dell'interrogazione, intesa come interpretazione e messa in forma dei dati da parte dell'utente nel darsi della stessa attività ermeneutica in un circolo organico di domande e risposte che mettono in luce alcuni aspetti del deposito culturale. Sotto questo punto di vista, l'archivio, come nell'«Usignolo di Keats» richiamato da Borges in *Altre inquisizioni*, è universale e imperituro nella specie perché informato dall'individuo interrogante a partire, va ribadito, da una soglia (minima) della realtà (Eco 1997; Ferraris 2018). La dimensione realista di tale relazione si esplica infatti proprio nella funzione di interrogabilità e di consultabilità dell'archivio: una funzione di ordine superiore rispetto alla singolarità interrogante che, pure, non deriva dai concetti ma li precede, facendoli emergere e rendendo attuale il carattere trascendentale della funzione stessa. Poter interrogare l'archivio significa infatti individuare e di conseguenza selezionare una (o più) delle *possibilità* custodite al suo interno, escludendo le altre che restano invece in latenza.

## Conclusioni

Questa considerazione conduce a due evidenze conseguenti, una di natura teoretica e l'altra di natura etica. In primo luogo, l'interrogabilità dell'archivio ci pone sulle tracce di una protesi tecnologica grazie alla quale una funzione del possibile diventa agente. L'archivio cartaceo, quello ibrido e quello digitale hanno stessa struttura e stesse caratteristiche. Il processo di selezione non riguarda infatti soltanto l'aspetto strettamente archivistico, per cui si interviene con le pratiche di selezione e scarto ma interviene anche nel momento – a nostro avviso altrettanto determinante – di restituzione di ciò che l'archivio custodisce. Ma è in questa coincidenza di selezione e interrogazione che si esprime la *funzione del possibile*, che diventa appunto *agente*, dell'archivio. Detto in altri termini, l'archivio agisce funzionalizzando il possibile, producendo attualità quando viene interrogato, nella misura in cui in esso tutte le possibilità sono in potenza. Assistiamo, dunque, ad una sorta di ribaltamento: l'archivio è il 'luogo' degli oggetti intenzionali in atto, l'interrogabilità (e la selezione per esclusione che ne consegue) è il 'momento' di attuazione dei possibili. In questo quadro, si compie dunque un passaggio fondamentale, sotteso all'intera rappresentazione epistemologica ed ermeneutica dell'archivio, che fonda il passaggio da *archivio per tutti* ad *archivio per me*. Più specificamente, la distinzione tra archivio per tutti e archivio per me è una distinzione produttiva: nel primo caso – quello dell'*archivio per tutti* – l'archivio produce possibilità; nel secondo caso – quello dell'*archivio per me* – l'archivio produce attualità perché lo interrogo, attualizzando le possibilità in esso virtualmente contenute.

La cifra di tale ribaltamento è da rintracciarsi proprio nell'intenzionalità 'connaturata' agli oggetti informativi che compongono e giustificano ontologicamente l'archivio. Quest'ultimo, proprio perché intelletto agente, si presenta come intenzionalità collettiva e non come somma di intenzionalità individuali. Le intenzionalità individuali sono anzi 'agite', 'rese possibili' dall'intenzionalità collettiva, cioè dalla *natura sociale e condivisa dell'archivio*. Vale a dire: gli oggetti intenzionali non

sono la condensazione di intenzionalità individuali, bensì la rappresentazione di una intenzionalità collettiva, nel passaggio da *archivio per tutti* ad *archivio per me*. Come si evince, questo ‘*per me*’ non cancella l’archivio, allo stesso modo in cui l’oggetto intenzionale non cancella, né crea, né sostituisce l’oggetto reale.

La seconda evidenza, di natura etica, riguarda il fatto che il processo di selezione – tanto nella sua accezione strettamente archivistica quanto nella funzione di interrogabilità – comporta un richiamo fondamentale alla responsabilità: un impegno a formare che va ben al di là della semplice contemplazione, dall’esterno, di una struttura formale. Se selezionare è escludere, mantenendo in latenza ciò che si è escluso, ecco che la responsabilità emerge necessariamente nella definizione e nella individuazione di ‘ciò che è dentro’ e ‘ciò che è fuori’. Tale responsabilità è naturalmente sancita e regolata dalla professione archivistica a cui è attribuito il compito di scarto e di selezione dei documenti da inviare in conservazione<sup>10</sup>, ma essa non è di minore importanza dal lato utente, nella consultabilità e interrogazione che esso compie.

---

<sup>10</sup> Sul tema della responsabilità etica degli archivisti e degli archivi, rimandiamo al testo di Federico Valacchi, il quale sottolinea: “L’uso degli archivi, e nell’uso ricomprendo anche tutte quelle attività necessarie a renderli effettivamente fruibili, è innanzitutto un fatto etico perché presuppone scelte di ordine intellettuale e morale che subordinano la tecnica alla volontà di costruire modelli identitari, basati su una corretta circolazione di informazione quanto più possibile corretta e quanto più possibile ‘archivistica’” (Valacchi 2020, 159).

## Riferimenti bibliografici

- Barbuti, Nicola. 2020. “Ripensare i dati come risorse digitali: un processo difficile?.” In *La svolta inevitabile: sfide e prospettive per l’Informatica Umanistica*, a cura di Cristina Marras, Marco Passarotti, Greta Franzini, e Eleonora Litta, 19–22. Milano: Università Cattolica del Sacro Cuore.
- Briet, Suzanne. 1951. “Qu’est-ce que la documentation?.” Paris: EDIT.
- Buckland, Michael K. 1997. “What Is a “Document”?.” *Journal of the American Society for Information Science* 48 (9): 804–809.
- Buckland, Michael. 2018. “Document Theory.” *Knowledge Organization* 45(5): 425-436. 71 references. <https://doi.org/10.5771/0943-7444-2018-5-425>.
- Carucci, Paola. 2010. *Le fonti archivistiche: ordinamento e conservazione*. Roma: Carocci.
- Carucci, Paola, e Marina Messina. 1998. *Manuale di archivistica per l’impresa*. Vol. 19. Roma: Carocci.
- Cavaleri, Piero. 2013. *La biblioteca crea significato: thesaurus, termini e concetti*. Milano: Editrice Bibliografica.
- Chalmers, David J. 2022. *Reality+: Virtual worlds and the problems of philosophy*. London: Penguin.
- Eco, Umberto. 1997. *Kant e l’ornitorinco*. Milano: Bompiani.
- Eco, Umberto. 2011a. *Confessions of a young novelist*. Cambridge (MA): Harvard University Press.
- Eco, Umberto. 2011b. *Trattato di semiotica generale*. Milano: Bompiani.
- Eco, Umberto. 2016. *I limiti dell’interpretazione*. Milano: La Nave di Teseo.
- Eco, Umberto. 2023. *Opera aperta forma e indeterminazione nelle poetiche contemporanee. Nuova edizione con i materiali preparatori dell’autore*. A cura di Riccardo Fedriga. Milano: La nave di Teseo.
- Esposito, Elena. 2001. *La memoria sociale: mezzi comunicare e modi di dimenticare*. Roma: Laterza.
- Ferraris, Maurizio. 2014. *Documentalità: perché è necessario lasciar tracce*. Bari-Roma: GLF editori Laterza.
- Ferraris, Maurizio. 2018. “Ermeneutica neorealista.” *Giornale di metafisica* XL (2). <http://digital.casalini.it/4476939>.
- Guercio, Maria. 1997. “Definitions of Electronic Records, the European Perspective.” *Archives and Museum Informatics* 11 (3–4): 219–22. <https://doi.org/10.1023/A:1009029211533>.
- Lodolini, Elio. 1987. *Archivistica: principi e problemi*. Milano: FrancoAngeli.
- Manna, Leonardo. 2021. “Il pensiero come relazione o intero semantico? Intorno alla filosofia di Luciano Floridi.” *Phenomenology and Mind*, 20: 164–74.
- Nicolai, Gilda. 2017. “Dagli archivi tradizionali all’ambiente digitale: la valutazione e selezione nel contesto internazionale”, in *Archivi*, XII/1: 29-46
- Peirce, Charles S. 1980. *Semiotica: Testi scelti introdotti da Massimo A. Bonfantini, Letizia Grassi, Roberto Grazia*. Torino: Einaudi.

Pigliapoco, Stefano. 2016. *Progetto archivio digitale: metodologia, sistemi, professionalità*. Torre del Lago (Lucca): Civita Editoriale.

Salarelli, Alberto. 2012. *Introduzione alla scienza dell'informazione*. Milano: Editrice Bibliografica.

Sebastiani, Mario. 2005. Identificatori persistenti per oggetti digitali. In *DigItalia*, 0: 62-82.

Sebastiani, Mario. 2008. Il “documento digitale”. Analisi di un concetto in evoluzione. In *DigItalia*, III/1: 9-31.

Serrai, Alfredo. 1973. *Biblioteconomia come scienza: Introduzione ai problemi e alla metodologia*. Milano: Editrice Bibliografica.

Sokolov, Viktor. 2009. “The epistemology of documents”. in *Automatic Documentation and Mathematical Linguistics*, XXXXIII/2: 57-68.

Valacchi, Federico. 2020. *Gli archivi tra storia, uso e futuro: la rivoluzione tecnologica e le biblioteche*. Milano: Editrice Bibliografica.



# The Formation of the Idea of the Library as an Institution in 18<sup>th</sup>-Century Europe. A Qualitative and Quantitative Approach\*

Carlo Bianchini<sup>(a)</sup>, Lorenzo Mancini<sup>(b)</sup>, Fiammetta Sabba<sup>(c)</sup>

a) University of Pavia, <https://orcid.org/0000-0002-6635-6371>

b) National Research Council of Italy, Institute for the European Intellectual Lexicon  
and the History of Ideas (CNR-ILIESI), <https://orcid.org/0000-0001-6344-7000>

c) University of Bologna, <https://orcid.org/0000-0002-9786-6825>

**Contact:** Carlo Bianchini, [carlo.bianchini@unipv.it](mailto:carlo.bianchini@unipv.it); Lorenzo Mancini, [lorenzo.mancini@cnr.it](mailto:lorenzo.mancini@cnr.it);  
Fiammetta Sabba, [fiammetta.sabba@unibo.it](mailto:fiammetta.sabba@unibo.it)

**Received:** 27 December 2023; **Accepted:** 19 February 2024; **First Published:** 15 May 2024

## ABSTRACT

The paper illustrates the LIBMOVIT project – Libraries on the Move: Scholars, Books, Ideas Traveling in Italy in the 18th Century – whose main research focus is the European Eighteenth century socio-cultural framework in which the library as an institution acquired an historical, social, public and dynamic dimension. This context will be analysed through a study of the Eighteenth century sources connected to the learned journey experience of the Grand Tour, in particular those contained in the Angiolo Tursi collection – one of the largest travel literature collections in Italy – held at the Marciana national library in Venice. The paper presents the planned approach of the research: first, a classification and an organization of a corpus of relevant documents for the knowledge of travel literature in connection to the libraries world will be created; in particular, the sources will be identified, further bibliographical information will be added, and new sources will be integrated to the corpus and selected documents will be digitized. After that, the research will proceed through a double analysis – traditional and computational – of the texts collected in the corpus is to be developed. First, all the library and bibliographical aspects described by travellers will be studied according to the traditional approach in humanities research to collect important information about the history of libraries (location, decoration, catalogues, opening hours, access, collections, cited books and documents), the travellers and their companions (professions, nationality, reason to travel), the people met (scholars, librarians, superintendents) and the subjects and ideas discussed during the visits in the libraries. Second, the texts will be computationally analysed through several Natural Language Processing (NLP) techniques, starting from the automatic text recognition until arriving to more complex lexical and terminological analysis and Named Entity Recognition (NER). This work is meant to support the previously described qualitative study and will also allow to produce Linked open data about the domain entities (e.g. libraries, people, books) in view of their publication in the semantic web in order to ease and promote their exploration, visualisation and reuse.

## KEYWORDS

Library history; Travel literature; Natural Language Processing; Wikidata; Qualitative research; Quantitative research.

\* The authors cooperated in the redaction and revision of the article in full. Nevertheless, some authors mainly authored some parts of the article: Lorenzo Mancini: section “Research Detailed Tasks Program”; and Fiammetta Sabba: sections “Background of research” and “Research context, sources, and methodologies”. Moreover, Lucia Sardo, who is part of the research team, authored the Gantt Diagram (Figure 2) at page 97.

© 2024, 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

This paper presents a research project that aims to historically reconstruct the formation of the idea of the library as an institution in the 18th century through oediporic sources and by applying qualitative and quantitative methodology. On the one hand, the purpose of this publication is to highlight the key points that enabled the project to receive a national research grant (PRIN 2022) and, on the other hand, to foster the project to a wider audience, with the goal to identify possible additional partners and stakeholders.

The project title is “Libraries on the move: scholars, books, ideas in 18th century Italy” (LIBMOVIT) and involves three research units: the University of Bologna (host institution of the Principal Investigator), the University of Pavia and the Institute for the European Intellectual Lexicon and the History of Ideas of the National Research Council of Italy (CNR-ILIESI).

The LIBMOVIT<sup>1</sup> project focuses primarily on the socio-cultural background of eighteenth-century Europe, examining the historical, social, public, and dynamic dimensions of libraries as institutions during this period. The investigation delves into the context through an exploration of eighteenth-century sources related to the scholarly journeys of the *Grand Tour*, with a particular emphasis on the Angiolo Tursi collection. This collection, housed at the Biblioteca Nazionale Marciana in Venice, stands as one of Italy’s most extensive collections of travel literature.

In the initial phase of the project, a corpus of pertinent documents, illuminating the intersection of travel literature and libraries, will be assembled. The documents will undergo classification and organization based on various criteria, such as publication type, language, origin, literary genre, and accessibility. This stage involves source identification, additional bibliographical details, corpus expansion with new sources, and the digitization of selected documents.

The second phase incorporates a dual analysis – traditional and computational – of the collected texts. Traditional humanities research methods will be applied to examine library and bibliographical aspects described by travellers. This includes an exploration of elements crucial for reconstructing library use and assessing the impact of library visits on intellectual development, the European scholarly communication network, and the formation of bibliographic collections and library institutions. The computational analysis will leverage Natural Language Processing (NLP) techniques, ranging from automatic text recognition to advanced lexical and terminological analysis, as well as Named Entity Recognition (NER). This computational aspect aims to support the overall research and facilitate the indexing of domain entities (e.g., libraries, individuals, books) for publication in the semantic web. The goal is to enhance exploration, visualization, and reuse of the data.

## Background of research

The Grand Tour reached its zenith in the 18th century, with Italy standing out as one of its preferred destinations. Extensive scientific literature exists on this historical and cultural phenome-

---

<sup>1</sup> Progetto finanziato dall’Unione Europea – NextGenerationEU a valere sul Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Istruzione e ricerca – Componente 2 Dalla ricerca all’impresa – Investimento 1.1, Avviso Prin 2022 indetto con DD N. 104 del 2/2/2022, Progetto dal titolo LIBMOVIT – Libraries on the move: scholars, books, ideas traveling in Italy in the 18th century, codice proposta 2022CP88KY.

non, with numerous studies delving into specific interests and research areas related to the Grand Tour. However, studies examining the relationship between this extensive travel experience and the world of libraries, a key participant in this era, are sporadic.

While travel documentation has been acknowledged as a potential source for library history, particularly by authors such as Ottino-Fumagalli, Predeek, Buzas, and Serrai (Ottino and Fumagalli 1889; Predeek 1928; Buzas 1976; Serrai 1999) – as well as used for limited interventions by Lodovica Braida (Braida 2002; 2018), Loretta De Franceschi (De Franceschi 2013), Graziano Ruffini (Ruffini 2012), Gianfranco Tortorelli (Tortorelli 2012) e Vincenzo Trombetta (Trombetta 1994; 2014; Andrés 1997)- a comprehensive historical perspective on libraries has only recently emerged. In addition to numerous essays on the topic which delve into specific testimonies or highlight relevant aspects, (Sabba 2018a; 2019; 2021a; 2021b; 2021c; 2021d; 2023; Conterno and Sabba 2022) a notable recent study by Sabba (Sabba 2018b) presented a broader historiographical approach with a focus on library history. Nonetheless, this study analysed a limited sample of sources, including only printed sources, travel reports, and letters from foreign travellers.

Against the backdrop of early modern European culture, this project aims to scrutinize the role of libraries more extensively and precisely than previous studies. It seeks to underscore the centrality of libraries in scholarly communication, pivotal in shaping cultural models revered by aristocratic and bourgeois elites throughout the early modern era. The project also endeavours to explore and document the 18th-century European origins of the contemporary concept of the library as a “conversation” (Lankes 2011) and a space for the elaboration and production – beyond preservation – of knowledge. In pursuit of these goals, travel memoirs serve as a privileged source due to their spontaneous, culturally diverse, and inclusive qualities. These “original” sources remain underutilized in the context of library history. Their disciplinary significance is evident as the history of 18th-century Italian libraries has predominantly focused on individual collections and institutions, lacking a comprehensive approach. Even if previous research initiative already approached this field using Digital Humanities methodologies (De Caprio and Meschini 2006; Meschini 2011; Anderson et al. 2017; Carey, Gelléri, and Ingram 2020), this project aims to fill the absence of a specialized textual database, offering comprehensive information on works, authors/travellers, companions, places, libraries, collections, and individuals encountered. Furthermore, there is a scarcity of structured, open, free, and reusable publication of data in the field of bibliography and library history, hindering innovative approaches to access, exploration, and visualization.

## Research context, sources, and methodologies

The Grand Tour phenomenon wielded a significant influence on shaping the European Republic of Letters, marking travel as an intellectual and formative experience. In the 18th century, travel evolved into a multifaceted encounter, impacting both individual and societal perspectives. It became a widespread practice, recognized across European society for its instrumental role in shaping ruling elites, scholars, men of science, and the intellectual class. The Republic of Letters, well-established by the late 17th century, owed its consolidation to various communication channels like literary and scientific journals and scholarly correspondence and to the opportunities for personal interaction and exchange facilitated by travel. This renewed context witnessed movements of people, books, and ideas.

Italian libraries emerged as preferred destinations for foreign travellers, due to their architectural and artistic significance and as hubs of literary and scientific information. Librarians played a crucial role, disseminating knowledge through correspondence and personal interactions with library visitors. Despite the pivotal role of libraries in shaping the cultural identity of modern Europe, there is currently a scarcity of studies that analyse and underscore their significance as centres of erudite communication.

The research will delve into the cultural model associated with the 18th-century birth of the public library – a dynamic institution serving societal advancement, documentary selection and conservation, and the generation, production, and dissemination of knowledge.

Uniquely, the project adopts a library history perspective and relies on odeporic sources, which will be documented and made publicly accessible through an online platform.

The primary resource for building the corpus for study and analysis is the Angiolo Tursi (1885-1977) collection, donated to the Biblioteca Nazionale Marciana in Venice in 1968 (Tursi 1956; Vianello 1968). Tursi's collection, collected over nearly 40 years, stands as the richest and most systematic archive on travel in Italy, comprising about 26,000 documents. Noteworthy features include an 80,000-card catalogue alphabetically arranged and enriched by a systematic classification of travellers and resource types. A portion of the collection is digitized and available online on the "Internet Culturale" portal ('Home – Internet Culturale' 2023; Ciancio 2010).

The bibliographic and bibliological richness, coupled with the meticulous arrangement, positions the Tursi collection as an invaluable resource for scholars. Its unique qualities also make it a foundational element for an essential annotated bibliography on travellers in Italy, deserving a dedicated bibliographic study. It is essential to recognize that collector-assembled collections are the most comprehensive and carefully curated, representing a paradigmatic perspective from a bibliographic standpoint.

Analysing the Tursi collection will expand the documentary base for research, broadening the scope of sources to include travel guides and apodemic texts, in addition to the more conventional reports, diaries, and letters. Another expansion will concern both manuscripts and printed documents, and the study will not only consider writings by foreign travellers but also those penned by Italian authors, providing a comprehensive exploration of the subject matter.

Furthermore, the LIBMOVIT project seeks to evaluate a new methodology for source analysis, considering the increasing prevalence of rendering cultural relations, encompassing people and bibliographic, archival, or museum objects. This is pursued through the publication of data in the form of Linked Open Data (LOD). This approach introduces a transformative paradigm for utilizing, exploring, and visualizing data, proving valuable for both research and content dissemination (Barbera 2013; Boccone et al. 2017; Klic et al. 2017; Molinié 2019; 2020; Zavattoni 2020; Bonora and Pompilio 2021). While some limited experiences exist in developing data analysis and visualization for articles in Italian librarianship journals (Bianchini 2021; Bianchini, Marchitelli, and Moi 2023), expertise in publishing LOD in the history of libraries is still to be exploited.

To align the project with LOD practices, the sources corpus will undergo quantitative and computational analysis. Beginning with the 18th-century sources selected in the initial research stage, textual data will be extracted using techniques and tools developed in the Natural Language Processing (NLP) context (Piotrowski 2012). This operation will furnish information to structure subsequent tasks, including the reconstruction of a specific travel experience terminology. With-

in this framework, the identification of terms linked to key concepts in the history of ideas (e.g., otherness, knowledge, freedom) or associated with philosophical literature using travel metaphors will be undertaken (Van den Abbeele 1992; Menzio 1994). The computational analysis will also facilitate data processing related to individuals, themes, itineraries, places, institutions, collections, and texts, along with their interrelationships.

Ultimately, all the aforementioned data will be published on the semantic web to capitalize on its capabilities, including identification, enrichment, integration, data quality, software independence, accessibility, exploration, sharing and interoperability, decentralization, reusability, visualization, and analysis. Technologically, the research will leverage tools provided by the Wikimedia community, offering additional advantages beyond those associated with semantic web technologies. The publication of data on Wikimedia platforms – or Wikibase instances – brings several additional benefits, complementing the advantages already highlighted for semantic web technologies:

- Cost-Free Tools: Wikimedia tools – and Wikibase websites – are freely accessible;
- Up-to-date Visualization Tools: Access to constantly updated free tools for visualization and disseminating data in the form of timelines, graphs, and interactive maps;
- Interoperability Guarantee: Ensured interoperability of data with those from other institutions;
- Modularity: Easy integration and enrichment of data in future projects, thanks to a modular approach;
- Collaborative Environment: The platform fosters a free and collaborative environment;
- Long-Term Data Guarantee: Assurance of data preservation and compatibility with various formats over the long term;
- Third-Party Reusability: Facilitated reuse of data by third-party institutions;
- Effortless Dissemination: Simple dissemination, communication, and utilization of scientific and cultural content by an extensive audience, including the educational sector, as Wikimedia platforms stand as the most significant and visited non-profit website globally.

Furthermore, adhering to an open science perspective (Giglia 2019; Angiolini 2021), data and information related to the sources under study in this research will be published in the form of datasets. They will be made accessible to the scientific community by uploading them to European research infrastructures, such as *Zenodo*.<sup>2</sup>

## Content Objectives of the research and Related Tasks

The primary content objective is segmented into fundamental objectives, along with subsequent tasks to be pursued until the publication of results. These are detailed as follows:

1. In-depth Study of Odeporic Memories:
  - a. Conduct an exhaustive examination of research on odeporic memories utilizing the Angiolo Tursi collection;
  - b. Utilize the extensive collection to precisely identify bibliographic and library-related aspects significant to both foreign and Italian travellers;

---

<sup>2</sup> <https://zenodo.org/>.



- c. Investigate the influence of visits to Italian libraries on the construction of collections and library organization;
2. Extension of Research to Additional Bibliographic Typologies:
  - a. Extend research to underexplored travel literature types such as guides and apodemic texts;
  - b. Enhance understanding of the communicated and conveyed external library idea in comparison with perceived and real libraries from personal travel memories;
3. Creation of an Open Access Portal:
  - a. Establish a portal for international collaboration and dissemination of research results;
  - b. Provide a platform for scholars interested in travel history and odeporic documentation;
4. Creation of a Bibliographic and Documentary Repertoire:
  - a. Identify odeporic sources in the Tursi collection from the 18th century;
  - b. Develop a bibliographic tool containing digitized publications and cataloging data;
5. Textual Database Creation and Computational Analysis:
  - a. Enrich bibliographic data with the full text of selected works;
  - b. Establish a textual database for computational analyses and semantic enrichment;
6. Publication of Linked Open Data (LOD) on semantic web:
  - a. Publish data in LOD format on Wikimedia platforms for visibility and accessibility;
  - b. Provide modular data for future research integration in the field.

Primary content objectives imply the following descending objectives and Related Tasks:

7. Study on Librarians in Travel Literature:
  - a. Investigate the presence of librarians in travel literature as travellers or individuals met during journeys;
  - b. Assess the impact of travels and library encounters on the circulation of bibliographic and librarianship models;
8. Creation of Multilingual Textual Corpus:
  - a. Establish a multilingual textual corpus of travel literature;
  - b. Identify terminology and conduct terminological searches on the intersection of travel and philosophical research;
9. Double Census from Odeporic Texts:
  - a. Conduct a register of unknown or forgotten libraries;
  - b. Compile a register of librarians not recorded in specialized biographical repertoires.

These tasks aim to contribute significantly to the history of libraries and travel literature, providing valuable resources for future research.

## Expected primary general and specific outcomes

The project aims to investigate the evolving role of libraries in the 18th century, transitioning from scholarly and traveler-centric centers to the public-oriented and service-driven contemporary library. Tangible results for the scientific community and the public include: the creation of a census of 18th-century Travel Records (starting from the Tursi collection); the creation of a bibliographic index by the identification and cataloging of odeporic sources relevant to the research, including digitized full texts; a computational analysis of texts to create a marked full-text multilingual database and the harvesting of data on people, themes, etc., with LOD publication on Wikidata-like platforms; the creation of a web portal for data access, project presentation, ongoing initiatives, and open-access publication of scientific results. These outputs represent an innovative historical approach, leveraging diverse source types and adhering to open science principles, ensuring structured and reusable information accessibility. The project also incorporates Natural Language Processing (NLP) techniques for text analysis, offering insights into linguistic, terminological, and quantitative aspects of odeporic literature.

Specific outcomes of the research include educational events, such as seminars and hackathons for students and professionals to disseminate research results and describe methodologies, a final conference for presenting research outcomes and inviting contributions from institutions and scholars from various disciplines, and open access publications, available through the project website, institutional repositories, and research infrastructure. So, outcomes ensure a comprehensive understanding of libraries in the 18th century, and contribute to technological innovation and the potential integration of qualitative and computational research methods. The use of the semantic web warrants global accessibility, facilitating data reuse, collaboration, and the project's visibility.

## Research Detailed Tasks Program

The research program was defined in detail, to identify actions, timetables, resources, and interactions between them. First, a project management and monitoring process, to organize operational units and provide initial training for new personnel, conduct a kick-off meeting to introduce the project to the community<sup>3</sup>, create a project website<sup>4</sup> and activate communication channels and constantly monitor activities, update social channels and the website.

The second activity – involving all research units – provides for the creation of the census of sources and their classification: sources from the Tursi collection are selected by means of classification and bibliographic records, are integrated by other catalogues of the Biblioteca Nazionale Marciana, Venice, and of other specialized collections. A database based on identified sources, classified according to relevant features, is set up.

A third step includes the use of available metadata to create a repertoire, checking and enriching existing records, and adding sources not yet catalogued in the National Librarian Service (SBN) catalogue, to publish a specific database on the project website.

---

<sup>3</sup> The Kick-off meeting was held in Ravenna, on 7 December, 2023.

<sup>4</sup> <https://libmov.it/>.

The next activity is the identification and the content analysis of relevant sources, to identify textual and editorial features of classified sources, and recognize passages and information of library and bibliographic interest. Following well established workflow in the field of textual analysis (Piotrowski 2012; Del Grosso et al. 2018), after digitalization, when necessary, and automatic text recognition, identified sources need to undergo to a basic XML/TEI markup. The linguistic analysis will then provide a lemmatized corpus, necessary to perform better Named Entities Recognition (NER), both automatic and manual, and also ease the construction of a terminology resource (Bellandi et al. 2017) related to travel experience with its philosophical connections.

The identification of entities allows to model an ontology of the most interesting entities (e.g. cities, libraries, catalogues, opening hours, etc.), to define necessary metadata for LOD publication on Wikidata-like platforms, to process and validate data, and to link them to existing projects taking advantage of Wikidata as a hub of the semantic web (Allison-Cassin and Scott 2018; Association of Research Libraries 2019; Linked Data for Production 2020).

All the activities need to be properly communicated and publicized, through a well-defined and modular website, the organization of an operational workshop in the first year and of a final conference with an international call for papers, the open access publication of a special monograph and of the proceedings of the conference, and the organization of an online bibliographic exhibition, teaching activities, and third mission activities.

Actions, timetables, resources, and interactions between them required by the research are illustrated in the Flowchart (figure 1) and in the Gantt Diagram (figure 2) below.

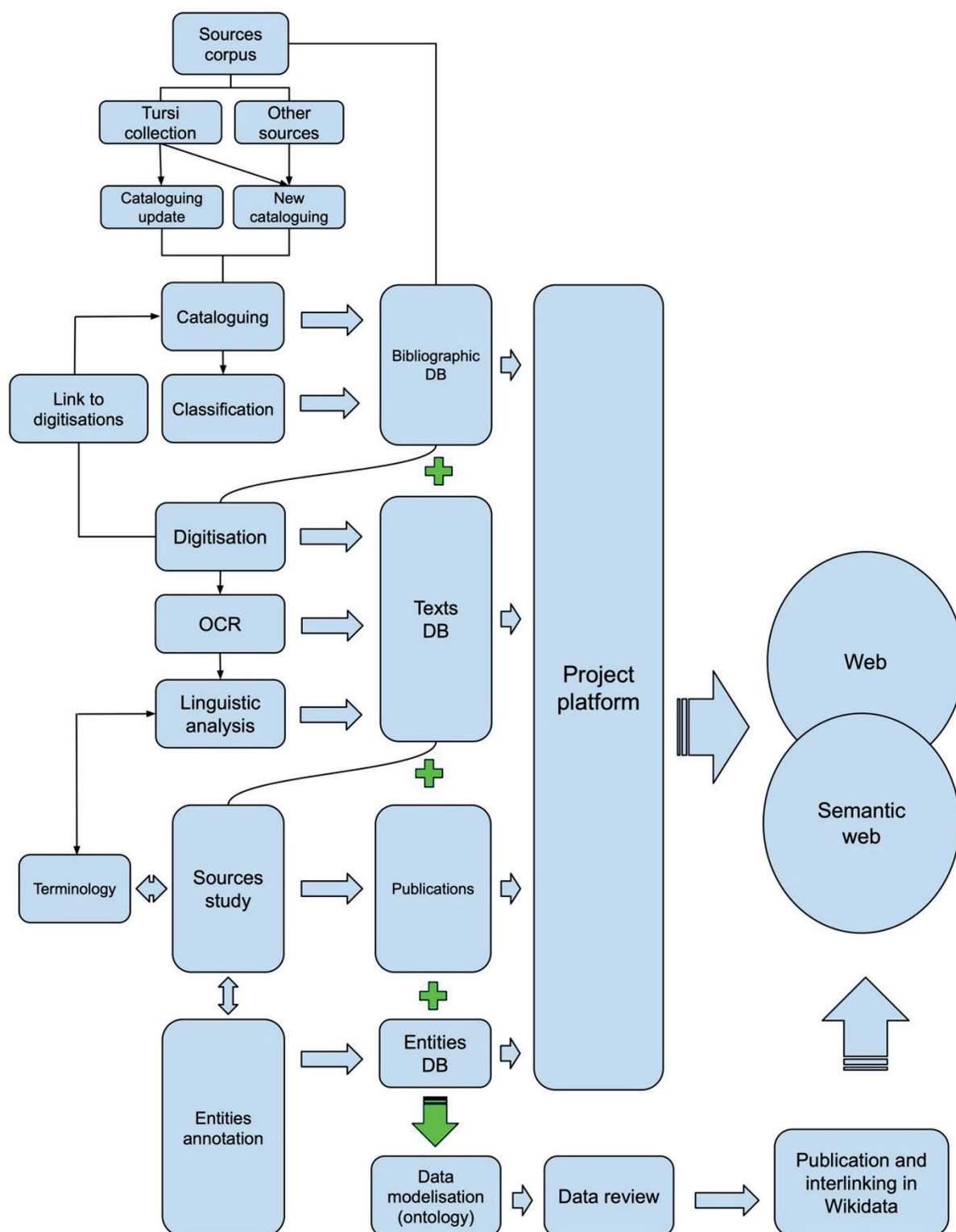


Figure 1. Flowchart of activities of the research project.

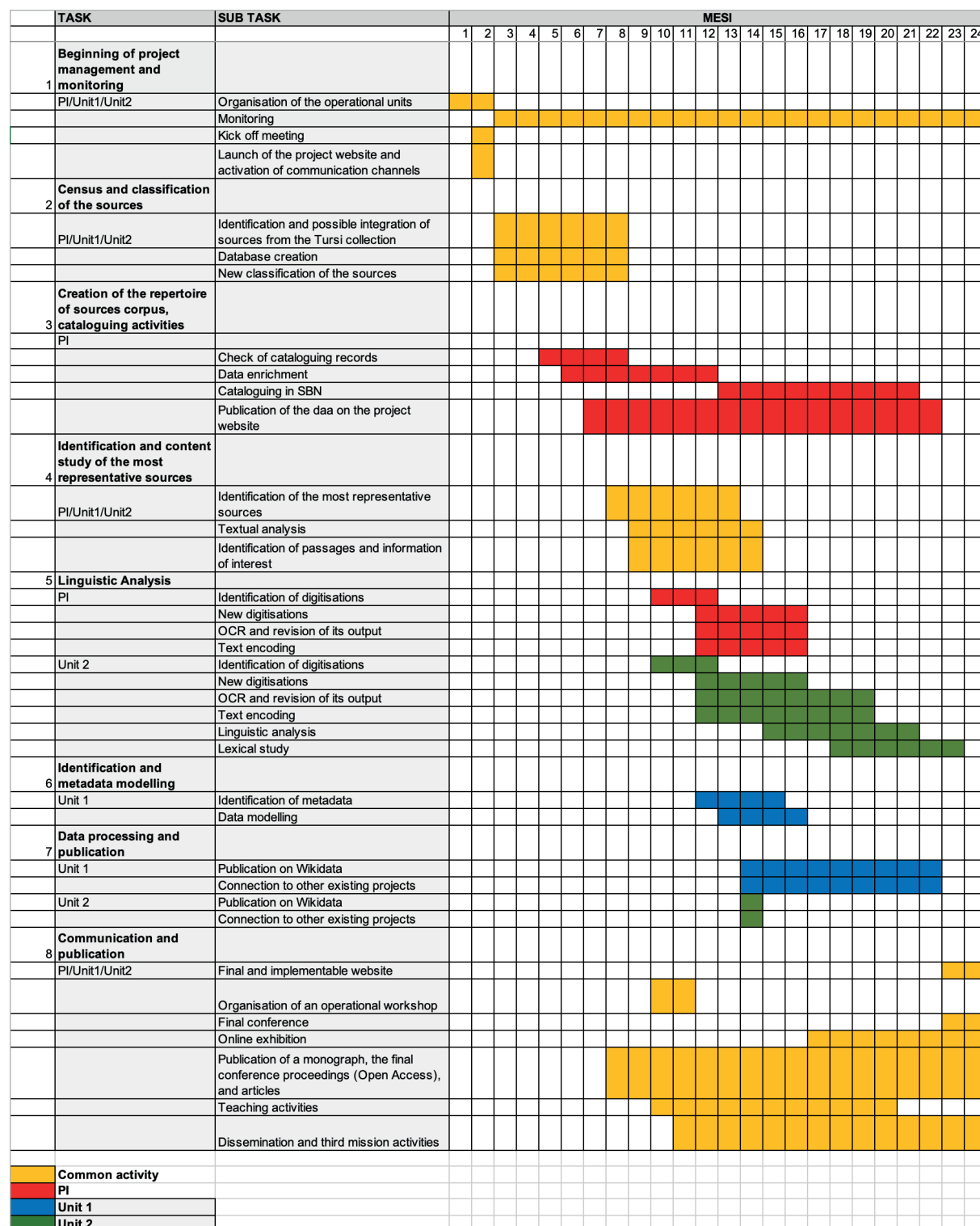


Figure 2. Gantt diagram with the timetable of the research project.



## Expected impact of the research project

The impact of the project primarily lies in making a substantial amount of data and information related to the history of libraries and travel literature available to the scientific community. This is particularly significant for understanding the 18th-century origins of the modern public library as a space for exchange and conversation, rather than mere preservation.

Additionally, the project will contribute significantly by creating a free and online bibliographic repertoire of odeporic sources. The availability of full-text selections from these sources will serve to enhance knowledge in this domain. While these sources are primarily studied for historical-librarianship purposes, their interdisciplinary nature, reflecting diverse cultural experiences through travel, adds broader academic value.

Moreover, the recognition and systematization of sources from the Angiolo Tursi collection will not only enrich the odeporic collections of the Biblioteca Nazionale Marciana, Venice, but also exemplify how the scientific community can collaborate with cultural institutions to enhance and promote existing collections. It will serve as a model for the collaborative development of similar source corpora across Italian and European institutions.

The significance of these collections extends beyond individual libraries. Notable Grand Tour collections in various Italian cities, such as those in Florence, Rome, Naples, Turin, and Milan, represent valuable tools for promoting the internationalization of research. These collections offer a comprehensive view of how libraries were perceived by travellers over a broader chronological spectrum, enhancing the study of library perceptions.

Furthermore, odeporic literature, by its nature, is international and multilingual, aligning with the spirit of the Grand Tour. The project's results will facilitate international scientific relations by involving scholars from other European countries and experts in computational text analysis in languages other than Italian.

Within the Italian research context, the project sets an example of desired integration, distributed across the territory, between the university's scientific community and public research bodies. Such integration is relatively uncommon, especially in the field of humanities research.

Finally, the research results will be disseminated through the promotion of an interdisciplinary study approach, integrating humanities with technological applications, including teaching; the dissemination of basic research data from an open science perspective, utilizing platforms like Zenodo and other European research infrastructures; the dissemination on a Wikidata-like platform of structured, open, searchable, and reusable data under a free license. This includes data visualizations tools to enhance accessibility and impact on a non-specialist audience at national and international levels.

Embracing an open-access approach aligns with Open Science principles, aiming to involve society in the research community and share the resulting benefits in economic, social, and cultural terms.

## First steps of the project

As stated at the beginning, the objective of these pages was to present to a specialized audience the background, the goals, and the expected results of a project worthy to be funded in a competitive

national research programme. The LIBMOVIT project officially started in October 2023 and in December the Cultural Heritage Department of the University of Bologna hosted the kick-off meeting. During this event<sup>5</sup> the team laid out the project and invited Lodovica Braida and Giorgio Montecchi to attend and comment its potentialities for the fields of Book, Library and Cultural History; in addition, Orsola Braides, special collections librarian at the Biblioteca Nazionale Marciana, was invited to present the Angiolo Tursi collection, which constitutes the main LIBMOVIT bibliographic source. The kick-off meeting was also the opportunity to promote the project website (available at: <https://www.libmov.it/>, where more information about the project, a page with a specialized bibliography, and a “News” section are available, to stay updated about the research progress and events.

---

<sup>5</sup> <https://www.libmov.it/?p=1>.

## References

- Allison-Cassin, Stacy, and Dan Scott. 2018. 'Wikidata: A Platform for Your Library's Linked Open Data.' *Code4Lib Journal*, 4 May 2018. <https://journal.code4lib.org/articles/13424>.
- Anderson, Carrie, Giovanna Ceserani, Christopher Donaldson, Ian N. Gregory, Melanie Hall, Adam T. Rosenbaum, and Joanna E. Taylor. 2017. 'Digital Humanities and Tourism History.' *Journal of Tourism History* 9 (2–3): 246–69. <https://doi.org/10.1080/1755182X.2017.1419455>.
- Andrés, Juan. 1997. *Gl'incanti di Partenope*. Edited by Vincenzo Trombetta. Napoli: A. Guida.
- Angiolini, Andrea. 2021. 'Open to Whom : The Open Science in the Quest for Readers.' *JLIS.it*: 12 (3). <https://doi.org/10/gnbnx5>.
- Association of Research Libraries. 2019. *ARL White Paper on Wikidata. Opportunities and Recommendations*. <https://www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf>.
- Barbera, Michele. 2013. 'Linked (Open) Data at Web Scale: Research, Social and Engineering Challenges in the Digital Humanities.' *JLIS.it* 4 (1): 91–104. <https://doi.org/10.4403/jlis.it-6333>.
- Bellandi, Andrea, Emiliano Giovannetti, Silvia Piccini, and Anja Weingart. 2017. 'Developing LexO: A Collaborative Editor of Multilingual Lexica and Terminology-Ontological Resources in the Humanities.' In *Proceedings of Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS 2017)*, edited by Francesca Frontini, Larisa Grčić Simeunović, Špela Vintar, Anas Fahad Khan, and Artemis Parvisi. Montpellier, France: Association for Computational Linguistics. <https://aclanthology.org/W17-7010>.
- Bianchini, Carlo. 2021. 'Wikidata for JLIS.it : a new step forward mapping Italian library and information science journals.' *JLIS.it* 12 (1): 29–38. <https://doi.org/10/gndpxj>.
- Bianchini, Carlo, Andrea Marchitelli, and Alessandra Moi. 2023. 'Metodi e strumenti di un progetto di valorizzazione delle riviste italiane di biblioteconomia in Wikidata.' *AIB studi* 63 (2): 313–35. <https://doi.org/10.2426/aibstudi-13893>.
- Boccone, Alessandra, Claudio Forziati, Tania Maio, and Remo Rivelli. 2017. 'Valorizzazione dei fondi privati in una biblioteca accademica: divulgazione, comunicazione, ricerca.' *Bibliothecae.it* 6 (2): 255–84. <https://doi.org/10.6092/issn.2283-9364/7703>.
- Bonora, Paolo, and Angelo Pompilio. 2021. 'Corago in LOD. The Debut of an Opera Repository into the Linked Data Arena.' *JLIS.it* 12 (2): 54–72. <https://doi.org/10.4403/jlis.it-12699>.
- Braida, Lodovica. 2002. 'Circolazione Del Libro e Pratiche Di Lettura Nel Settecento.' In *Biblioteche Nobiliari e Circolazione Del Libro Tra Settecento e Ottocento*, edited by Gianfranco Tortorelli, 11–37. Bologna: Pendragon.
- Braida, Lodovica. 2018. 'Il Ricorso All'anonimato Nel Settecento.' *La Bibliofilia* 120 (2): 259–78.
- Buzas, Ladislaus. 1976. *Deutsche Bibliotheksgeschichte Der Neuzeit (1500-1800)*. Wiesbaden: Reichert.
- Carey, Daniel, Gábor Gelléri, and Anders Ingram. 2020. 'The Art of Travel (1500-1850) Database.' *Viatica*, 7. <https://doi.org/10.52497/viatica1370>.

Ciancio, Laura. 2010. “Internet Culturale. Cataloghi e Collezioni digitali delle biblioteche italiane” La nuova versione del portale.’ *DigItalia* 5 (2): 123–34. <https://digitalia.cultura.gov.it/article/view/244/155>.

Conterno, Chiara, and Fiammetta Sabba, eds. 2022. *Il Patrimonio Culturale Della Biblioteca Universitaria Di Bologna e Della Città Allo Specchio Dei Viaggiatori Europei. Esplorazioni Tra La Prima Modernità e l'era Contemporanea. / Das Kulturelle Erbe Der Universitätsbibliothek von Bologna Sowie Der Ganzen Stadt Im Spiegel Europäischer Reisender. Streifzüge Zwischen Früher Neuzeit Und Moderne*. Bologna: BUP.

De Caprio, Vincenzo, and Federico Meschini. 2006. ‘Scritture Di Viaggio e Informatica Umanistica: L'esperienza Di AVIREL.’ In *Viaggiatori Dell'Adriatico. Percorsi Di Viaggio e Scrittura*, 187–202. Bari: Palomar.

De Franceschi, Loretta. 2013. ‘Viaggiando per Biblioteche Nel Settecento: Un Percorso Guidato.’ *Bibliologia* 8: 63–84.

Del Grosso, Angelo Mario, Andrea Bellandi, Emiliano Giovannetti, Simone Marchi, and Ouafae Nahli. 2018. ‘Scanning Is Just the Beginning: Exploiting Text and Language Technologies to Enhance the Value of Historical Manuscripts.’ In *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, 214–19. Marrakech: IEEE. <https://doi.org/10.1109/CIST.2018.8596373>.

Giglia, Elena. 2019. ‘OPERAS: Bringing the Long Tail of Social Sciences and Humanities into Open Science.’ *JLIS.it* 10 (1): 140–56. <https://doi.org/10.4403/jlis.it-12523>.

‘Home – Internet Culturale.’ 2023. <https://www.internetculturale.it/>.

Klic, Lukas, Matt Miller, Jonathan K. Nelson, Cristina Pattuelli, and Alessandra Provo. 2017. ‘The Drawings of the Florentine Painters: From Print Catalog to Linked Open Data.’ *Code4Lib Journal* 38. <https://journal.code4lib.org/articles/12902>.

Lankes, R. David. 2011. *The Atlas of New Librarianship*. Cambridge, Mass.: The MIT Press.

Linked Data for Production. 2020. ‘Wikidata as a hub for identifiers.’ Google Docs. 11 June 2020. <https://t.ly/677QP>.

Menzio, Pino. 1994. *Il Viaggio Dei Filosofi: La Metafora Del Viaggio Nella Letteratura Filosofica Moderna*. Geneve ; Moncalieri: Slatkine ; Centro interuniversitario di ricerche sul viaggio in Italia.

Meschini, Federico. 2011. ‘Avirel. Un Archivio Digitale per l'Odeporica.’ In *Libri Di Viaggio, Libri in Viaggio. Studi in Onore Di Vincenzo de Caprio*, edited by Stefano Pifferi and Cinzia Capitoni, 292–303. Viterbo: Sette città.

Molinié, Christelle. 2019. ‘Objectif LOD (Linked Open Data): comment et pourquoi ouvrir et lier les données des musées? L'expérience du musée Saint-Raymond, musée des Antiques de Toulouse.’ In *Capitaliser les ressources documentaires en Musée*, 139–56. Dijon: Ocim. <https://fr.calameo.com/read/005777060ac36dc7396bc>.

Molinié, Christelle. 2020. ‘Les Vase Communicants: Collaboration Entre Musées et Projets Wikimédia Pour Le Partage Du Patrimoine Culturel.’ Wikimedia France. <https://www.wikimedia.fr/collaboration-musees-wikimedia-partage-patrimoine-culturel/>.

Ottino, Giuseppe, and Giuseppe Fumagalli. 1889. *Bibliotheca Bibliographica Italica: Catalogo Degli Scritti Di Bibliologia, Bibliografia e Biblioteconomia Pubblicati in Italia e Di Quelli Risguardanti l'Italia Pubblicati All'estero*. 2 voll. Roma; Torino: Loreto Pasqualucci; Clausen. <http://books.google.it/books?id=Dd95ZnH-6AYC>.

Piotrowski, Michael. 2012. *Natural Language Processing for Historical Texts*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-031-02146-6>.

Predeek, Albert. 1928. 'Bibliotheksbesuche Eines Gelehrten Reisenden Im Anfänge Des 18. Jahrhunderts.' *Zentralblatt Für Bibliothekswesen* 45: 221–65, 342–54, 393–407.

Ruffini, Graziano. 2012. *La Chasse Aux Livres. Bibliografia e Collezionismo Nel Viaggio in Italia Di Etienne-Charles de Loménie de Brienne e François-Xavier Laire (1789-1790)*. Firenze: Firenze University Press.

Sabba, Fiammetta. 2018a. 'Le Biblioteche Negli "Itinera Erudita et Bibliothecaria": Riflessioni Su Grand Tour e Turismo.' In *Le Biblioteche Anche Come Musei: Dal Rinascimento Ad Oggi, Atti Del 'Colloquio Internazionale/Colloque International Les Bibliothèques Aussi Comme Musées' (Roma, Biblioteca Nazionale Centrale, 16-17 Novembre 2016)*, edited by Andrea De Pasquale and Silvana De Capua, 105–23. Roma: Biblioteca nazionale centrale di Roma.

Sabba, Fiammetta. 2018b. *Viaggi Tra i Libri: Le Biblioteche Italiane Nella Letteratura Del Grand Tour*. Pisa: Fabrizio Serra editore.

Sabba, Fiammetta, ed. 2019. *Patrimonio Culturale Condiviso, Viaggiatori Prima e Dopo Il Grand Tour. Atti Del Convegno Internazionale (Ravenna, Dipartimento Di Beni Culturali, 21-23 Novembre 2018)*. Napoli: Viaggiatori.

Sabba, Fiammetta. 2021a. 'I Costi Del Consumo Culturale a Napoli a Fine Settecento. Informazioni Da Un Documento Allegato al "Diario Di Viaggio Nel Sud Italia" Di Angelo Maria Bandini.' In *Scaffali Come Segmenti Di Storia. Studi in Onore Di Vincenzo Trombetta*, edited by Rosa Parlavecchia and Paola Zito, 129–34. Roma: Quasar.

Sabba, Fiammetta. 2021b. 'Il Grand Tour: Da Officina Delle Idee e Dell'apprendimento a Motore Urbano, Sociale, Economico e Culturale Attraverso Libri e Documenti.' In *Viaggio Tra i Libri Di Viaggio*, edited by Paolo Tiezzi Maestri, 47–51. Torrita di Siena: Biblioteca di Villa Classica.

Sabba, Fiammetta. 2021c. 'Il Viaggiatore Erudito Montesquieu in Italia: Saggio Del Suo Incontro Con Uomini, Libri e Biblioteche.' *Biblioteca Di Via Del Senato* 13: 77–82.

Sabba, Fiammetta. 2021d. 'La Biblioteca Nel Grand Tour: Luogo Di Incontro Fra Uomini e Libri.' In *La Biblioteca: Crocevia e Connessione Di Mondi, (Bagni Di Lucca, Settembre 2018)*, edited by Laura Giovannelli and Roberta Ferrari, 57–78. Pisa: ETS.

Sabba, Fiammetta. 2023. 'Ambrogio Traversari, Camaldolese Romagnolo: Una Precoce Testimoniaza Odeporica.' In «*Ex Libris... Ne Pereant*». *Cultura Libraria e Archivistica Tra Umanesimo e Rinascimento. Miscellanea Di Studi in Onore Dell'avv. Paolo Tiezzi Mazzoni Della Stella Maestri*, edited by Manlio Sodi and Mario Ascheri, 317–25. Firenze: Leo S. Olschki.



Serrai, Alfredo. 1999. 'Itinera Erudita. Itinera Bibliothecaria.' In *Storia Della Bibliografia*, edited by Alfredo Serrai, IX: Manualistica, Didattica, e Riforme nel sec. XVIII:534–52. Roma: Bulzoni.

Tortorelli, Gianfranco, ed. 2012. *Viaggiare con i libri: saggi su editoria e viaggi nell'Ottocento*. Le sfere. Bologna: Pendragon.

Trombetta, Vincenzo. 1994. 'Viaggiatori Stranieri Nelle Biblioteche Napoletane Del Settecento.' *Rivista Italiana Di Studi Napoleonici* 31: 143–68.

Trombetta, Vincenzo. 2014. 'Le Biblioteche Italiane Nei Voyages Di Antoine-Claude Valery.' In *Il Libro al Centro. Percorsi Fra Le Discipline Del Libro in Onore Di Marco Santoro*, edited by Rosa Marisa Borraccini, Alberto Petrucciani, Carmela Reale, and Paola Zito, 3–16. Napoli: Liguori.

Tursi, Angiolo. 1956. 'Di Una Bibliografia Di Viaggiatori Stranieri in Italia.' *Nuova Rivista Storica* XL: 1–13.

Van den Abbeele, Georges. 1992. *Travel as Metaphor: From Montaigne to Rousseau*. Minneapolis: University of Minnesota Press.

Vianello, Nereo. 1968. *La Raccolta Di Angiolo Tursi Nella Biblioteca Marciana*. Venezia: Stamperia di Venezia.

Zavattoni, Francesco. 2020. 'Using Correspondence to Underline Changes in a Historiographical Network. Ideas for an Inter-Archival Analysis Starting from the Correspondence between Armando Saporì, Gino Luzzatto, Roberto S. Lopez and Angelo Sraffa.' *JLIS.it* 11 (1): 106–29. <https://doi.org/10.4403/jlis.it-12582>.

## Beyond transformative agreements. Library strategies towards Open Science

Maria Cassella<sup>(a)</sup>

a) Università di Torino, Biblioteca "Norberto Bobbio", <https://orcid.org/0000-0002-7229-8921>

**Contact:** Maria Cassella, [maria.cassella@unito.it](mailto:maria.cassella@unito.it)

**Received:** 13 December 2023; **Accepted:** 29 February 2024; **First Published:** 15 May 2024

### ABSTRACT

Transformative agreements are increasingly supporting the amount of open access articles, but they also pose big dilemmas to libraries, as costs are increasing and they have a time limit.

This is the rationale for libraries to imagine new strategies that go beyond the transformative agreements, i.e. the support to the Diamond Open Access, the development of new tools for a national repository network, the commitment to policies that foster innovative practices in the research assessment and so on.

The author argues that collaboration among libraries, funders and publishers is the key to move forward with the open access achievements. Assuming that there will always be costs involved in supporting the Gold Road, libraries, publishers, and research funders will need to conceive in the future new sustainable agreements, both economically and ethically.

### KEYWORDS

Transformative agreements; Open Science; Open Access; Gold Road.

## Oltre i *transformative agreements*. Le strategie delle biblioteche verso la scienza aperta

### ABSTRACT

Gli accordi trasformativi stanno sostenendo la crescita degli articoli ad accesso aperto, ma pongono anche grossi dilemmi alle biblioteche, in quanto hanno una scadenza temporale e comportano costi crescenti per le biblioteche accademiche.

Per questo motivo le biblioteche devono cercare strategie alternative che vadano oltre gli accordi trasformativi, ad esempio, il sostegno al Diamond Open Access, lo sviluppo di accordi per la costruzione di una rete di repository, il sostegno alle politiche e agli strumenti alternativi per la valutazione della ricerca e così via.

L'autrice sostiene che la collaborazione tra biblioteche, finanziatori ed editori è la chiave per sostenere l'Open Science. Partendo dal presupposto che il supporto alla Gold Road comporterà inevitabilmente dei costi, le biblioteche dovranno costruire, insieme agli editori e ai finanziatori della ricerca, accordi alternativi economicamente ed eticamente sostenibili.

### PAROLE CHIAVE

Accordi trasformativi; Open Science; Open Access; Gold Road.

La concezione della biblioteca aperta è *in primis* l'affermazione di un principio di giustizia sociale e l'affermazione della centralità del lettore rispetto alla collezione, tanto da poter dichiarare che la biblioteca o è open o non è.

(Mauro Guerrini, *Biblioteconomia come impegno civile*, Lectio magistralis, 2023)

Il recente successo dei contratti trasformativi (*transformative agreements*) ma anche l'aumento vertiginoso dei costi per sostenerli ha portato con sé inevitabili interrogativi nel movimento della scienza aperta e ha fatto sorgere dubbi sul futuro, più che sull'efficacia, dei contratti stessi e sulle scelte strategiche che le biblioteche hanno fin qui adottato per sostenere l'accesso aperto. Nonostante il successo del modello trasformativo che, come emerge dai dati raccolti dall'*ESAC Initiative*, il progetto di monitoraggio dei contratti coordinato dalla Max Planck Digital Library,<sup>1</sup> sta sostenendo concretamente la crescita dell'accesso aperto, appare sempre più evidente come i contratti trasformativi, ideati per favorire il passaggio delle riviste ad un modello aperto, per limitare il fenomeno del cosiddetto *double dipping*, e per contenere la spesa delle biblioteche per l'accesso ai contenuti, non siano riusciti a sciogliere i nodi fondamentali che ruotano intorno alla scienza aperta. La stessa caratteristica di transitorietà dei contratti dovrebbe imporre alle biblioteche, così come agli enti finanziatori, alle governance degli Atenei, agli autori accademici ecc. di interrogarsi sul futuro dell'accesso aperto e sul superamento dei contratti stessi.

Occorre, dunque, allargare la prospettiva e operare scelte strategiche di diversa natura: sostenere le attività di sensibilizzazione verso le comunità scientifiche, formare alla pratica della scienza aperta i giovani ricercatori, creare incentivi concreti per chi pratica spontaneamente la scienza aperta in tutte le sue declinazioni, innovare i meccanismi di valutazione della ricerca, migliorare gli strumenti per la condivisione dei risultati della ricerca, potenziare l'infrastruttura tecnologica attraverso scelte tecniche e strategiche (ad esempio tramite accordi mirati con network scientifici certificati) in grado di rendere i repository istituzionali sempre più interconnessi, visibili e usabili, così da riuscire ad inserirli pienamente nel flusso della ricerca, concentrando i propri sforzi sul processo più che sul prodotto finale.

## La lunga strada verso l'accesso aperto: dalla via verde e dalla via aurea ai contratti trasformativi (passando per la via rossa)

All'inizio del percorso verso la scienza aperta l'attenzione dei ricercatori, delle biblioteche di ricerca, nonché di tutti i numerosi stakeholder della comunicazione scientifica era focalizzata sulla risposta a due necessità:

- quella di utilizzare la rete per superare il problema dei costi crescenti delle riviste scientifiche;

---

<sup>1</sup> L'*ESAC Initiative* è un progetto collaborativo concepito al fine di rendere pubbliche le informazioni sui contratti trasformativi, sui costi e sulle clausole degli stessi e sul grado di avanzamento dell'accesso aperto nel mercato della comunicazione scientifica.

- quella di eliminare le barriere economiche e legali che rendevano la conoscenza un bene inaccessibile.

Due le strade che nel 2002 venivano suggerite dalla Budapest Open Access Initiative (BOAI) per rendere accessibili le pubblicazioni generate dai progetti di ricerca: quella del *self-archiving* ovvero dell'archiviazione degli articoli già sottoposti a revisione tra pari (*peer-reviewed*) in repository certificati, istituzionali o disciplinari, (la cosiddetta via verde) e quella della pubblicazione in riviste ad accesso aperto (la via aurea).

La prima strategia è rimasta praticamente immutata negli ultimi venti anni: l'infrastruttura tecnologica si è evoluta, si sono sviluppate reti di cooperazione e di certificazione intorno ai repository (ad esempio COAR, la Confederation of Open Access Repositories – l'associazione che riunisce i repository managers che lavorano allo sviluppo e all'implementazione dei repository) e comunità di utenti e sviluppatori (la DSpace community o la Islandora community, per citarne due tra le più attive) ma l'idea è rimasta sostanzialmente la stessa. L'archiviazione nei repository istituzionali è stata anche sostenuta negli anni da policy più o meno efficaci, più o meno "mandatarie". Da quella pioniera della Faculty of Arts & Sciences di Harvard del 12 febbraio 2008, nella quale gli autori concedevano alla Facoltà il diritto di rendere disponibili le proprie pubblicazioni, fino alle più blande policy italiane che suggeriscono, consigliano, raramente obbligano, chi fa ricerca a mettere a disposizione una copia delle proprie pubblicazioni nel repository istituzionale mantenuto dal proprio ateneo.<sup>2</sup>

Dal canto loro, i sistemi bibliotecari di ateneo hanno promosso la redazione e la pubblicazione delle policy a favore dell'accesso aperto e messo a disposizione, con tempi e modelli organizzativi diversi, personale, competenze e servizi per sostenere la pratica dell'archiviazione dei prodotti della ricerca, termine non felicissimo ma ormai consolidato in ambito accademico con il quale vengono denominati gli articoli, le monografie, le curatele, i dataset, le miscellanee, i *proceedings* ecc. tutto ciò che i ricercatori producono e pubblicano come risultato del loro percorso di ricerca. In Italia la via verde si è intersecata e sovrapposta da tempo con il processo di valutazione della ricerca: grazie ad IRIS, ossia alla suite di applicazioni per la ricerca e la valutazione sviluppata e mantenuta dal CINECA, i repository per l'accesso aperto delle università italiane si sono trasformati in CRIS (Current Research Information System), l'acronimo anglosassone che indica gli archivi certificati con prioritarie funzioni di supporto ai processi di valutazione della ricerca: VQR, ASN, Terza missione ecc.<sup>3</sup> In prospettiva diacronica è difficile valutare se questa scelta, praticamente obbligata per gli Atenei italiani che utilizzavano IRIS, abbia giovato o meno all'accesso aperto; di fatto, resta l'impressione che i CRIS vengano utilizzati dai ricercatori per archiviare i soli metadati delle pubblicazioni, mentre il full text resti nella maggior parte dei casi ad accesso riservato, disponibile internamente per i processi di valutazione e pubblicamente solo attraverso un pulsante di richiesta all'autore.<sup>4</sup>

---

<sup>2</sup> Le policy italiane sono raccolte sulla pagina wiki mantenuta dall'Università di Pisa [http://wikimedia.sp.unipi.it/index.php/OA\\_Italia/Regolamenti\\_e\\_Policy\\_sull%27Open\\_Access](http://wikimedia.sp.unipi.it/index.php/OA_Italia/Regolamenti_e_Policy_sull%27Open_Access).

<sup>3</sup> I CRIS italiani sono mantenuti da CINECA. L'elenco delle installazioni IRIS in Italia si trova sulla pagina wiki del CINECA <https://wiki.u-gov.it/confluence/display/public/UGOVHELP/List+of+IRIS+Installations++Elenco+delle+installazioni+IRIS>. A marzo 2022, ultima data di aggiornamento del sito, si contavano 80 installazioni.

<sup>4</sup> Ad esempio, sul portale IRIS dell'università di Torino alla data del 6 febbraio 2023 su un totale di 259.111 prodotti 60.660 sono ad accesso aperto. Sul portale IRIS dell'università di Bologna su 294.297 prodotti 33.560 sono open access; sul portale IRIS dell'università di Firenze su 237.520 prodotti totali risultano 37.558 prodotti ad accesso aperto.

Di fatto, ciascuna università raccoglie annualmente il dato percentuale dei prodotti archiviati ad accesso aperto nei rispettivi CRIS, ma non c'è una statistica nazionale che rilevi quale sia la crescita percentuale di pubblicazioni ad accesso aperto archiviate nei CRIS italiani.

Poco efficace per incrementare il numero di pubblicazioni<sup>5</sup> ad accesso aperto si è rivelata anche la decisione dell'ANVUR di inserire nel bando dell'ultimo esercizio nazionale di valutazione, la VQR 2015-2019, l'obbligo di rendere disponibili in full text nei repository istituzionali i prodotti della ricerca conferiti per la valutazione (art. 8 del Bando *Valutazione Qualità della Ricerca 2015-2019*), sia perché era possibile derogare all'obbligo per gli articoli risultati di una ricerca finanziata per una quota inferiore al 50% con fondi pubblici,<sup>6</sup> sia perché i 182.648 prodotti valutati nella VQR 2015-2019 rappresentavano solo una parte dell'intera produzione scientifica degli Atenei italiani nel quinquennio sottoposto a valutazione (ANVUR 2022).<sup>7</sup>

Nel nuovo bando per il prossimo esercizio di valutazione (VQR 2020-2024) pubblicato il 31 ottobre 2023 è stato nuovamente inserito da ANVUR l'obbligo per gli Atenei di rendere disponibili ad accesso aperto i prodotti sottoposti a valutazione (art. 8 del *Bando Valutazione Qualità della Ricerca 2020-2024*) (ANVUR 2023).<sup>8</sup> Anche in questo caso sono previste deroghe all'obbligo (periodo di embargo ed esclusione per gli articoli risultati di una ricerca finanziata per una quota inferiore al 50% con fondi pubblici).

La via aurea si è rivelata molto più complessa e tortuosa della prima nel sostenere il percorso verso la scienza aperta.

Da un lato, il supporto alla pubblicazione di riviste e di monografie ad accesso aperto si è concretizzato nella realizzazione di piattaforme di pubblicazione digitali grazie allo sviluppo di software open source per la pubblicazione quali, ad esempio, OJS o Lodel.<sup>9</sup> Dall'altro le biblioteche hanno investito una parte considerevole delle proprie risorse per sostenere i costi della pubblicazione ad accesso aperto favorendo la transizione di articoli e di monografie verso un modello aperto.

---

<sup>5</sup> I prodotti conferiti all'ANVUR per la VQR sono di diverse tipologie; oltre le pubblicazioni tra i prodotti valutati rientrano anche brevetti, traduzioni, commenti scientifici, curatele ecc.

<sup>6</sup> La legge 112/2013 obbligherebbe a rendere disponibili ad accesso aperto le pubblicazioni finanziate al 50% più 1 con fondi pubblici e che abbiano almeno due uscite annue, cioè – nella maggior parte dei casi – articoli su rivista.

<sup>7</sup> Per ottenere alcuni dati significativi è sufficiente accedere alla relazione finale dell'ANVUR sulla VQR 2015-2019. Da questa si evince che il totale dei prodotti valutati nel quinquennio è stato di 182.648; in pratica, una media di 2,5 prodotti per singolo ricercatore nel quinquennio di riferimento.

Di contro, e nonostante l'Italia sia ben al di sotto della media europea nell'investimento sulla ricerca scientifica (vedi l'articolo: *A proposito di innovazione, quanto spende l'Italia in ricerca e sviluppo*, "Il Sole 24 Ore. Economia", 3 dicembre 2023 [https://www.infodata.ilsole24ore.com/2022/12/03/a-proposito-di-innovazione-quanto-spende-litalia-in-ricerca-e-sviluppo/?refresh\\_ce=1](https://www.infodata.ilsole24ore.com/2022/12/03/a-proposito-di-innovazione-quanto-spende-litalia-in-ricerca-e-sviluppo/?refresh_ce=1)) l'Italia, si colloca ormai da alcuni anni tra i Paesi occidentali con più alta produzione scientifica.

Lo conferma anche la *Relazione sulla ricerca e sull'innovazione in Italia* del Consiglio Nazionale delle Ricerche: l'ultima edizione del 2021 del Rapporto mostra una crescita costante delle pubblicazioni indicizzate in Web of Science: nel 2021 erano quasi 160.000. Considerando che, per quanto ampia, la copertura della produzione scientifica del Web of Science è comunque parziale, il numero di prodotti sottoposti per la VQR 2015-2019 corrisponderebbe ad un quinto del totale delle pubblicazioni di ricerca prodotte in Italia nel quinquennio.

<sup>8</sup> [https://www.anvur.it/wp-content/uploads/2023/10/Bando-VQR-2020-2024\\_31ottobre.pdf](https://www.anvur.it/wp-content/uploads/2023/10/Bando-VQR-2020-2024_31ottobre.pdf).

<sup>9</sup> OJS è un software open source per la gestione e la pubblicazione di riviste elettroniche, sostenuto e reso disponibile dal Public Knowledge Project sotto la GNU General Public License. Lodel <https://www.openedition.org/10905?lang=en> è un Content Management System creato nel 2000 per la piattaforma francese *Revue.org* per gestire la pubblicazione di riviste online.



La prima strategia è stata concettualmente di supporto per le università al lancio di numerose piattaforme digitali di pubblicazione, tra le altre: Riviste UniMI, Università di Milano,<sup>10</sup> SIRIO, Università di Torino, OJS Alma DL Journals, Università di Bologna, ESE Publications, Università del Salento, SHARE riviste, Università di Napoli ecc. – e alla costituzione di alcune University press italiane.

Sulle University press si potrebbe aprire, in realtà, un lungo capitolo a parte.

Nel 2024 le University press italiane sono 17: alcune nascono come naturale evoluzione delle piattaforme istituzionali di pubblicazione, altre si sono sviluppate indipendentemente dai sistemi bibliotecari di ateneo, sotto la spinta del corpo docente e grazie a finanziamenti straordinari. In Italia, si raggruppano nel coordinamento UPI, l'associazione delle University press italiane. Le University press pubblicano prevalentemente, anche se non esclusivamente, monografie di ricerca, manualistica e periodici nelle scienze umane e sociali;<sup>11</sup> anche se si tratta di case editrici no profit che si ispirano nei regolamenti ai principi dell'accesso aperto,<sup>12</sup> adottano un modello ibrido per l'accesso alle proprie pubblicazioni, con alcune case editrici fortemente orientate all'accesso aperto (RomaTre Press e Firenze University Press, ad esempio) e altre su posizioni decisamente più tradizionali (Pisa University Press, Editrice LAS dell'Università Pontificia Salesiana).

Non diversamente anche negli Stati Uniti un recente rapporto della società di ricerche e consulenza Ithaka SR e dell'Associazione delle University Press statunitensi mette in evidenza come una percentuale molto bassa di titoli (5%) venga pubblicata ad accesso aperto dalle case editrici universitarie. Tra i fattori che concorrono a questo risultato negativo vengono citati dagli autori: la resistenza culturale degli umanisti a pubblicare ad accesso aperto e unicamente in formato digitale, la difficoltà ad attingere a fondi per coprire i costi della pubblicazione ad accesso aperto, la difficoltà per le piccole case editrici di gestire e mantenere piattaforme editoriali digitali.

Still, with all of this activity, open access book publishing remains on the fringe of most university press book programs. While a few presses such as MIT and University of Michigan have made a significant commitment to publish the monographs on their list in open access editions, and there are a number of newer presses founded to be all-OA such as Athabasca and Amherst, on the whole, OA titles make up less than five percent of the total new monographic output of university presses.

Why is this the case? For one thing, there are still pockets of cultural resistance among humanists who worry that OA publication will be viewed less favorably than a traditional print monograph in the tenure and promotion process. Beyond these cultural barriers, there are financial concerns. At the highest level, the level of funding available for research in the humanities lags well behind what we have seen for STEM. As a result, it has been challenging to develop open access business models for books similar to the grant-supported APCs and transformative agreements that journal publishers have adopted.

---

<sup>10</sup> Di recente l'Università di Milano ha fondato una sua University press. Le riviste pubblicate sulla piattaforma Riviste UNIMI dedicata fin dal 2008 alle riviste online di ateneo con oltre 50 titoli attualmente disponibili si possono considerare a pieno titolo il primo nucleo della neonata Milano University Press.

<sup>11</sup> Cataloghi e numeri sulle pubblicazioni delle University press italiane si trovano pubblicati sul sito dell'UPI all'url: <https://www.universitypressitaliane.it/>.

<sup>12</sup> Si vedano, ad esempio, i regolamenti di Genova University Press, della Bozen University Press o di Firenze University Press, la prima casa editrice universitaria italiana.

Moreover, smaller and even mid-sized university presses often lack the capital to experiment with new models. Perhaps more daunting still, with the decline in monograph sales over the last couple of decades, margins on academic books are so thin that publishers may fear that anything that threatens to cannibalize anticipated print sales of a scholarly title, such as a freely available open edition, is a threat to its viability (Brown et al. 2023).

Per quanto case editrici universitarie e piattaforme digitali abbiano raccolto negli anni un numero crescente di contenuti e assunto un valore per le comunità di ricerca, è ormai diffusa la consapevolezza che non risolvano i nodi fondamentali della comunicazione scientifica. Non è solo il dato relativo alla mancata adozione del modello open che penalizza le case editrici italiane. Una ricognizione realizzata dall'autrice a partire dal sito di UPI tra i cataloghi editoriali delle 17 University press italiane associate rivela, infatti, come si tratti di canali di pubblicazione utilizzati prevalentemente da chi fa ricerca nelle scienze umane e sociali.

Case editrici universitarie italiane e piattaforme istituzionali di pubblicazione risultano essere poco utilizzate dai ricercatori del settore biomedico che vi si affidano quasi esclusivamente per pubblicare manualistica a scopo didattico. La ricerca biomedica, infatti, si serve per comunicare i risultati della propria attività scientifica di un circuito internazionale e, per questo, è legata, più di altre discipline, a pochi grandi editori internazionali, ormai vere e proprie multinazionali nel campo dell'editoria come Elsevier, SpringerNature, e Wiley che detengono ampie quote del mercato mondiale delle pubblicazioni accademiche (ad esempio, secondo quanto riporta l'*ESAC Market Watch*<sup>13</sup> ovvero l'osservatorio sul mercato delle pubblicazioni scientifiche dell'ESAC, negli Stati Uniti, Elsevier controlla il 21% del mercato, SpringerNature l'11,6% e Wiley il 10,4%. Queste percentuali variano in modo non significativo da un Paese all'altro).

## Articoli e riviste

La strategia di sostenere la pubblicazione di articoli ad accesso aperto nelle riviste si è rivelata in assoluto la meno convincente per le biblioteche in quanto lunga nei tempi e tremendamente costosa. Gli editori, infatti, hanno escogitato soluzioni diverse, a volte decisamente ingegnose, per trasformare l'accesso aperto in nuovi modelli commerciali.

Sono nate in breve tempo le riviste ibride ovvero quelle riviste in abbonamento che chiedono agli autori il pagamento di una quota (*Article Processing Charge = APC*) per pubblicare un articolo ad accesso aperto. La pratica, estremamente remunerativa per gli editori, è stata poi estesa da questi anche alle monografie con costi variabili, non sempre esplicitamente dichiarati e notevolmente più alti rispetto alle APC richieste per pubblicare ad accesso aperto un articolo (per la pubblicazione di una monografia i costi partono dai 12.000 euro a salire).<sup>14</sup> Sostenere il pagamento delle APC

<sup>13</sup> *ESAC Market Watch*: <https://esac-initiative.org/market-watch/>

<sup>14</sup> Poche le fonti pubbliche sui costi della pubblicazione di monografie ad accesso aperto. Sul sito dell'editore SpringerNature si trova la cifra di 13.700 euro per la pubblicazione open access di una monografia di 400 pagine <https://www.springernature.com/gp/open-research/journals-books/books/pricing>; sul sito dell'editore Routledge (Gruppo Taylor and Francis) il prezzo per la pubblicazione ad accesso aperto di una monografia è di 12.000 euro <https://www.routledge.com/our-products/open-access-books/publishing-oa-books/book-publication-charges#:~:text=Prices%20for%20a%20full%20monograph,between%2025%2D50%2C000%20words>

ha dato vita al controverso fenomeno editoriale del *double dipping* ovvero il doppio pagamento da parte delle università sia per leggere che per pubblicare ad accesso aperto e si è così rapidamente trasformata in una via rossa almeno fino a quando, con un salto temporale di più di dieci anni, nel 2015 nell'ambito della Conferenza sull'Open Access di Berlino, organizzata dal 2003 annualmente dalla Max Planck Gesellschaft, prende corpo una nuova ipotesi e cioè che l'ammontare delle risorse annualmente investite nelle sottoscrizioni dalle biblioteche fosse in grado di sostenere economicamente l'editoria ad accesso aperto:

“Tutto lascia pensare che il denaro già investito nel sistema editoriale della ricerca sia sufficiente per consentire una trasformazione che sia sostenibile per il futuro. È necessario condividere il principio che i fondi attualmente bloccati nel sistema di abbonamento alle riviste devono essere ritirati e reimpiegati per i servizi editoriali ad accesso aperto” (Schimmer et al. 2015).

## Il paradosso dei contratti trasformativi

A partire dalla Conferenza di Berlino del 2015 si fa strada l'idea dei contratti trasformativi (Borrego, Anglada, e Abadal 2021; Anderson et al. 2022; Parmhed e Säll 2023; Capaccioni 2021; Cassella 2023).

Nel settembre 2018 viene, infatti, lanciata da alcuni enti che finanziano la ricerca cOAlitionS, un'iniziativa molto concreta sostenuta dalla Commissione Europea e dallo European Research Council (ERC), l'organizzazione che finanzia la ricerca di frontiera in Europa, per ottenere in tempi molto brevi, più precisamente entro il 2021, accesso aperto e immediato ai risultati della ricerca finanziata con fondi pubblici o privati.

Tra gli strumenti che cOAlitionS sostiene vi è anche quello dei contratti trasformativi ovvero quei contratti in base ai quali gli editori mettono a disposizione l'accesso alle proprie riviste (*read*) a fronte di un valore del contratto stabilito sulla previsione di articoli ad accesso aperto che gli autori affiliati ad una istituzione o ad un consorzio di istituzioni potranno pubblicare nel periodo oggetto del contratto (*publish*).

I contratti trasformativi sono stati adottati in quanto considerati una strategia efficace per favorire la transizione delle riviste verso l'accesso aperto, purché transitori e di breve durata.

Si sono rivelati, tuttavia, un grande paradosso per le biblioteche. Attraverso questi contratti, infatti, le biblioteche sono diventate alleate consapevoli degli editori nel promuovere e sostenere, non solo finanziariamente, questo nuovo modello editoriale che altro non è se non un nuovo modello commerciale. Sulla base di questi accordi, le biblioteche alimentano di fatto un circuito che le obbliga a promuovere i servizi offerti dagli editori, facendo così crescere i costi di un contratto, favorendo ancora una volta il sistema oligopolistico dei principali editori internazionali.

Di contro è innegabile che, grazie ai contratti trasformativi, la percentuale di pubblicazioni ad accesso aperto sia in crescita costante, come si evidenzia dai dati forniti dall'*ESAC Initiative*.

A settembre 2023 l'*ESAC market watch* registra numerosi Paesi europei che si stanno avvicinando ad una percentuale del 90% di pubblicazioni disponibili ad accesso aperto: al primo posto nella classifica dell'*ESAC* compaiono la Finlandia e la Svezia con una percentuale di pubblicazioni ad accesso aperto pari all'84% del totale delle pubblicazioni, seguite dalla Norvegia (82%) e dall'O-

landa (82%) e così via. In questa classifica di Paesi virtuosi, l'Italia si colloca al ventiduesimo posto con una percentuale di pubblicazioni ad accesso aperto pari al 40% della produzione scientifica nazionale e ciò nonostante la CRUI abbia concluso nel secondo semestre del 2023 due nuovi contratti di questo tipo: uno con Elsevier, l'altro con l'Association for Computing Machinery.

L'altra faccia della medaglia di questa crescita vertiginosa degli articoli ad accesso aperto in alcuni Paesi del Nord Europa parla di un 68% delle riviste (1.589 titoli) che non hanno completato la transizione verso il modello aperto (cOAlitionS 2023), di costi crescenti (di natura sia economica che amministrativa) per le università e per le biblioteche, grazie ad un meccanismo in grado di autoalimentarsi ed estremamente gradito agli autori, di forti dubbi sulla tenuta dei contratti e sulla loro sostenibilità nel medio termine e di perplessità verso un'editoria commerciale sempre più aggressiva a fronte di decisi segnali politici che vanno nella direzione opposta e che vengono da interlocutori forti come il Consiglio dell'Unione Europea. Quest'ultimo un anno fa (maggio 2023) ha pubblicato una nota a favore di un sistema di editoria scientifica di qualità, trasparente, open, affidabile e equa.

Per non parlare degli effetti distorsivi sul mercato dell'editoria scientifica e dell'ulteriore rafforzamento degli oligopoli editoriali. Un tema che è stato evidenziato dall'Associazione Italiana per la Scienza Aperta (AISA) già nel 2020, in un post pubblicato sul sito dell'associazione a luglio 2020 e poi aggiornato a marzo 2022 (Pievatolo 2020).

AISA critica anche il sistema stesso delle negoziazioni che rende difficile concludere contratti che non "siano più che favorevoli agli oligopolisti dell'editoria scientifica". Nel dibattito che ne è seguito sono intervenuti anche alcuni attori accademici (Della Sala e Cubelli 2021), che, in un articolo pubblicato sul *Giornale Italiano di Psicologia*, hanno messo in evidenza i rischi di una capillare diffusione del modello trasformativo, l'Associazione Italiana Editori (Attanasio 2022) e la comunità bibliotecaria, divisa tra sostenitori e detrattori degli accordi trasformativi.<sup>15</sup>

Nonostante il successo, il futuro dei contratti trasformativi appare, comunque, concettualmente incerto oltre che poco sostenibile. Non a caso, la transitorietà resta uno dei requisiti fondamentali di questo tipo di contratti.

Se allarghiamo la prospettiva e rivolgiamo lo sguardo al dibattito internazionale, negli Stati Uniti e in Europa, e anche a quello nazionale, non c'è dubbio che i contratti trasformativi abbiano catalizzato l'attenzione e le risorse degli addetti ai lavori nella comunicazione scientifica a partire dal 2018, lasciando nell'ombra progetti e strategie complementari che, invece, soprattutto nel caso di università fortemente orientate alla ricerca, vanno adeguatamente sostenute dalla governance universitaria e, di rimando, dalle biblioteche: il Diamond Open Access, la via verde e il collegamento con i network disciplinari e di pratica utilizzati dalle comunità scientifiche, la creazione di incentivi per chi pubblica ad accesso aperto, il sostegno economico tramite fondi dedicati per i ricercatori che scelgono di pubblicare in riviste non coperte da accordi trasformativi ecc.

Si vedano, ad esempio, le strategie di biblioteche come quelle del MIT (Massachusetts Institute of Technology Libraries), negli Stati Uniti, o della biblioteca dell'EPFL (École Polytechnique Fédérale de Lausanne), in Europa, due università che ispirandosi ai principi della scienza aperta

---

<sup>15</sup> Si legga, ad esempio, il *thread*: *Oltre la "trasformazione"?* sulla lista di discussione OA-Italia, <https://liste.cineca.it/cgi-bin/mailman/listinfo/oa-italia>. Il primo messaggio di Paola Galimberti è del 22 novembre 2023.

scelgono di lavorare su più filoni contemporaneamente all'interno di un quadro istituzionale (*framework*) sull'open science ampio, monitorato e documentato.

Punto di forza delle biblioteche del MIT nel sostegno all'open science è avere adottato una policy Open Access già a partire dal 2009, la *MIT Faculty Open Access Policy*, nella quale è inserita una clausola che obbliga gli autori a cedere all'università in modalità non esclusiva i propri diritti di pubblicazione.<sup>16</sup> In virtù di questa policy open access il repository istituzionale del MIT – *DSpace@MIT* – è popolato da 51.546 documenti ad accesso aperto a gennaio 2024. Le MIT Libraries pubblicano sul proprio sito ogni mese il numero di download dal proprio repository istituzionale e mettono in evidenza gli articoli più scaricati. È possibile, inoltre, ottenere le statistiche dei download per autore, per dipartimento e per Paese. Le statistiche quantitative sono, infine, completate da storie ossia da narrazioni di coloro che utilizzano il repository con successo.

Nel caso dell'EPFL l'azione della biblioteca è rafforzata da quella dell'Open Science Unit e delle Facoltà.

Biblioteca e Open Science Unit collaborano in modo costante e interconnesso a seconda dei temi o dei progetti in questione.

Il quadro strategico verso l'accesso aperto dell'EPFL prevede: il sostegno all'Open Access publishing nelle sue diverse articolazioni con fondi dedicati, un Research Data Management Support, con un'ampia palette di servizi offerti e con un'attenzione particolare alle pratiche di apertura e all'applicazione concreta dei principi FAIR, un repository istituzionale attivo dal 2003 (Infoscience). Quest'ultimo è lo strumento utilizzato dall'EPFL per allinearsi alla *Switzerland's national Open Access strategy* che prevede che il 100% delle pubblicazioni finanziate con fondi pubblici divengano accessibili entro il 2024.

Le politiche verso la scienza aperta dell'EPFL sono coordinate da un Open Science Strategic Committee (OSSC), nominato dal Rettore, che propone alla direzione le strategie per implementare la scienza aperta in tutte le sue declinazioni. L'OSSC "è un gruppo di riflessione interno il cui obiettivo è sviluppare una visione lungimirante e ambiziosa sull'Open Science, con l'intento esplicito di rendere l'EPFL un'istituzione leader per l'Open Science e di radicare saldamente l'Open Science nella cultura della ricerca dell'EPFL".<sup>17</sup>

## Oltre i contratti trasformativi

Con il limite temporale del 2024 per i contratti trasformativi si materializzano in relazione al futuro della via aurea nuovi interrogativi.

Ancora volta sono gli enti finanziatori che stanno imprimendo una svolta al percorso della conoscenza verso la scienza aperta dopo la decisione di cOAlitionS di non sostenere più dalla fine del 2024 e, comunque, non oltre il 2027, per quelli già in essere, nuovi accordi trasformativi.

---

<sup>16</sup> Il testo della policy open access del MIT si trova pubblicato sul sito delle MIT Libraries nella sezione *Scholarly Communication* alla seguente URL: <https://libraries.mit.edu/scholarly/mit-open-access/open-access-policy/>. Nel 2017 il MIT ha anche adottato una clausola *opt-in* che concede agli autori che non sono coperti dalla *MIT Faculty Open Access Policy* la possibilità di cedere in modalità non esclusiva i propri diritti all'università.

<sup>17</sup> <https://www.epfl.ch/about/vice-presidencies/vice-presidency-for-academic-affairs-vpa/academic-committees/open-science-strategic-committee-oss/>.



Le attenzioni degli enti finanziatori si stanno, quindi, rivolgendo verso nuovi traguardi: il Diamond Open Access, il finanziamento volto a rafforzare l'infrastruttura tecnologica e le piattaforme istituzionali di pubblicazione, il sostegno alla pubblicazione in accesso aperto per i *grant* concessi dagli enti (vedi, ad esempio, l'Open Access Block Grant che l'Austrian Science Fund si appresta a lanciare nel 2024), il sostegno alle iniziative di pubblicazione di monografie ad accesso aperto, sia finanziando direttamente le case editrici universitarie, sia supportando iniziative quali, ad esempio, Knowledge Unlatched, The Open Library of Humanities, OpenBook Publishers ecc., il sostegno ad iniziative volte a rinnovare i meccanismi di valutazione della ricerca, quali, ad esempio, COARA (Coalition for Advancing Research Assessment).<sup>18</sup>

In Italia continuano a sussistere una serie di criticità che ostacolano la transizione verso l'accesso aperto, criticità che si sommano ad uno scenario che, come abbiamo descritto, è già di per sé molto complesso.

Gli enti finanziatori sono pochi, a parte il Ministero, qualche Fondazione bancaria, poche fondazioni private no-profit (ad es. Fondazione Telethon) e non hanno la forza economica che hanno in altri Paesi finanziatori come, per esempio, i Research Councils in UK, la Deutsche Forschungsgemeinschaft (DFG) in Germania, l'Austrian Science Fund in Austria ecc.

Quanto al Ministero dell'Università e Ricerca ha pubblicato con il Decreto Ministeriale n. 268 del 28 febbraio 2023 il Piano Nazionale sulla Scienza Aperta (PNSA) 2021-2027.

Il Piano individua cinque assi di intervento: pubblicazioni scientifiche, dati, valutazione della ricerca, partecipazione e apertura dei dati della ricerca su Covid -19.

Quanto al primo asse gli obiettivi indicati dal Piano sono i seguenti:

- fornire accesso aperto immediato alle pubblicazioni scientifiche finanziate con fondi pubblici;
- incentivare il ricorso a forme non commerciali di pubblicazione in accesso aperto
- dotare l'Italia di un quadro normativo organico in materia di diritto d'autore che renda possibile l'accesso aperto alle pubblicazioni scientifiche;
- dotare l'Italia di un sistema di monitoraggio sull'attuazione del principio dell'accesso aperto alle pubblicazioni scientifiche;
- razionalizzare e rendere trasparenti i contratti di abbonamento alle riviste scientifiche delle istituzioni accademiche e di ricerca;
- promuovere e incentivare la creazione di Risorse formative aperte (*Open Educational Resources*).

Il PNSA non è stato, tuttavia, fin qui sostenuto da finanziamenti ministeriali che fossero in grado di renderlo operativo.

- Sul versante dell'infrastruttura tecnologica, i repository istituzionali sono fermi alla funzione di sostegno alla valutazione della ricerca; vengono utilizzati come CRIS e poco si investe in termini strategici ed economici nella loro reingegnerizzazione per cercare di renderli maggiormente funzionali al processo di ricerca.

---

<sup>18</sup> A ottobre 2023 COARA è stata firmata da 644 tra istituzioni, enti finanziatori, università, agenzie di valutazione della ricerca – tra le quali per l'Italia l'ANVUR – associazioni e accademie di ricerca,

- Tranne poche lodevoli eccezioni,<sup>19</sup> non si raccolgono e, soprattutto, non si pubblicano dati ufficiali sulla crescita dell'Open Access per indirizzare le scelte strategiche della governance universitaria. Quest'ultima sembra considerare non prioritarie le azioni che sostengono la scienza aperta, nonostante le policy sull'accesso aperto già adottate, la firma da parte di 71 università italiane della Dichiarazione di Messina 2.0 nel Decennale della Dichiarazione di Messina (2004-2014) e, più di recente, il PNSA.

Resta, dunque, come soluzione pressoché unica in Italia quella di abbracciare la strategia dei contratti trasformativi; cosa seguirà dopo non è chiaro. Verosimilmente, quale che sia il modello (*membership*, *Subscribe to Open*, altro?), le università continueranno a pagare per sostenere i costi dell'accesso aperto così come una volta sostenevano gli abbonamenti.

È il caso, quindi, di interrogarsi fin da ora sul futuro del modello trasformativo.

Così come sta accadendo in altri Paesi. In Svezia, ad esempio, dove già nel 2021 in seno alle Università è stato costituito un gruppo di lavoro sul superamento del modello trasformativo: *Beyond transformative agreements*.<sup>20</sup>

L'idea è quella di trovare un modello aperto e sostenibile per accedere alla conoscenza scientifica per indirizzare le scelte del consorzio svedese Bibsam, che negozia i contratti per tutte le università in Svezia, adottando un piano strategico nazionale a medio termine e collaborando con gli editori e con gli enti finanziatori.

Questa collaborazione è la chiave per procedere in futuro ad uno sviluppo dell'accesso aperto. Preso atto che il sostegno alla via aurea comporta dei costi, le biblioteche dovranno costruire nel tempo, insieme agli editori e agli enti finanziatori, accordi e percorsi sostenibili alternativi ai contratti trasformativi.

---

<sup>19</sup> Ad esempio, l'Università di Milano, l'Università di Modena e Reggio Emilia e l'Università di Bolzano riversano i propri dati in OpenAPC, l'iniziativa internazionale promossa dalla Bielefeld University Library. Così fanno anche il CNR e l'Istituto di Oncologia del Veneto.

<sup>20</sup> <https://www.su.se/english/news/open-access-need-to-move-away-from-transformative-agreements-1.683787>.

## Riferimenti bibliografici

Anderson, Graham, Jade Heyman, e Maggie Simmons. 2022. "How transformative agreements are actually transforming the subscription systems: a society publisher's perspective." *Insights* 35. <https://doi.org/10.1629/uksg.579>.

ANVUR (Agenzia nazionale di valutazione del sistema universitario e della ricerca). 2022. *Valutazione della Qualità della Ricerca VQR 2015-2019. Rapporto finale ANVUR: statistiche e risultati di compendio*. [https://www.anvur.it/wp-content/uploads/2022/07/VQR-2015-2019\\_Rapporto\\_Finale\\_EC\\_21luglio2022.pdf](https://www.anvur.it/wp-content/uploads/2022/07/VQR-2015-2019_Rapporto_Finale_EC_21luglio2022.pdf).

ANVUR (Agenzia nazionale di valutazione del sistema universitario e della ricerca). 2023. *Bando Valutazione Qualità della Ricerca 2020-2024*. [https://www.anvur.it/wp-content/uploads/2023/10/Bando-VQR-2020-2024\\_31ottobre.pdf](https://www.anvur.it/wp-content/uploads/2023/10/Bando-VQR-2020-2024_31ottobre.pdf).

Attanasio, Piero. 2022. "I punti di vista sull'accesso aperto." *Giornale italiano di psicologia* 49 (3): 499-507. <https://www.rivisteweb.it/doi/10.1421/105465>.

Borrego, Angel, Lluís Anglada, e Ernest Abadal. 2021. "Transformative agreements: do they pave the way to open access?." *Learned publishing* 34 (2). <https://doi.org/10.1002/leap.1347>.

Brown, Laura, Maya Dayan, Brenna McLaughlin, Roger C. Schonfeld, John Sherer, e Erich van Rijn. 2023. *Print revenue and open access monographs: a university press study*. <https://doi.org/10.18665/sr.319642>.

Capaccioni, Andrea. 2021. "Oltre gli abbonamenti: che cosa sono i contratti trasformativi." *JLIS.it* 12 (1): 47-53. <https://doi.org/10.4403/jlis.it-12664>.

Cassella, Maria. 2023. "I contratti trasformativi: un paradosso per le biblioteche." *Biblioteche oggi* 41 (5): 12-20. <http://dx.doi.org/10.3302/0392-8586-202305-012-1>.

cOAlitionS. 2023. *Transformative Journals: analysis from the 2022 reports*. cOAlitionS. <https://www.coalition-s.org/blog/transformative-journals-analysis-from-the-2022reports/>.

Della Sala, Sergio, e Roberto Cubelli. 2021. "La beffa e il danno delle nuove politiche dell'editoria scientifica." *Giornale italiano di psicologia* 48 (3): 599-608.

Parmhed, Sara, e Johanna Säll. 2023. "Transformative agreements and their practical impact: a librarian perspective." *Insights* 36. <https://doi.org/10.1629/uksg.612>.

Pievatolo, Maria Chiara. 2020. *Un'offerta che non si può rifiutare?* (blog). <https://aisa.sp.unipi.it/accordi-trasformativi-unofferta-che-non-si-puo-rifutare/>.

Schimmer, Ralf, Kai Karin Geschuhn, e Andreas Vogler. 2015. *Disrupting the subscription journals' business model for the necessary large-scale transformation to open access*. <http://dx.doi.org/10.17617/1.3>.

Si ringraziano Alessandra Bianchi e Lorenza Salvatori per le informazioni sulle politiche e le attività di Open Science presso l'EPFL.

# Towards a semi-automatic classifier of malware through tweets for early warning threat detection\*

Claudia Lanza<sup>(a)</sup>, Lorenzo Lodi<sup>(b)</sup>

a) University of Calabria, <https://orcid.org/0000-0002-3018-1987>

b) Zanasi & Partners

**Contact:** Claudia Lanza, [claudia.lanza@unical.it](mailto:claudia.lanza@unical.it); Lorenzo Lodi, [lorenzo.lodi@zanasi-alessandro.eu](mailto:lorenzo.lodi@zanasi-alessandro.eu)

**Received:** 29 November 2023; **Accepted:** 18 March 2024; **First Published:** 15 May 2024

## ABSTRACT

This paper presents a method for developing a malware ontology structure by detecting malware instances on Twitter. The ontology represents a semi-automatic classifier fed by the data extracted from tweets. In particular, the automatic part of the presented methodology relies on a pattern-based approach to detect trigger expressions leading to new information about malware, whilst the manual one covers the evaluation of the results by domain-experts, who also validate the reliability of the semantic relationships within the ontology framework. We present preliminary results on the application of our methodology to tweets extracted from MalwareBazaar database showing how the documents' collection analysis, through Natural Language Processing (NLP) tasks, can support the knowledge retrieval and documents' classification procedures for building early warning system of detected malware. Results obtained from this research paper within the time framework of 2023 are referred to the previous version of the current social network X.

## KEYWORDS

Malware; Classification; NLP; Twitter; Text Mining.

\* The authors, although they have jointly worked on the paper, have specifically covered the following sections: Claudia Lanza specifically dealt with sections "Methodology", "Corpus design", "Terminological extraction", "Rule-based pattern recognition", "Classification tool", Lorenzo Lodi sections "Related works" and "Results". Section "Conclusion" is to be attributed to both authors.

This paper is the result of the research activities carried out within the framework of the UE co-financing – FSE REACT-EU, PON Ricerca e Innovazione 2014-2020.

© 2024, 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

This paper will describe a preliminary study on a method to detect new data about malware and structure them in an ontology model. The ontology represents a means through which is possible to build a classifier able to structurally organize the knowledge behind the malicious events within the cyber-sphere. The novelty brought from our approach can be envisaged in the source documentation taken into account to construct the malware classifier. More specifically, not only the set of documents considered, i.e., tweets, but also the techniques applied to retrieve the information about the malware can constitute the originality of this work. In detail, we propose an approach which, through the Natural Language Processing (NLP) tasks over the normalized group of tweets is meant to systematize the informative set of obtained data into an ontology framework. The ontology represents the classifier created in a semi-automatic way and can help cyber analysts in creating a conceptual structure to infer knowledge about malicious events as well as in supporting malware triaging operations from a semantic point of view.

## Related Works

The detection of new events from Twitter represents a common research branch and usually is focused on the interpretation of tweets' content from a topic-based approach, as Twitter Stand platform created at Maryland University (Sankaranarayanan 2009) shows by capturing the late breaking news from tweets becoming popular topics per each country. Regarding the cyber threats detection from Twitter, Gaglio (2015) proposed an extension of Soft Frequent Pattern Mining (SFPM) through an improved topic detection algorithm with the presentation of Twitter Live Detection Framework (TLDF) able to face the new incoming data from a topic detection perspective. Cordeiro (2012) presented a work on topic inference events from the social platform by using the Latent Dirichlet Allocation topic inference model based on Gibbs Sampling. Concione et al. (2017) also proposed a methodology to detect, and give an alert on, new malware using the data coming from reliable Twitter's subscribers by means of a Bayes naive classifier. Specifically, they worked with the "Bayes classifier trained on a set of tweets containing an equal number of i) events related to security attacks, viruses, malware, and ii) generic messages", and realized "groups of tweets discussing the same topic, e.g, a new malware infection, are summarized in order to produce an alert". The authoritativeness of users selected by the authors has been based on an "influence metric" which links the users' interaction with the community in terms of retweets, feelings, answers and number of likes. Another study covering cyber threat detection from Twitter is that of Sabottke (2015), where the authors specifically refer to the exploit detection by creating a Twitter-based exploit detector. This system detects on Twitter the use of exploits against known vulnerabilities by looking within the tweets containing texts mentioning vulnerabilities and comparing, as ground truth, to CVE IDs as well as ExploitDB and classifying them using the SVM classifier.

Given the increase in the variants of malware, a resource able to analyze similarities and gather these features as informative elements in a classification structure becomes a valid means for the enhancement of cybersecurity predictive actions. In the literature review, malware classification has been considered as an urgent and evolving study to foster and a wide range of techniques has



been proposed within the scientific community. The most common way to identify increasingly complicated malware is signature-based, (Akhtar and Feng 2022) offer a literary review of new machine learning based techniques which aim at analyzing the efficacy of those approaches in the identification of Polymorphic malwares.

The method we hereby propose may show potential benefits for the malware identification independently from their code, but, conversely, by analyzing the contents of tweets dealing with new malware outbreaks. A comprehensive overview of Deep Learning (DL) tasks used to classify malware events is offered by Mathews (2019). Still referring to DL approaches, Kalash (2018) proves the better performance of Convolutional Neural Networks (CNN) in identifying and properly classifying malware, as well as Tekerek and Yapici (2022), Adem and Yapici (2022), and Habibi (2023). Amongst the main tested malware classification procedures, the call graph clustering approach by Kinable and Kostakis (2011) has been executed for the detection of structural similarities between malicious behaviour samples. Echo state networks (ESNs) and recurrent neural networks (RNNs) have been exploited, among a vast number of experts, by Pascanu (2015) to detect malicious files. Annachhatre (2015) applies hidden Markov models and cluster analysis to discover malware classes, whereas Tang (2023) uses LightGBM to identify malware families. Mirza (2018) proposes a combination of machine learning approaches employed over a group of features extracted from a wide corpus made of benign and malicious files through a bespoke feature extraction tool. A set of studies focus on semantics in malware code detection under the lens of obfuscation used by attackers to hide the actual code and the behaviour of malware (Singh 2018; Sahu 2014; Christodorescu 2005). Against this background one cannot fail to mention the main classification systems which guide the comprehension and representation of the malware families, their features and targets in their attacking processes, both released by the Mitre Corporation, an American association that supports government structures specifically with respect to the cybersecurity area. The first one is the MITRE ATT&CK platform, a web-based tool that helps in enhancing knowledge on threat tactics and techniques applicable to several operative systems: it is subdivided into 14 tactics and 188 techniques representing the ways by which attackers can perform a cyber attack against the infrastructures. This tactics' representation supports either the acquisition of a knowledge base in the cyber adversaries' techniques and a dictionary modeling of this information under a classification perspective to overcome cyber threats (Xiong et al. 2022; Georgiadou 2021; Kwon 2020). The second one is the Common Attack Pattern Enumeration (CAPEC) catalogue that gathers under a tree-like configuration a range of attacks' mechanisms and attacks' domain by merging the patterns according to the common features they share (Kotenko and Doynikova 2015). This structure becomes essential to understand the adversaries' behaviours and to create a common dictionary and a classification taxonomy of the attacks' patterns to be used by analysts or developers working within the cyber defense field (Andrei 2019).

In this paper we propose a new method to detect in a predictive way the new malware denomination and classify them by a hierarchical structure through an ontological configuration . The proposed method will provide institutions working within the cybersecurity strategy plans with a classification tool to manage their knowledge base when exposed to cyber threats. This will be empowered by an updated classification system, i.e., an ontology, covering the connections with the malware families' attributes starting from a tweets analysis.

With respect to other malware ontologies existing and largely used in the scientific community (see for instance those of Rastogi 2020; Zareen 2016; Huang 2010), the one proposed in our work will aim at potentially including the zero-day cyber-attack events through the detection of the semantic information within tweets in a father-son classes' relationship and object properties, by exploiting the advantages of OWL language (McGuinness and Van Harmelen 2004; Antoniou and Van Harmelen 2004; Wang 2004). The main contributions of this paper can be summarized as follows: (i) a tweets dataset taken into consideration to realize the classification tool, (ii) the terminological analysis of the texts extracted from Twitter (iii) that leads to (iv) the employment of NLP methods, particularly based on a patterns-based approach (León-Araúz 2013; Auger and Barrière 2008) to find trigger expressions in tweets, used to retrieve in a predictive way the upcoming malware classes within the cybersecurity spectrum.

## Methodology

Our methodology relies on a source corpus containing a set of documents constituted by tweets, which, for their intrinsic nature, are marked by a regular configuration as well as by an unstructured way to formulate texts from a linguistic perspective (Arora and Kansal 2019; Kumar and Das 2013). The reason why tweets have been chosen as source corpus is related to the predictive goal of this work towards early malware detection and their subsequent classification: tweets are a real-time information that can allow a punctual acquisition of knowledge to be studied. Indeed, "Tweets are free text micro blogging posts of no more than 140 characters, used by millions of people around the world; with one important characteristic, its real-time nature. Although their length per post is limited, the variety of words that can be used is high. If we take in account that each single word represents a different variable, a tweet is considered a high dimensional data."(-Gutierrez 2014, 168)

Moreover, as observed by Barnard (2016), each tweet owns an inner narrative form of communication that synthetically highlights salient information to be retrieved. Thanks to the character limit behind the posts' publication the information within tweets can be spread in an immediate way without facing the semantic noise that generally implies a massive removal of unnecessary text (Gupta and Rao 2020). On the basis of using the Twitter microblogging as a major source of information (Bakliwal 2012) our approach that leads to a malware predictive detection and their classified configuration can be described by the steps depicted in Figure 1.

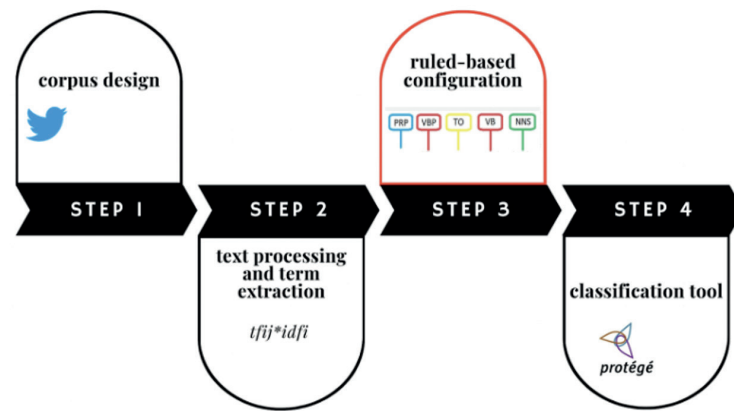


Figure 1. Our methodology steps for tweets processing, extraction and classification

## Corpus design

The collected corpus related to the predictive discovery of new malware denominations is composed by a set of tweets' contents published by few reporters identified as a test set to experiment the methodology. Ranking the social activity of Twitter users has represented a research interest in several areas of study and it mainly focuses on the computation of the most influential profiles on Twitter given certain topics. For instance, amongst the others (e.g., Lo 2016; Bartoletti 2016; Drakopoulos 2016; Noro 2013; Subbian and Melville 2011; Das Sarma 2010), Montangelo and Furini (2015) proposed a method based on the algorithm TRank that connects the user's activity, i.e., tweets, and the profile itself in order to reveal the user's level of influence; Cappelletti and Sastry (2012) set out a technique based on the IARank ranking model that orders the information about Twitter users in a real-time span, the logic behind it is to compute the average of users' influence by taking into account the retweet and mentions as information amplification sources being both the features proving how a user is likely to be retweeted and mentioned; Yamaguchi (2010) published a work presenting an algorithm called TURank (Twitter User Rank) based on the connection existing between users and tweets, both of them represented by a user-tweet graph. For the purpose of our initial study on how the new malware occurrences are publicly shared on Twitter, alongside the support of experts in the cybersecurity field of knowledge, the platform MalwareBazaar has been chosen as a first resource from which to begin to test the extraction of Twitter users' profiles. These latter usually share information about the new malware generation through their posts on social media. This platform offers to cyber analysts statistical means through which it is possible to be informed about the latest cyber threat reported by determined users identified as Top Reporters. Therefore, the first task addressed the crawling of tweets published by the main profiles indicated as 'reporters' on the MalwareBazaar portal, which are in total 10597. The crawling executed through a custom Twitter API client gave in output a list of files, each one of them including a list of meta-data about users' activity in a tabular format. The columns within each file contain: tweet Id, Text, Name, Screen Name, Date, Favorites, Retweets, Language, Client, Tweet type (e.g., retweet, reply, tweet), URL, number of Hashtags, number of Mentions, Media type (e.g., photo) and Media URLs. The generated files are then parsed in a next step. In the extraction phase, just the column referred to the tweets' content (Text) from the crawling output has been performed. Each column extracted

has represented a separate file considered as a single document to put into the source corpus to be semantically analyzed. For instance, the tweet text column of a selected user, e.g., tolisec user, has represented a single document containing the 121 tweets published by this user. Successively, through Python, specifically with NLKT and SPACY packages, the texts have been cleaned, this step specifically addressed the removal of stopwords as well as of symbols and emoticons in order to make the documents processable for the term extraction tool, as shown in Table 1.

Text unprocessed	Text processed
🚫 #QakBot malware active again 🙄🙄🙄 https://t.co/H6WFmcUamo https://t.co/6IzKwvNwLo	#QakBot malware active again
RT @Europol : 12 suspects have been targeted in 🇷🇺 for carrying out aggressive #ransomware attacks against critical infrastructure. 🌐 1800 high-stake victims in 71 countries 🇺🇦 6 #Europol specialists deployed to Ukraine to assist @CyberpoliceUA	RT @Europol: 12 suspects have been targeted in for carrying out aggressive #ransomware attacks against critical infrastructure. 1800 high-stake victims in 71 countries 6 #Europol specialists deployed to Ukraine to assist @CyberpoliceUA

Table 1. Details of text processing of collected tweets

Figure 2 depicts the steps followed after the compilation of the source corpus and the employment of the text included in the tweets to be used as a starting point from which to begin the classification process.

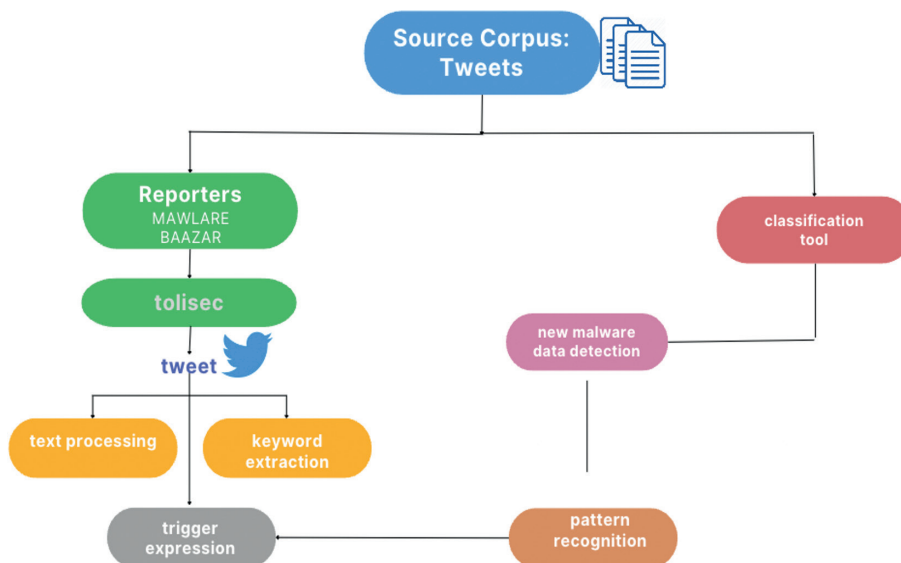


Figure 2. A comprehensive view of the processing and classification steps of the tweets<sup>1</sup>

<sup>1</sup> Tolisec is a Twitter user included in the graph as to provide a starting use-case.

## Terminological extraction

The term extraction has been realized through the software SketchEngine (Kilgarriff 2008; Jakubíček et al. 2014), a corpus analytic tool which gives in output a ranked list of the most representative terms included in the source corpus. Indeed, by using a semantic extractor it is possible to analyze the knowledge domain information under a terminological perspective and see which are the most frequent terms in the documents selected to reflect the information under study and apply reasoning techniques. For what concerns the frequency and the relevance of terms with respect to a specialized corpus, we used the Term Frequency Inverse Document Frequency (TF/IDF) (Qaiser 2018) measure. This formula allows to have in the first position terms that are very specific to the domain under study, in this case the cybersecurity one, and in the last one those most commonly used in the general language. This measure supported the identification of the most representative lexical units and the building of a network of co-occurrences that have guided the systematization of the trigger expressions. Indeed, these latter have constituted the semantic means to discover new information about malware denominations in tweets meant to represent the new entities within the classification system.

Table 2 shows an extract of the lemmas retrieved in the source documentation through the integration of a stopword list in English and the definition of the minimum frequency threshold at 1 in order to have as much terms as possible to detect.

Item	Frequency
malware	735
email	710
url	642
sample	628
com	601
new	587
exe	579
use	545
dll	531
payload	508
file	399
domain	399
report	386
zip	357
host	317
see	307
abuse	298
ransomeware	294

Table 2. Term extraction output



## Rule-based pattern recognition

The next step addresses the definition of the trigger expressions by relying on the terminological analysis, followed by their normalization through a rule-based pattern recognition (Babicm 2008; Anicic 2010). In detail, the objective of this phase is strictly connected to the one of realizing a classification tool on malware because it is aimed at establishing a set of rules to be applied over the corpus made up of tweets to retrieve new information about malware to be included into the classification tool, i.e., ontology, through the terms in the list obtained by the extraction (relying on the TF/IDF scores). The process continued by checking the co-occurrences identified within the terms in the source corpus. In detail, each term in SketchEngine can be analyzed according to the concordance terms show in the semantic distribution thanks to the syntactic connective structures within the source corpus through the Word Sketch function.

For instance, according to the output given by the semantic tool, the first two terms identified to generate a first list of regular expressions, alongside the support of domain experts, have been malware and ransomware. Therefore, the collocations represented the key combinations that could give back a networked kind of knowledge information through which to detect new names or to empower the cyber attack classification. A list of few collocations used in this preliminary stage is the following:

- Modifiers of malware: infect victims with – using a, the most resilient, the most dangerous/interesting malware
- Malware + verbs: attack (malware attacking), call (malware (also) called), identify (malware identified as)

These collocations have been used as trigger expressions to run the automatic identification of unknown malware and information about them meant to be included in the classification tool. Each of these expressions has been transformed into a syntactic structure following the rule-based pattern recognition (Xu and Cai 2021). In this way the automatic extractor tool is able to detect the morpho-syntactic structures to be mapped with the new information to be retrieved and populate the malware classifier in the form of ontology. Once understood that terms are accompanied by expressions that can lead to knowledge-domain discovery (Ervert 2008), the following phase of this research activity covered the creation of the trigger expressions to be implemented in the tool in order to represent the training set for the alert on new malware denominations sorted by time. The definition of the rule-based pattern recognition system is based on the exploitation of the regular expressions, created with the use of SpaCy library (Vasiliev 2020) in Python. This configuration has been based on the discovery of new denominations of malware starting from the information contained in the documents (the tweets of the users selected through the Malware Bazaar platform) interpreted under the lens of clues to be considered as alerts of new data to import. In contrast to the fixed pattern matching of regular expressions, this method allows us to match tokens according to some pre-set patterns. Additionally, it includes features such as parts-of-speech analysis, entity types, dependency parsing, lemmatization, and a great deal more. In addition to this, this further bolsters regular expression patterns. The token pattern matcher provided by the SpaCy library takes advantage of the word level features proper to this linguistic toolkit such as LOWER, LENGTH, LEMMA, and SHAPE as well as flags, such as IS PUNCT, IS DIGIT, LIKE URL. An input text may be given and rules can be defined in order to parse

the text and determine whether or not it includes the appropriate morpho-syntactic objects in the appropriate sequence. In order to provide some practical examples of results obtained by our method, we present the outputs from the execution of the first group included in the list of regular expressions presented in Section 2.4. The sentence “infect victims with... malware” represents the starting point from which to discover new information about the malware being discussed. The resulting information retrieved by executing the tool over the corpus will be the key entity to be integrated in the classification system. In detail, regarding the case of the aforementioned sentence, the pattern-base code instructs the SpaCy library to recognize sentences that begin with one or more verbs, followed by one or more nouns, and by one or more prepositions or postpositions, possibly a determiner, the “malware” lemma, and then end with a noun that will be identified as the new potential malware name in this case. This enables us to identify sentences that contain different words but are constructed in the same syntactic manner.

## Results

As a consequence of our work, various instances of patterns as well as their representation in the form of spacey regexes are hereby presented. These regexes have the potential to be utilized in the filtering of pre-processed and normalized tweets. The following short list represents an overview of the source expressions, selected under the basis of the terminological analysis outcomes, used to construct the patterns to be employed over the crawled tweets for the detection of new malware-related data.

Expression	RegEx
(malware) infect victims with...	<i>NOUN+ VERB (infect) + NOUN+ADP+NOUN</i>
(malware) known as... / (malware) formerly known as...	<i>NOUN+ VERB ADV+VERB (know) +ADP (as) +PROPN</i>
identified as...	<i>VERB (Identify)+ADP</i>
encrypted by	<i>VERB (encrypt) + ADP + NOUN</i>

## Classification tool

In the literature many malware classification schemes have been configured, such as the Common Taxonomy for law Enforcement and the National Network of CSIRTs published by the Europol Public Information, which describes a range of incidents according to their class and type and then support the “mapping each type of incident with the pertinent article of the international legislative framework” (Europol Public Information 2017, 5); MISP Taxonomies ; CIRCL taxonomy schemes of classification in Incident response and detection ; OSINT Open Source Intelligence; The VERIS Framework, Vocabulary for event recording and Incident sharing; Kaspersky which reports the types of malware by behaviors.

Our proposal relies on the construction of an ontology structure as a classification tool starting from the entities discovered by executing the pattern-based approach to the tweets’ contents. Ontology is a: “[...] hierarchically structured set of concepts describing a specific domain of knowledge that can be used to create a knowledge base.” (Blomqvist and Sandkuhl 2005, 1)

With reference to the meaning of ontology in the informatics area, Gaurino (2009, 2) gives a clear definition by stating that “Computational ontologies are a means to formally model the structure of a system, i.e., the relevant entities and relations that emerge from its observation, and which are useful to our purposes. An example of such a system can be a company with all its employees and their interrelationships. The ontology engineer analyzes relevant entities and organizes them into concepts and relations, being represented, respectively, by unary and binary predicates. The backbone of an ontology consists of a generalization/specialization hierarchy of concepts, i.e., a taxonomy.”

An interesting study, specifically oriented to the cybercrime field using ontologies, has been reported by Donalds and Osei-Bryson (2019) who, besides offering an extensive overview of the cybercrime classification schemes existing in the literature, describe in a practical way the realization of a high-level ontology for the cybercrime events, specifically called cybercrime classification ontology (CCO) through the use of Protégé platform (Sivakumar and Arivoli 2011) as our study will do. The author starts by isolating the main cybercrime-related concepts (i.e., attack event, vulnerability, tool and technique, objective, offence, location, complainant, victim, target, impact, attacker) and continues by enhancing the parent-child relationships through the use of the object properties which help in improving the attack events classification.

Our work regarded the construction of the classification scheme on the basis of tweets’ analysis where the structure of the ontology follows the hierarchical configuration of classes and sub-classes and the association of each new malware discovered as an individual, as the following example demonstrates: Cyber\_attacks hasSubclass Malware; Malware hasSubclasses Zombie, Crypro\_miner\_malware, Trapdoor, Trojan\_horse, Banking\_malware, Virus, Logic Bomb, Worm, Ransomware. In this regard, we targeted the inclusion of new denominations of malware in the classification scheme as the extreme leaf of this tree-like configuration, hence the matching will be between, for instance, Banking\_malware hasIndividual: new name of worm. This process has been executed by running the semantic analysis over the tweets and confirmed by cyber defense experts as well as by relying on the information included in the aforementioned malware classification tools.

Indeed, through the implementation of the regular expressions we have obtained encouraging results, some of which are the following:

pattern	tweet	pattern retrieved	new information	Ontology
<i>NOUN+ VERB- BLADV+VERB (know) +ADP (as) +PROP</i>	@msftsecintel: we have detected and are <b>now</b> blocking a <b>new family</b> of ransomware being used after an initial compromise of unpatched on-premises exchange servers. microsoft protects against this threat known as ransom:win32/ doejocrypt.a, and also as dearcry.	threat known as ransomwin32	ransomwin32	<i>Class: Cyber_attacks Sub-Class: Malware Sub-sub-Class: Ransomware Individual: Ransomwin32</i>
<i>NOUN+ VERB- BLADV+VERB (know) +ADP (as) +PROP</i>	rt @craiu : while looking at the #solarwinds #sunburst backdoor, <b>we discovered</b> several features that overlap with a previously identified backdoor known as kazuar, used by turla. our analysis: <a href="https://t.co/3ef6y-2g5ly">https://t.co/3ef6y-2g5ly</a> #unc2452 #darkhalo	backdoor known as kazuar	kazuar	<i>Class: Cyber_attacks Sub-Class: Backdoor Individual: Kazuar</i>
<i>NOUN+ VERB (infect) + NOUN+A- DP+NOUN</i>	malspam uses national bank of argentina (banco de la nación argentina) as a lure to infected user with #nanocore rat exe: <a href="https://t.co/i3pusmwnws">https://t.co/i3pusmwnws</a> nanocore rat c2:185.140.53.11:6532 inetnum: 185.140.53.0 - 185.140.53.255 netname: freedom_of_speech_vpn <a href="https://t.co/ng7yee4ajx">https://t.co/ng7yee4ajx</a>	malspam infect user with #nanocore rat	nanocore rat	<i>Class: Cyber_attacks Sub-Class: Malware Sub-sub-Class: Spam seeAlso Malspam<sup>2</sup></i>  <i>Class: Cyber_attacks Sub-Class: Malware Sub-sub-Class: Trojan Individual: Nanocore rat</i>  Object Property: Malspam USES Nanocore rat
<i>VERB (encrypt) + ADP + NOUN</i>	@govcert_ch : don't open fax messages pretending to come per email within a zip file, or you will have your data encrypted by ransomware family sodinokibi (e.g. email from hellofax with subject "sie haben ein fax"). <a href="https://t.co/ai6unad2bv">https://t.co/ai6unad2bv</a>	encrypted by ransomware family sodinokibi	sodinokibi	<i>Class: Cyber_attacks Sub-Class: Malware Sub-sub-Class: Ransomware Individual: Sodinokibi</i>  Object Property: Data EncryptedBy Sodinokibi
<i>VERB (identify)+ADP</i>	@cyb3rops @n0le_ptr your yara rule <b>hits</b> on 10 samples on malware bazaar by @abuse_ch all identified as blackcat before :+1: : <a href="https://t.co/bl6duhhuwi">https://t.co/bl6duhhuwi</a>	10 samples identified as blackcat	blackcat	<i>Class: Cyber_attacks Sub-Class: Malware Sub-sub-Class: Ransomware Individual: Blackcat</i>

<sup>2</sup> SeeAlso stands for the synonymy relationship.

s demonstrated in these examples another element that can contribute to the retrieval of novelty trait can be the attention to some verbs, e.g., discover or hit, adjectives, e.g., new, or adverbs such as ‘now’ held up by gerundive constructions (‘are now blocking’), which can lead to the inference of current data meant to be included in the ontology. The ontology offers a comprehensive way to represent the classified information by structuring it according to the generic-specific principle and associating to each class a set of instances, in our case study the new malware denominations, as Figure 3 depicts.

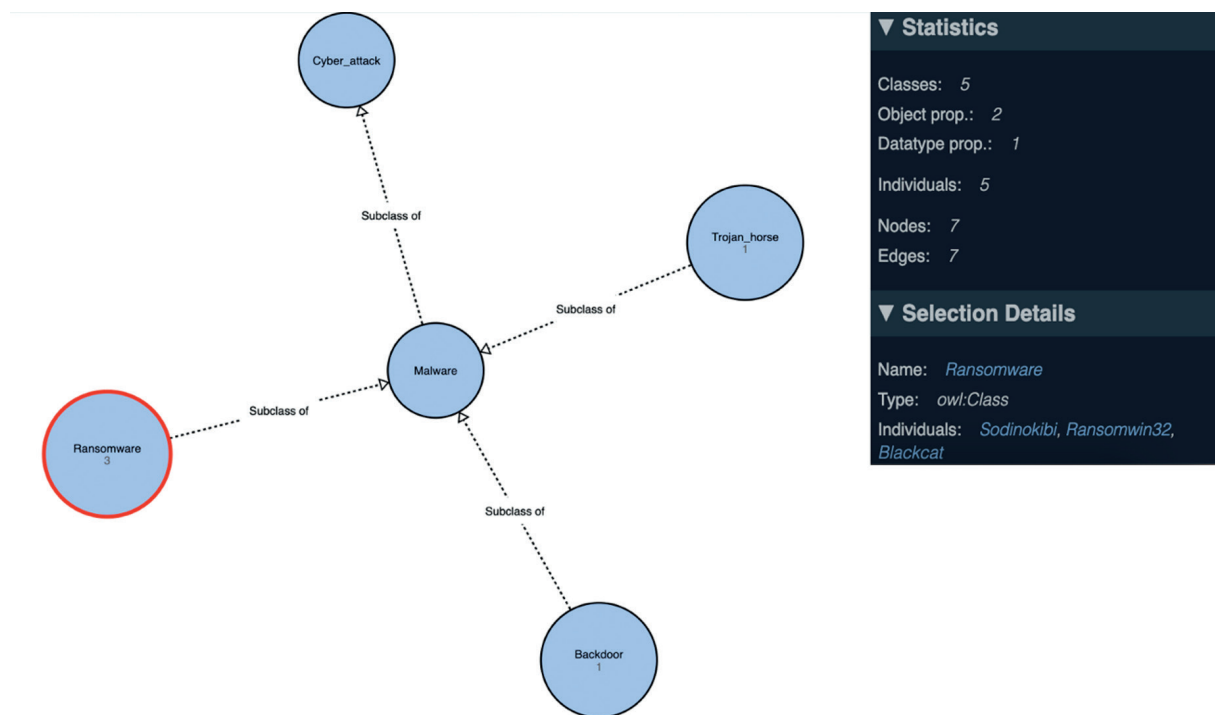


Figure 3 Ontology structure for malware instances<sup>3</sup>

The object properties expressed through OWL language relate pairs of entities (Glim 2014) and are the means of organizing the informative data by specific connections that explicit the conceptualization of the knowledge base. This will be the focus of our next research activity which will be performed using the additional data we crawled from Twitter.

## Conclusion

This paper develops a new method to predict malware appearance through the analysis of tweets shared by active Twitter users identified in the MalwareBazaar database and spreading information on new malware instances. In this work we conducted a semantic analysis of the isolated users' tweets after a crawling operation and the configuration of a set of identified trigger expressions, normalized in the form of regular expressions, that are used as a knowledge inference task

<sup>3</sup> Realized with WebVowl platform (Lohmann 2015) <https://service.tib.eu/webvowl/>



to tune data. The establishment of the trigger expressions followed a terminological selection of terms to be used as morphosyntactic units within the regular expression rules. We then exploited the entities retrieved by the trigger expressions over the tweets dataset to be used to define a classification tool. The classifier is considered to represent the connections of new malware, detected by implementing the above-mentioned steps, show within a hierarchical structure.

One of the future perspectives will address the continuous enhancement of the malware classifier (ontology instances) which will be also fed by the CSIRT *Settimana cibernetica* enabling the mapping with the new attacks included in the MITRE official framework of CAPEC and the associated vulnerabilities. This triangular interconnection could support the association of new types of malwares with existing networked semantic flow of information related to the vulnerabilities present in the hardware, software or protocols infrastructures. This activity could represent a forecastable knowledge platform to be used by companies when it comes to considering the elements meant to be analyzed to reduce the risk of being exposed to cyberthreats.

## References

- Adem, Tahir, and Muhammed Mutlu Yapici. 2022. "A Novel Malware Classification and Augmentation Model Based on Convolutional Neural Network." *Computers & Security* 112. <https://doi.org/10.1016/j.cose.2021.102515>.
- Akhtar, Muhammad Shoaib, and Tao Feng. 2022. "Malware Analysis and Detection Using Machine Learning Algorithms." *Symmetry* 14 (11): 2304.
- Andrei, Brazhuk. 2019. "Semantic Model of Attacks and Vulnerabilities Based on CAPEC and CWE Dictionaries." *International Journal of Open Information Technologies* 7 (3): 38-41.
- Anicic, Darko, Paul Fodor, Sebastian Rudolph, Roland Stühmer, Nenad Stojanovic, and Rudi Studer. 2010. "A Rule-Based Language for Complex Event Processing and Reasoning." In *Web Reasoning and Rule Systems. RR 2010. Lecture Notes in Computer Science*, edited by Pascal Hitzler, and Thomas Lukasiewicz, vol 6333, 4: 42–57. Berlin, Heidelberg: Springer. [https://doi.org/10.1007/978-3-642-15918-3\\_5](https://doi.org/10.1007/978-3-642-15918-3_5).
- Annachhatre, Chinmayee, Thomas H. Austin, and Mark Stamp. 2015. "Hidden Markov Models for Malware Classification." *Journal of Computer Virology and Hacking Techniques* 11: 59–73. <https://doi.org/10.1007/s11416-014-0215-x>.
- Antoniou, G., van Harmelen, F. (2004). "Web Ontology Language: OWL". In *Handbook on Ontologies. International Handbooks on Information Systems*, edited by Steffen Staab, and Rudi Studer. Berlin, Heidelberg: Springer. [https://doi.org/10.1007/978-3-540-24750-0\\_4](https://doi.org/10.1007/978-3-540-24750-0_4).
- Arora, Monika, and Vineet Kansal. 2019. "Character Level Embedding with Deep Convolutional Neural Network for Text Normalization of Unstructured Data for Twitter Sentiment Analysis." *Social Network Analysis and Mining* 9: 12. <https://doi.org/10.1007/s13278-019-0557-y>.
- Auger, Alain, and Caroline Barrière. 2008. "Pattern-based Approaches to Semantic Relation Extraction: A State-of-the-Art." *Terminology* 14 (1). <https://doi.org/10.1075/term.14.1.02aug>.
- Babic, Bojan, Nenad Nestic, and Zoran Miljkovic. 2008. "A Review of Automated Feature Recognition with Rule-based Pattern Recognition." *Computers in Industry* 59 (4): 321–337.
- Akshat Bakliwal, Piyush Arora, Senthil Madhappan, Nikhil Kapre, Mukesh Singh, and Vasudeva Varma. 2012. "Mining Sentiments from Tweets." In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, 11–18. Jeju, Korea: Association for Computational Linguistics.
- Barnard, Josie. 2016. "Tweets as Microfiction: On Twitter's Live Nature and 140-Character Limit as Tools for Developing Storytelling Skills." *New Writing* 13 (1): 3–16. <https://doi.org/10.1080/14790726.2015.1127975>.
- Bartoletti, Massimo, Stefano Lande, and Alessandro Massa. 2016. "Faderank: An Incremental Algorithm for Ranking Twitter Users." In *Web Information Systems Engineering–WISE 2016: 17th International Conference, Shanghai, China, Proceedings, Part II* 17, 55–69. Springer International Publishing.
- Blomqvist, Eva, and Kurt Sandkuhl. 2005. "Patterns in Ontology Engineering: Classification of Ontology Patterns." *ICEIS* 3: 413–416.

- Brazhuk, Andrei. 2019. "Semantic Model of Attacks and Vulnerabilities Based on CAPEC and CWE Dictionaries." *International Journal of Open Information Technologies* 7(3): 38–41.
- Cappelletti Rafael, and Sastry Nishanth. 2012. "IARank: Ranking Users on Twitter in Near Real-Time, Based on Their Information Amplification Potential." *International Conference on Social Informatics*, 70–77. Alexandria, VA, USA. <https://doi.org/10.1109/SocialInformatics.2012.82>.
- Christodorescu, Mihai, Sanjit Jha, Sanjit A. Seshia, Dawn Song, and Randal E Bryant. 2005. "Semantics-Aware Malware Detection." *IEEE Symposium on Security and Privacy (S&P'05)*, Oakland, CA, USA, 2005, 32–46. <https://doi.org/10.1109/SP.2005.2032-46>.
- Concone, Mário. 2012. "Twitter Event Detection: Combining Wavelet Analysis and Topic Inference Summarization." *DSIE'12, Doctoral Symposium on Informatics Engineering*, 1: 11–16.
- Das Sarma, Anish, Atish Das Sarma, Sreenivas Gollapudi, and Rina Panigrahy. 2010. "Ranking Mechanisms in Twitter-Like Forums." In *Proceedings of the Third ACM International Conference on Web Search and Data Mining WSDM'10, 21–30, February 4-6*. New York City, New York, USA: Association for Computer Machinery.
- Das, Tushar Kant, and P. Mohan Kumar. 2013. "BIG Data Analytics: A Framework for Unstructured Data Analysis." *International Journal of Engineering and Technology* 5: 153–156.
- Donalds, Charlette, and Kweku-Muata Osei-Bryson. 2019. "Toward a Cybercrime Classification Ontology: A Knowledge-Based Approach." *Computers in Human Behavior* 92: 403–418.
- Drakopoulos, Georgios, Andreas Kanavos, and Athanasios K Tsakalidis. 2016. "Evaluating Twitter Influence Ranking with System Theory." *WEBIST* 1: 113–120.
- Europol Public Information. 2017. "Common Taxonomy for Law Enforcement and The National Network of CSIRTs." [https://www.europol.europa.eu/cms/sites/default/files/documents/common\\_taxonomy\\_for\\_law\\_enforcement\\_and\\_csirts\\_v1.3.pdf](https://www.europol.europa.eu/cms/sites/default/files/documents/common_taxonomy_for_law_enforcement_and_csirts_v1.3.pdf).
- Evert, Stefan. 2008. "Corpora and Collocations." In *Corpus Linguistics: an international handbook* 2, 1212–1248. Berlin, New York: De Gruyter Mouton.
- Gaglio, Salvatore, Giuseppe Lo Re, and Marco Morana. 2016. "A Framework for Real-Time Twitter Data Analysis." *Computer Communications* 73: 236–242.
- Georgiadou, Anna, Spiros Mouzakis, and Dimitris Askounis. 2021. "Assessing MITRE AT-T&CK Risk Using a Cyber-Security Culture Framework." *Sensors* 21(9): 3267.
- Glimm, Birte, Ian Horrocks, Boris Motik, Rob Shearer, and Giorgos Stoilos. 2012. "A Novel Approach to Ontology Classification." *Journal of Web Semantics* 14: 84–101.
- Guarino, Nicola, Daniel Oberle, and Steffen Staab. 2009. "What Is an Ontology?." *Handbook on Ontologies* 1–17. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-540-92673-3>.
- Gupta, Rishabh, and Rajesh N Rao. 2020. "Towards Semantic Noise Cleansing of Categorical Data Based on Semantic Infusion." <https://doi.org/10.48550/arXiv.2002.02238>.
- Gutierrez, Carlos Enrique, Mohammad Reza Alsharif, Katsumi Yamashita, and Mahdi Khosravy. 2014. "A Tweets Mining Approach to Detection of Critical Events Characteristics Using Random Forest." *Int J Next-Gener Comput* 5(2): 167–176.

- Habibi, Omar, Mohammed Chemmakha, and Mohamed Lazaar. 2023. "Performance Evaluation of CNN and Pre-trained Models for Malware Classification." *Arabian Journal for Science and Engineering*: 1–15.
- Huang, Hsien-Der, Tsung-Yen Chuang, Yi-Lang Tsai, and Chang-Shing Lee. 2010. "Ontology-based Intelligent System for Malware Behavioral Analysis." In *International Conference on Fuzzy Systems*, 1–6, Barcelona, Spain. doi: [10.1109/FUZZY.2010.5584325](https://doi.org/10.1109/FUZZY.2010.5584325).
- Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý, and Vít Suchomel. 2014. "Finding Terms in Corpora for Many Languages with the Sketch Engine." In *Proceedings of the demonstrations at the 14th conference of the european chapter of the association for computational linguistics*, 56-56. Gothenburg, Sweden: Association for Computational Linguistics. <https://doi.org/10.3115/v1/E14-2014>.
- Kang, Boojoong, KimTaekeun, Heejun Kwon, Yangseo Choi, and Eul Gyu Im. 2012. "Malware Classification Method via Binary Content Comparison." In *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, 316–321, New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/2401603.2401672>.
- Kalash, Mahmoud, Mrigank Rochan, Noman Mohammed, Neil D.B. Bruce, Yang Wang, and Farkhund Iqbal. 2018. "Malware Classification with Deep Convolutional Neural Networks." In *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 1–5. Paris, France. <https://doi.org/10.1109/NTMS.2018.8328749>.
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž, and David Tugwell. 2008. "The Sketch Engine." *Practical lexicography: a reader*: 297–306.
- Kinable, Joris, and Orestis Kostakis. 2011. "Malware classification based on call graph clustering." *Journal in Computer Virology* 7(4): 233–245. <https://doi.org/10.1007/s11416-011-0151-y>.
- Kotenko, Igor, and Elena Doynikova. 2015. "The CAPEC based generator of attack scenarios for network security evaluation." In *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 436–441. Warsaw, Poland. <https://doi.org/10.1109/IDAACS.2015.7340774>.
- Kwon, Roger, Ashley Travis, Jerry Castleberry, Penny Mckenzie, and Sri Nikhil Gupta Gouriseti. 2020. "Cyber Threat Dictionary Using MITRE ATT&CK Matrix and NIST Cybersecurity Framework Mapping." *Resilience Week (RWS)*, 106–112.
- León-Araúz, Pilar, Antonio San Martín, and Pamela Faber. 2016. "Pattern-based Word Sketches for the Extraction of Semantic Relations." In *Proceedings of the 5th International workshop on Computational Terminology (Computerm2016)*, 73–82. Osaka, Japan.
- Lo, Siaw Ling, Raymond Chiong, and David Cornforth. 2016. "Ranking of High-value Social Audiences on Twitter." *Decision Support Systems* 85: 34–48.
- Lohmann, Steffen and Vincent Link, Eduard Marbach, and Stefan Negru. 2015. "WebVOWL: Web-based Visualization of Ontologies." In *Knowledge Engineering and Knowledge Management: EKAW 2014 Satellite Events, VISUAL, EKM1, and ARCOE-Logic, Linköping, Sweden, November 24-28, 2014. Revised Selected Papers*, 19: 154–158. Springer International Publishing.

- Mathews, Sherin Mary. 2019. "Explainable Artificial Intelligence Applications in NLP, Biomedical, and Malware Classification: A Literature Review." *Intelligent Computing. CompCom 2019. Advances in Intelligent Systems and Computing*, 998. Cham: Springer. [https://doi.org/10.1007/978-3-030-22868-2\\_90](https://doi.org/10.1007/978-3-030-22868-2_90).
- Mirza, Qublai K. Ali., Irfan Awan, and Muhammad Younas. 2018. "CloudIntell: An Intelligent Malware Detection System." *Future Generation Computer Systems* 86: 1042–1053.
- Montangero, Manuela, and Marco Furini. 2015. "Trank: Ranking Twitter Users According to Specific Topics." In *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, 767–772. Las Vegas, NV, USA. <https://doi.org/10.1109/CCNC.2015.7158074>.
- Noro, Tomoya, Fei Ru, Feng Xiao, and Takehiro Tokuda. 2013. "Twitter User Rank Using Keyword Search." *Information Modelling and Knowledge Bases XXIV. Frontiers in Artificial Intelligence and Applications* 251: 31–48.
- Pascanu, Razvan, Jack W. Stokes, Herminie Sanossian, Mady Marinescu, and Anil Thomas. 2015. "Malware Classification with Recurrent Networks." In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1916–1920. South Brisbane, QLD, Australia. <https://doi.org/10.1109/ICASSP.2015.7178304>.
- Qaiser, Shahzad, and Ramsha Ali. 2018. "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents." *International Journal of Computer Applications* 181(1): 25–29.
- Rastogi, Nidhi, Sharmishtha Dutta, Mohammed J. Zaki, Alex Gittens, and Charu Aggarwal. 2020. "Malont: An Ontology for Malware Threat Intelligence." In *International Workshop on Deployable Machine Learning for Security Defense*, 28–44. Cham: Springer International Publishing.
- Sabottke, Carl, Octavian Suci, and Tudor Dumitras. 2015. "Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting {Real-World} Exploits." In *24th USENIX Security Symposium (USENIX Security 15)*, 1041–1056.
- Sahu, Manish Kumar, Manish Ahirwar, and A. Hemlata. 2014. "A Review of Malware Detection Based on Pattern Matching Technique." *International Journal of Computer Science and Information Technologies (IJCSIT)* 5 (1): 944–947.
- Sankaranarayanan, Jagan, Hanan Samet, Benjamin E. Teitler, Michael D. Lieberman, and Jon Sperling. 2009. "Twitterstand: News in Tweets." In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 42–51. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/1653771.1653781>.
- Singh, Jagsir, and Jaswinder Singh. 2018. "Challenge of Malware Analysis: Malware Obfuscation Techniques." *International Journal of Information Security Science* 7(3): 100–110.
- Sivakumar, Ramakrishnan, and P.V. Arivoli., 2011. "Ontology Visualization PROTÉGÉ Tools—A Review." *International Journal of Advanced Information Technology (IJAIT)* 1: 1-11. <http://dx.doi.org/10.5121/ijait.2011.1401>.
- Subbian, Karthik, and Prem Melville. 2011. "Supervised Rank Aggregation for Predicting Influencers in Twitter." In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust*



and 2011 IEEE Third International Conference on Social Computing, 661–665. Boston, MA, USA. <https://doi.org/10.1109/PASSAT/SocialCom.2011.167>.

Vasiliev, Yuli. 2020. *Natural Language Processing with Python and spaCy: a practical introduction*. San Francisco, California, USA: No Starch Press.

Tang, Yonghe, Xuyan Qi, Jing Jing, Liu Chunling, and Weiyu Dong. 2023. “BHMD: A Byte and Hex N-gram Based Malware Detection and Classification Method.” *Computers & Security* 103118.

Tekerek, Adem, and Muhammed Mutlu Yapici. 2022. “A Novel Malware Classification and Augmentation Model Based on Convolutional Neural Network.” *Computers & Security* 112: 102515.

Zareen, Syed, Padia Ankur, Tim Finin, Lisa Mathews, and Joshi Anupam. 2016. “UCO: A Unified Cybersecurity Ontology.” In *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*. Palo Alto, California, USA: AAAI Press.

Wang, Xiao Hang, D. Qing Zhang, Tao Gu, and Hung, Keng Pung. 2004. “Ontology-Based Context Modeling and Reasoning Using OWL.” In *IEEE Annual Conference on Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second*, 18–22. Orlando, FL, USA. <https://doi.org/10.1109/PERCOMW.2004.1276898>.

Xiong, Wenjun, Emeline Legrand, Oscar Åberg, and Robert Lagerström. 2022. “Cyber Security Threat Modeling Based on the MITRE Enterprise ATT&CK Matrix.” *Software and Systems Modeling* 21.1: 157–177.

Xu, Xin, and Hubo Cai. 2021. “Ontology and Rule-Based Natural Language Processing Approach for Interpreting Textual Regulations on Underground Utility Infrastructure.” *Advanced Engineering Informatics* 48, 101288.

Yamaguchi, Yuto, Tsubasa Takahashi, Toshiyuki Amagasa, and Hiroyuki Kitagawa. 2010. “Tur-ank: Twitter User Ranking Based on User-Tweet Graph Analysis.” In *Web information systems engineering—WISE 2010: 11th International Conference, Hong Kong, China, December 12-14, 2010. Proceedings*, 11, 240–253. Springer Berlin Heidelberg.

## The evolution of Data.bnf.fr: past, present and future of the BnF linked open data project

Elisa Grimaldi<sup>(a)</sup>

a) University of Florence, <https://orcid.org/0009-0001-3431-7888>

**Contact:** Elisa Grimaldi, [elisa.grimaldi@unifi.it](mailto:elisa.grimaldi@unifi.it)

**Received:** 17 October 2023; **Accepted:** 07 January 2024; **First Published:** 15 May 2024

### ABSTRACT

In 2011, the Bibliothèque nationale de France (BnF) launched the Data.bnf research and development project to meet 3 main goals: to bring together in a single portal the data produced with different formats of the catalogue and the various applications of the BnF, to increase its visibility and promote its reuse by applying the technologies of the semantic web and to experiment a new way of structuring information based on IFLA's conceptual models, thus foreshadowing the catalogue of the future.

The contribution aims to trace the innovations that are partially appreciable and partially in the process of being implemented by Data.bnf, which, while remaining stable in its primary goals, has continued to evolve to improve accessibility and functions and adapt to the needs of users.

### KEYWORDS

BnF; Data.bnf.fr; Semantic web; Linked Open Data; Bibliographic transition; IFLA-LRM.

## L'evoluzione di Data.bnf.fr: passato, presente e futuro del progetto linked open data della BnF

### ABSTRACT

Nel 2011 la Bibliothèque nationale de France (BnF) ha lanciato il progetto di ricerca e sviluppo Data.bnf per rispondere a 3 obiettivi principali: riunire in un unico portale i dati prodotti con formati diversi del catalogo e dei vari applicativi della BnF, aumentarne la visibilità e favorirne il riutilizzo applicando le tecnologie del web semantico e sperimentare un nuovo modo di strutturare l'informazione basato sui modelli concettuali dell'IFLA, prefigurando così il catalogo del futuro.

Il contributo si propone di ripercorrere le innovazioni in parte già apprezzabili e in parte in fase di realizzazione di Data.bnf, che pur rimanendo stabile nei suoi obiettivi primari ha continuato a evolversi per migliorare l'accessibilità e le funzioni e adattarsi alle esigenze degli utenti.

### PAROLE CHIAVE

BnF; Data.bnf.fr; Web semantico; Linked Open Data; Transizione bibliografica; IFLA-LRM.

L'evoluzione tecnologica ha modificato sensibilmente l'approccio alla ricerca e alla conoscenza, mettendo in secondo piano le biblioteche e le altre istituzioni culturali. Molto spesso, i primi a essere interrogati per soddisfare un bisogno informativo legato all'universo bibliografico non sono i bibliotecari, bensì i motori di ricerca.

La prima conseguenza è un calo nella frequentazione delle biblioteche e dell'utilizzo dei loro servizi, anche digitali, soprattutto perché la maggior parte dei cataloghi delle biblioteche e delle banche dati che potrebbero dare delle risposte alle domande formulate su Google rimangono nella *deep web*<sup>1</sup> e non compaiono dunque nei risultati di ricerca. Come possono le biblioteche riavvicinare a sé i lettori e recuperare il proprio ruolo di mediatori? La Bibliothèque Nationale de France (d'ora in poi BnF) ha raccolto la sfida più di un decennio fa, presentando un progetto sperimentale che ha il pregio di orientare gli utenti in un contesto informativo sempre più ampio e dispersivo, spaventosamente somigliante alla biblioteca-universo descritta da Borges<sup>2</sup>.

Dal 2011 Data.bnf è il portale open data della BnF. Il suo obiettivo principale, rimasto invariato, è quello di riunire i dati prodotti dalla BnF in un unico punto di accesso e aumentarne la visibilità. I dati del catalogo della BnF, di Gallica, di BnF archives et manuscrits in primo luogo e successivamente anche delle mostre virtuali realizzate dalla BnF<sup>3</sup> e di altri repository, sono accessibili e interrogabili da un unico portale.

Per essere raggiungibili direttamente dai motori di ricerca, dove iniziano (e spesso finiscono) le ricerche degli utenti, è cruciale come abbiamo visto essere presenti sul web. Questo è il secondo importante obiettivo di Data.bnf, recentemente rafforzato dall'accesso a un finanziamento del Ministero della cultura francese (DGMIC) e al Ministero della cultura e delle comunicazioni del Québec per migliorare la discoverability dei contenuti culturali francofoni<sup>4</sup>. Data.bnf si inserisce così in un contesto più vasto di quello di partenza, aprendosi al web semantico e ai suoi standard; in questo modo persegue non solo la disseminazione dei dati dell'istituzione ma anche lo scambio e il riuso di metadati. Per connettere e integrare i dati – non solo i propri ma anche quelli provenienti da altre fonti – sul web, la BnF fa uso dei linked open data, la tecnologia utilizzata per la realizzazione del web semantico (Guerrini 2019, 62).

L'interesse per i metodi e le tecnologie del web semantico è condiviso da diverse istituzioni culturali che stanno adattando i loro servizi al nuovo ambiente e parallelamente hanno intrapreso progetti di ricerca e sperimentazione. Nel 2011 ad esempio la Biblioteca Nacional de España (BNE) ha lanciato datos.bne.es<sup>5</sup>, servizio di dati basato sulle entità di FRBR e sui linked open data che integra le risorse della BNE e risorse selezionate provenienti da altre fonti.

---

<sup>1</sup> Il *deep web* ("web profondo") è la porzione del World Wide Web non indicizzata dai tradizionali motori di ricerca.

<sup>2</sup> "Affermano gli empi che lo sproposito è normale nella Biblioteca, e che il ragionevole (come anche la umile e pura coerenza) è un'eccezione quasi miracolosa. Parlano (lo so) della 'Biblioteca febbrile, i cui volumi precari corrono il rischio incessante di mutarsi in altri, e che tutto affermano, negano e confondono come una divinità delirante'" (Borges 2003, 74).

<sup>3</sup> <http://expositions.bnf.fr/>.

<sup>4</sup> <https://www.culture.gouv.fr/Thematiques/Industries-culturelles-et-creatives/Les-actions-en-faveur-de-la-decouvrabilite-des-contenus-culturels-francophones>.

<sup>5</sup> <https://datos.bne.es/inicio.html>.

## Data.bnf e il web semantico

Il portale è stato lanciato a luglio del 2011, con l’obiettivo di rendere interoperabili e accessibili sul web i dati della BnF provenienti da diversi cataloghi e banche dati. Le risorse della BnF sono prodotte con formati diversi, a seconda della tipologia: ad esempio le risorse del catalogo sono descritte in formato MARC, mentre le raccolte di documenti d’archivio registrate su BnF archives et manuscrits necessitano di una struttura gerarchica e utilizzano il formato XML-EAD (Simon *et. al.* 2013, 564). Data.bnf rende i dati della BnF – e quindi anche i suoi servizi – raggiungibili direttamente dai motori di ricerca adottando le tecnologie e le buone pratiche del web semantico. In particolare, Data.bnf utilizza URI HTTP (ad esempio gli identificatori ARK<sup>6</sup>), adotta processi di allineamento e raggruppamento e genera pagine comprensibili e utilizzabili sia dagli esseri umani che dalle macchine. Data.bnf si può così definire come la porta d’ingresso alle risorse della BnF, una sorta di “passerella” tra i silos altrimenti separati (Dalbin *et. al.* 2011, 50). Nel 2016, l’82% delle visite a Data.bnf provenivano dai motori di ricerca e due terzi dei visitatori si dirigevano ad altri applicativi della BnF (Bermès *et. al.* 2016, 3).

Il sito è basato su CubicWeb, una piattaforma open source che permette di estrarre dati da database diversi non prodotti con gli stessi formati e di pubblicarli come pagine HTML per l’essere umano e come RDF per la macchina (Wenz e Simon 2012, 41). È la stessa piattaforma utilizzata da France Archives<sup>7</sup> per integrare le risorse provenienti da archivi nazionali e locali. Già dal 2011 Data.bnf si inserisce nel movimento degli open data adottando la “Licence Ouverte/Open Licence” (Figura 1) di Etalab<sup>8</sup>, il dipartimento francese che coordina la politica di apertura e condivisione dei dati pubblici (open data). La licenza aperta di Etalab, raccomandata dallo Stato francese, è paragonabile a una licenza CC-by o ODC-by, in quanto consente la riproduzione e il riutilizzo dei dati richiedendo unicamente l’attribuzione della fonte (Boulet 2013, 9).



Figura 1. Logo della Licence Ouverte / Open License

Per quanto riguarda il riutilizzo dei metadati tutti gli scopi sono consentiti, anche quelli commerciali; ciò non è sempre valido per le risorse che possono essere sottoposte a restrizioni: ad esempio le risorse digitalizzate di Gallica possono essere riutilizzate liberamente, ma se a scopo commerciale l’istituzione prevede una remunerazione (Barré 2013, 148). Alla fine del 2011 i dati di

<sup>6</sup> Archival Resource Key. Per approfondire si veda (Manzoni 2022).

<sup>7</sup> <https://francearchives.gouv.fr/>.

<sup>8</sup> <https://www.etalab.gouv.fr/>.

Data.bnf sono stati integrati in data.gouv.fr, portale governativo francese, diventando i primi dati culturali a essere inclusi sulla piattaforma che mira a valorizzare i dati pubblici prodotti dall'amministrazione francese<sup>9</sup>.

Il riutilizzo dei dati è favorito dall'impiego delle ontologie del web semantico, tra cui FOAF per gli autori, SKOS per i concetti e RDA per gli elementi principali delle opere. Alcune proprietà e classi sono tuttavia espresse in un'ontologia specifica della BnF, bnf-onto<sup>10</sup>, come le proprietà BnFIdentifier per gli identificatori locali della BnF e mortPourLaFrance per lo specifico stato civile attribuito alle vittime di guerra dello Stato francese, che prolunga di ulteriori 30 anni la tutela del diritto d'autore<sup>11</sup>.

Per espandere le possibilità di recupero e riutilizzo dei dati, Data.bnf è dotato dal 2014 di uno SPARQL Endpoint, strumento che permette di effettuare delle query specifiche in SPARQL e di ottenere dei dati consultabili e manipolabili in RDF. Le potenzialità derivanti dalla rielaborazione dei dati semantici sono moltissime: come evidenzia un caso di studio proposto da Carlo Bianchini è possibile ad esempio utilizzare lo SPARQL Endpoint di Data.bnf per estrarre dei dati da manipolare in seguito con OpenRefine, strumento per la pulizia e la trasformazione di dati in altri formati, per produrre in maniera automatica i numeri della Colon Classification, seppure con alcune criticità (Bianchini 2019, 15-31).

Sono liberamente scaricabili anche i *dumps* di Data.bnf, in formato RDF/XML, NT e N3, tramite le Application Programming Interface (API) della BnF<sup>12</sup>. I dataset possono essere riutilizzati per nuovi progetti, come "Foucault fiches de lecture", che ha permesso tramite Data.bnf di indicizzare un corpus di migliaia di fogli di appunti di Michel Foucault destinati alla preparazione delle sue lezioni e dei suoi libri, digitalizzati e conservati alla BnF dal 2013<sup>13</sup>.

Inoltre, i dati sono collegati a dataset esterni, rispettando il quarto principio dei linked data, che suggerisce il collegamento con altri URI per scoprire più cose<sup>14</sup>. Sta infatti anche alle biblioteche favorire la scoperta delle risorse, riaffermando allo stesso tempo il loro ruolo di "garanti della qualità", esigenza sempre più forte nel mondo digitale (Morriello 2023, 11). Ad esempio, Data.bnf è allineato con il Virtual International Authority File (VIAF), che espone i propri dataset in linked open data.

## Uno strumento di accompagnamento della Transition bibliographique

Nel 2015 il progetto acquisisce una nuova forza, inserendosi nel contesto della nascente iniziativa Transition bibliographique<sup>15</sup>, avviata ufficialmente quell'anno dalle due agenzie bibliografiche francesi: l'Agence bibliographique de l'enseignement supérieur (ABES) e la BnF. Il programma nasce con lo scopo di facilitare la transizione dei cataloghi verso il web dei dati per migliorare

<sup>9</sup> Rapport d'activité 2011 de la BnF: [https://multimedia-ext.bnf.fr/pdf/rapport\\_2011.pdf](https://multimedia-ext.bnf.fr/pdf/rapport_2011.pdf).

<sup>10</sup> <https://data.bnf.fr/ontology/bnf-onto/>.

<sup>11</sup> Loi relative au code de la propriété intellectuelle, n. 92-597, article L 123-10, 1er juillet 1992.

<sup>12</sup> <https://api.bnf.fr/fr/dumps-de-databnffr>.

<sup>13</sup> <https://api.bnf.fr/fr/utiliser-les-donnees-de-databnffr-pour-lindexation-de-corpus-numeriques>.

<sup>14</sup> <https://www.w3.org/wiki/LinkedData>.

<sup>15</sup> <https://www.transition-bibliographique.fr/>.



l'esposizione dei propri dati. Data.bnf diventa in tal modo uno strumento di accompagnamento al passaggio dal record al dataset, realizzando in pieno il terzo obiettivo del progetto, ovvero essere un luogo dove sperimentare e prefigurare il catalogo del futuro.

Con l'adesione a FRBR prima e a IFLA-LRM (2017) poi, Data.bnf conferma la centralità dell'opera, entità spesso assente nei cataloghi delle biblioteche. Come accade nei principali siti Internet informativi da Wikipedia in poi, le informazioni su Data.bnf sono infatti organizzate intorno al concetto di opera (Wenz e Simon 2012, 40), a cui sono poi collegate le diverse manifestazioni e gli item, mostrando i benefici dei cataloghi basati sui modelli bibliografici dell'IFLA.

All'opera *René* di François-René de Chateaubriand (Figura 2 e Figura 3) sono collegate ad esempio l'edizione pubblicata da Hatier nel 2012, di cui è disponibile una copia alla BnF (ark:/12148/cb436874320), con collocazione 2014-21126 (Figura 4), e la versione digitalizzata dell'edizione Chaix (1865) delle *Oeuvres complètes*, comprendente anche il romanzo breve *René*, disponibile su Gallica<sup>16</sup> e consigliata da Data.bnf.

The screenshot shows the Data.bnf.fr interface for the work 'René' by François-René de Chateaubriand (1768-1848). The main content area includes a book cover image and the following metadata:

- Data recommends this digitized version**
- Language: **French**
- Category of work: **Textual works**
- Genre or work form: **Novel**
- Date: **1802**
- History of the work: **Récit romanesque prévu à l'origine comme un épisode des "Natchez". Première édition en compagnie d'Atala et du "Génie du christianisme", puis édition séparée en 1805**
- Field: **Littératures**
- Variants of the title: **René ou Les effets des passions : pour servir de suite à "Atala, ou Les amours de deux sauvages dans le désert" (français)  
Pevê (grec moderne (après 1453))  
Renato (portugais)**

The screenshot shows the 'Editions of René' section, indicating 37 resources in data.bnf.fr. The 'Books (36)' section is expanded, showing a list of editions with sorting and filtering options:

- Sort by
- Filter by language
- René (2012), François-René de Chateaubriand (1768-1848), Paris : Hatier , DL 2012
- René (2007), Christophe Bois, François-René de Chateaubriand (1768-1848), Paris : Hatier , DL 2007
- René (1997), François-René de Chateaubriand (1768-1848), Paris : Acamédia , 1997
- René (1991), François-René de Chateaubriand (1768-1848), Paris : Larousse , 1991
- René (1984), François-René de Chateaubriand (1768-1848), Paris : Bordas , 1984
- René (1980), François-René de Chateaubriand (1768-1848), Genève : Famot ; [La Seyne-sur-Mer] : diffusion F. Beauval , 1980

Figura 2 e Figura 3. Scheda della pagina di data.bnf.fr dedicata all'opera *René* di Chateaubriand (1802), consultata il 29/03/2024 ([https://data.bnf.fr/11966369/francois-rene\\_de\\_chateaubriand\\_rene/](https://data.bnf.fr/11966369/francois-rene_de_chateaubriand_rene/))

<sup>16</sup> <https://gallica.bnf.fr/ark:/12148/bpt6k23372s/f77>.

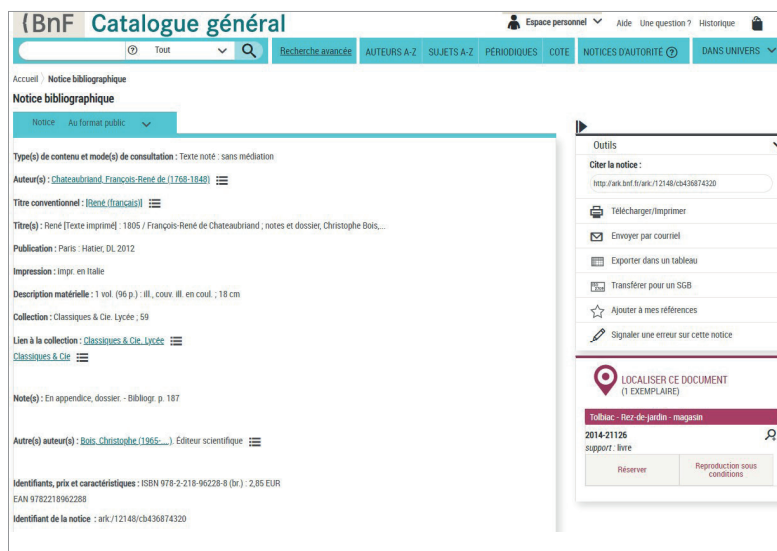


Figura 4. Notizia bibliografica linkata tra le edizioni dell'opera *René* in data.bnf.fr, consultata il 29/03/2024 (<https://catalogue.bnf.fr/ark:/12148/cb436874320>)

Nella scheda dell'opera, al centro della pagina, sono presenti i dati principali per l'identificazione dell'opera (come lingua, genere e data di pubblicazione), alcuni dei quali strutturati e navigabili: ad esempio, cliccando sulla data di pubblicazione, si possono raggiungere i dati relativi alle risorse indicizzate da Data.bnf pubblicate in quello stesso anno. Altri dati, non strutturati, completano la descrizione, come la storia (in senso filologico) dell'opera e le varianti del titolo. Spesso è proposta un'immagine dell'autore o di un'edizione dell'opera, che può rimandare a una versione digitalizzata di Gallica consigliata da Data.bnf. Questa funzionalità si chiama "Data vous conseille" e mette in evidenza le digitalizzazioni di Gallica, scelte manualmente per un corpus di opere. L'interfaccia offre a sinistra il menù navigabile delle sezioni in cui è suddivisa la pagina. Nel caso delle pagine opera, è possibile visualizzare l'elenco delle edizioni dell'opera (con link agli item posseduti dalla BnF); le risorse sull'opera (libri o altre risorse che hanno l'opera come soggetto); le persone e gli enti che hanno una responsabilità sull'opera (es. curatori, illustratori, traduttori ecc.) e i link a pagine esterne che identificano la stessa entità (es. record equivalente in Wikipedia o VIAF). A destra invece compaiono link utili per esplorare i servizi della BnF e per ripetere la ricerca in altri cataloghi e database, sia della BnF che esterni. Infine, in fondo alla pagina è possibile ottenere il permalink della pagina e condividerla, stamparla, esportarla in formato PDF o scaricare i dati in RDF o JSON.

Le pagine autori hanno una struttura simile (Figura 5). Le sezioni specifiche mostrano: le opere sulle quali l'autore o l'autrice ha una responsabilità, divise per tipologie; gli adattamenti (opere testuali, musical, spettacoli ecc. derivati dalle opere di un autore); le risorse sull'autore (libri o altre risorse che hanno la persona come soggetto); i temi collegati (ad esempio altri autori associati, luoghi ecc.); gli autori collegati e i link a pagine esterne. Le sezioni relative agli adattamenti e quella relativa ai temi associati sono state integrate nella versione di Data.bnf rilasciata a marzo 2021.

Questo tipo di pagina è evidentemente basato sull'agente di IFLA-LRM, entità "capace di azioni intenzionali, di godere di diritti e di essere ritenuta responsabile delle proprie azioni" (Riva, Le Boeuf, e Žumer 2020, 28) che raggruppa sia le persone che gli agenti collettivi.



(BnF) Data search in data.bnf.fr

### Violette Leduc (1907-1972)

*Le Temps de l'autobiographie Violette Leduc*

Country : **France**  
Language : **French**  
Gender : **Feminin**  
Birth : **Arras (Pas-de-Calais), 08-04-1907**  
Death : **Faucon (Vaucluse), 28-05-1972**  
Note : **Femme de lettres. - Prénoms complets : Violette, Thérèse, Andrée**  
Field : **Littératures**  
BnF bibliography : **Violette Leduc (1907-1972)**  
ISNI : **ISNI 0000 0001 1057 0048 (information about ISNI)**

Figura 5. Scheda della pagina di Data.bnf.fr dedicata all'autrice Violette Leduc (1907-1972), consultata il 29/03/2024.

La prima versione di Data.bnf propone circa 5000 pagine su autori e opere. Questo primo nucleo comprende gli autori e le opere citate negli indici dell'antologia della letteratura francese "Lagard et Michard", manuale di riferimento a lungo adottato nelle scuole secondarie francesi e di paesi francofoni (Boulet 2013, 9). Già alla fine del 2011 le pagine salgono a 10.000, per arrivare a 200.000 alla fine del 2012<sup>17</sup> e a 665.000 nel 2014<sup>18</sup>. Nel 2015 Data.bnf arriva a coprire il 100% dell'authority file del catalogo con 2,4 milioni di authority record e il 71% delle notizie bibliografiche con 8,5 milioni di record bibliografici collegati a quelli di autorità<sup>19</sup>. Nel 2020 più di 1 milione di pagine di opere vengono create in maniera automatica e integrate e alla fine del 2021 Data.bnf conta 4 milioni di record di autorità e 10 milioni di record bibliografici a questi collegati<sup>20</sup>. Attualmente, con la nuova versione (settembre 2023) Data.bnf raccoglie 2,5 milioni di agenti (persone ed enti) e 2 milioni di opere.

L'automazione ha svolto e svolge un ruolo fondamentale nell'espansione e nel consolidamento di Data.bnf. A permettere la crescita esponenziale dei dati forniti dal portale e quindi della visibilità dei dati prodotti dalla BnF è RobotDonnées, strumento creato dalla BnF che attraverso un algoritmo di calcolo automatico ha permesso di creare più di 1 milione di pagine di opere – il primo corpus trattato riguarda le opere a stampa del 20. secolo (Clavel 2017) – raggruppando per somiglianza le notizie bibliografiche. Se i bibliotecari avessero voluto creare manualmente le opere, avrebbero impiegato almeno 45 anni, calcolando un tempo di 2 minuti a record<sup>21</sup>. Non sempre infatti esistono nel catalogo BnF i legami tra la manifestazione descritta e il titolo uniforme: con il processo automatico sperimentato nella sandbox che di fatto è Data.bnf, le opere e i collegamenti saranno successivamente riversati nell'applicativo NOEMI (Nouer les Oeuvres, Expressions, Manifestations et Items), il nuovo strumento della BnF che permetterà di catalogare con il nuovo

<sup>17</sup> Rapport d'activité 2012 de la BnF: [https://multimedia-ext.bnf.fr/pdf/rapport\\_2012.pdf](https://multimedia-ext.bnf.fr/pdf/rapport_2012.pdf).

<sup>18</sup> Rapport d'activité 2014 de la BnF: [https://multimedia-ext.bnf.fr/pdf/rapport\\_2014.pdf](https://multimedia-ext.bnf.fr/pdf/rapport_2014.pdf).

<sup>19</sup> Rapport d'activité 2015 de la BnF: [https://multimedia-ext.bnf.fr/pdf/rapport\\_2015.pdf](https://multimedia-ext.bnf.fr/pdf/rapport_2015.pdf).

<sup>20</sup> Rapport d'activité 2015 de la BnF: [https://www.bnf.fr/sites/default/files/2023-06/rapport\\_activite\\_2022.pdf](https://www.bnf.fr/sites/default/files/2023-06/rapport_activite_2022.pdf).

<sup>21</sup> <https://data.bnf.fr/about#>.

codice di catalogazione francese RDA-FR<sup>22</sup>. Le pagine opera create con questo strumento sono riconoscibili perché presentano uno sfondo rosso e il link “automatically generated page” che rende conto del progetto sperimentale. RobotDonnées permette inoltre di correggere le duplicazioni di record di autorità.

Parallelamente, cresce anche il numero dei visitatori del portale: nel 2016 il sito è il 9° dataset più utilizzato, secondo un’indagine internazionale sui linked data, il primo non proveniente da un’area anglofona o internazionale<sup>23</sup>. Il picco è raggiunto nel 2020, con quasi 8 milioni di visitatori<sup>24</sup>, per poi scendere negli ultimi 2 anni a 5.562.739 nel 2021<sup>25</sup> e 4.980.549 nel 2022<sup>26</sup>. Gli utenti di Data.bnf non sono solo bibliotecari ed esperti del settore, ma anche un pubblico più vasto che riesce così a ottenere le risposte alle domande formulate al motore di ricerca, senza dover necessariamente iniziare la ricerca da un catalogo o da una biblioteca digitale.

Le pagine autori e opere, presenti sin dal lancio nel 2011, rappresentano tuttora il nucleo principale di Data.bnf. Non solo sono le più numerose, ma anche le più consultate, come è risultato da un’analisi dei log di Gallica e Data.bnf condotta da aprile 2016 a luglio 2017 (d’Alché-Buc *et. al.* 2017, 4).

Mentre il primo nucleo di autori e opere inizia a crescere, nel 2012 è introdotto un nuovo tipo di pagina, tema (*thème*), che integra i dati del vocabolario RAMEAU, il linguaggio di indicizzazione per soggetto utilizzato in Francia dalla BnF, dalle biblioteche accademiche e da molte biblioteche pubbliche, di ricerca e private<sup>27</sup>. Grazie all’esposizione secondo gli standard del web semantico, attraverso Data.bnf i soggetti di RAMEAU vengono riutilizzati da AbulEdu<sup>28</sup>, spazio dedicato all’insegnamento in cui la ricerca delle risorse è organizzata secondo descrittori strutturati, e da Isidore<sup>29</sup>, motore di ricerca specializzato nelle scienze umane e sociali del CNRS (Centre national de la recherche scientifique)<sup>30</sup>. Attualmente (2023) le pagine tema (Figura 6) prevedono: una scheda che indica la fonte, il campo o l’ambito (es. per “santuari”: religione) e le varianti (punti di accesso non preferiti); link interni a RAMEAU verso i termini apicali o più generali, termini più specifici e termini associati; link interni alle risorse che hanno quel termine come soggetto, link a immagini, video e film, con link al catalogo e a Gallica, se disponibile; link interni ad autori correlati al tema, divisi per tipo di responsabilità (es. autore, traduttore, curatore, illustratore, fotografo ecc.); infine link esterni ai record equivalenti in altri authority file: authority record della Biblioteca nacional de España tramite Datos.bne.es, DBpedia, authority record della Deutsche Nationalbibliothek, Library of Congress Authorities, Library of Congress Subject Headings, Nuovo Soggettario, Wikidata e Wikipedia Francophone. Grazie all’allineamento delle pagine tema su questi repertori, l’utente può ad esempio effettuare la stessa ricerca di libri sullo stesso soggetto anche nel catalogo della Library of Congress.

<sup>22</sup> <https://www.transition-bibliographique.fr/rda-fr/pdf-regles-rda-fr/>.

<sup>23</sup> Rapport d’activité 2016 de la BnF: [https://multimedia-ext.bnf.fr/pdf/rapport\\_2016.pdf](https://multimedia-ext.bnf.fr/pdf/rapport_2016.pdf).

<sup>24</sup> Rapport d’activité 2020 de la BnF: [https://www.bnf.fr/sites/default/files/2022-05/rapport\\_2020.pdf](https://www.bnf.fr/sites/default/files/2022-05/rapport_2020.pdf).

<sup>25</sup> Rapport d’activité 2021 de la BnF: [https://www.bnf.fr/sites/default/files/2023-06/rapport\\_2021.pdf](https://www.bnf.fr/sites/default/files/2023-06/rapport_2021.pdf).

<sup>26</sup> Rapport d’activité 2022 de la BnF: [https://www.bnf.fr/sites/default/files/2023-06/rapport\\_activite\\_2022.pdf](https://www.bnf.fr/sites/default/files/2023-06/rapport_activite_2022.pdf).

<sup>27</sup> Actualité du catalogue, avril 2014: <https://multimedia-ext.bnf.fr/lettres/produits/produits34.html>.

<sup>28</sup> <http://data.abuledu.org/wp/>.

<sup>29</sup> <https://isidore.science/>.

<sup>30</sup> Actualité du catalogue, avril 2014: <https://multimedia-ext.bnf.fr/lettres/produits/produits34.html>.

The screenshot shows the 'Sanctuaire' page on data.bnf.fr. The header includes the BnF logo and a search bar. The main content area is titled 'Sanctuaire' and features a small image of a church. Below the image, there is a metadata section with the following details:

- Topic: Sanctuaire
- Source file: RAMEAU
- Field: Religion, Construction, Architecture
- Variant subject headings: Lieux de pèlerinage, Lieux saints, Pèlerinage, Lieux de

Below the metadata, there are sections for 'related to this theme' (10 resources in data.bnf.fr), 'Documents on this topic' (319 resources in data.bnf.fr: see all these resources), and a list of document types: Books (241), Pictures (56), Videos, Films (21), and Multimedia documents (1).

Figura 6. Pagina di data.bnf.fr dedicata al tema “Sanctuaire” (“Santuari”), consultata il 29/03/2024 (<https://data.bnf.fr/en/11940527/sanctuaire/>)

Oltre alle opere (*œuvres*), agli autori (*auteurs*) e ai temi (*thèmes*), le tipologie di pagine presenti attualmente in Data.bnf sono: luoghi (*lieux*), date (*dates*), spettacoli (*spectacles*) e periodici (*périodiques*).

## Il futuro di Data.bnf<sup>31</sup>

Dopo un periodo di consolidamento ed espansione durato quasi dieci anni, Data.bnf si trova oggi in una fase nuova, caratterizzata ancora dalla sperimentazione. Un ricco programma di innovazioni, in parte già apprezzabili e in parte in fase di realizzazione, è esposto nel Feuille de route 2020-2024, un documento pubblicato nel 2019 sul sito di Data.bnf che fissa alcuni obiettivi da raggiungere<sup>32</sup>.

<sup>31</sup> Ringrazio Xavier Levoine, vicecapo del progetto Data.bnf, per il confronto sulle novità introdotte da Data.bnf e sui suoi obiettivi futuri (intervista condotta online il 1° settembre 2023).

<sup>32</sup> [https://data.bnf.fr/images/feuille\\_de\\_route\\_2020-2024.pdf](https://data.bnf.fr/images/feuille_de_route_2020-2024.pdf).



Il primo obiettivo di sviluppo è il consolidamento e il perfezionamento di ciò che già è in funzione. L'interfaccia rimarrà sostanzialmente incentrata sulle pagine autori e opere, ma sarà sviluppata affinché possa ospitare tutte le entità previste da IFLA-LRM, in particolar modo le espressioni. Del cosiddetto gruppo WEMI (Works, Expressions, Manifestations, Items), le espressioni sono le uniche assenti nelle pagine. Più in generale, il sito si arricchirà di filtri e faccette al fine di rendere più agile la navigazione, ad esempio per gestire ancora meglio le liste di risorse, molto più lunghe rispetto ai primi anni di vita del progetto. Un costante monitoraggio è necessario per adattare il sito ai cambiamenti occorsi nel tempo: le ontologie RDF si sono evolute e necessitano di un aggiornamento in Data.bnf. È inoltre in corso uno studio per migliorare l'indicizzazione del sito sui motori di ricerca. Il secondo obiettivo concerne il trattamento e l'accrescimento dei dati. Come si è visto, Data.bnf integra i documenti provenienti da BnF Archives et Manuscrites (BAM), ma l'integrazione è imperfetta, soprattutto perché il modello su cui si basa il sito, IFLA-LRM, è concepito per le risorse bibliografiche e non per gli archivi. L'intenzione è dunque migliorare il trattamento di questi dati, esplorando nuove ontologie. Data.bnf continuerà a espandersi e in futuro potrebbe inglobare altri dataset, interni – quali Mandragore<sup>33</sup>, il Catalogue Collectif de France (CCfr) e la Bibliographie nationale Française – ed esterni – dati della British Library, Europeana, Paris Musées e altri ancora. Il ruolo di accompagnamento della Transition bibliographique è ribadito dal terzo obiettivo, che si propone di completare l'adattamento dei dati a IFLA-LRM. Ciò comporta innanzitutto la creazione delle espressioni a cui si è già accennato sopra, attualmente non presenti in quanto non esistenti nel catalogo. Per il momento è possibile filtrare per lingua le edizioni e recuperare dunque le espressioni delle risorse testuali, ma le “vere” espressioni saranno integrate dal momento in cui saranno prodotte nativamente con NOEMI. In quanto parte della Transition bibliographique, Data.bnf seguirà l'evoluzione di RDA-FR, che utilizzerà come formato InterMarc-NG (InterMarc Nouvelle Génération)<sup>34</sup>. In particolare, è il gruppo “RDA en France” a occuparsi di adattare RDA alle pratiche e alle specificità francesi.

Infine, il quarto e ultimo obiettivo è incentrato sui bisogni degli utenti. Nel 2024 l'interfaccia cambierà ancora per migliorare l'accessibilità e l'ergonomia del sito. Sono previsti dei focus group e test di usabilità, al fine di migliorare l'interfaccia e implementare nuove funzionalità e servizi basati appunto sulle necessità degli utenti.

Nell'ambito di questo obiettivo, nel 2021 è stata implementata una nuova funzionalità per facilitare l'accesso al grafo della conoscenza di Data.bnf. Si tratta di Sparnatural<sup>35</sup>, uno strumento che consente di costruire *query* in un linguaggio più naturale di SPARQL, estendendo quindi la sua fruizione a un pubblico più eterogeneo, non necessariamente esperto di informatica e di linguaggio SPARQL.<sup>36</sup> Nell'esempio visibile nella Figura 7, è stato possibile chiedere di visualizzare i titoli di edizioni in lingua inglese di opere con autore Gustave Flaubert, senza dover utilizzare il complesso linguaggio SPARQL. Come si può vedere dall'esempio, Sparnatural permette di realizzare query in maniera semplice e intuitiva, scegliendo dai menù a tendina entità, attributi e relazioni. Molto diverso è l'aspetto della stessa query in SPARQL che si può visualizzare nella Figura 8.

<sup>33</sup> Mandragore è la banca dati iconografica dei manoscritti della BnF: <https://mandragore.bnf.fr/>.

<sup>34</sup> Actualités du catalogue, marzo 2021: <https://www.bnf.fr/fr/actualites-du-catalogue-no-51-mars-2021>.

<sup>35</sup> <https://data.bnf.fr/sparnatural/>.

<sup>36</sup> Actualités du catalogue, marzo 2021: <https://www.bnf.fr/fr/actualites-du-catalogue-no-54-mars-2022>.

The image shows a three-part interface for SPARNATURAL DATA.BNF.FR. The top part is a query builder with a visual graph showing relationships between 'Edition', 'Work', 'Person', and 'Language'. The middle part displays the generated SPARQL query. The bottom part shows a table with 8 results, including URIs and titles like 'Madame Bovary' and 'L'éducation sentimentale'.

**SPARNATURAL DATA.BNF.FR** Lang. ▾

Load example query... Querying <https://data.bnf.fr/sparql>

Graph components:  
- Edition (document icon) → edition of → Work (lightbulb icon)  
- Where: Work → author → Person → Gustave Flaubert  
- And: Work → title → Title  
- And: Edition → language → Language → eng

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX dcterms: <http://purl.org/dc/terms/>
4 PREFIX rdarelationshps: <http://rdvocab.info/RDARelationshipsWEMI/>
5 SELECT DISTINCT ?this ?title_3 WHERE {
6   ?this rdf:type <http://rdvocab.info/uri/schema/FR8RentitiesRDA/Manifestation>.
7   ?this <http://rdvocab.info/RDARelationshipsWEMI/work/Manifested> ?Work_1.
8   ?Work_1 rdf:type <http://rdvocab.info/uri/schema/FR8RentitiesRDA/Work>.
9   ?Work_1 <http://purl.org/dc/terms/creator> <http://data.bnf.fr/ark:/12148/cb11902894q#about>.
10  ?Work_1 <http://purl.org/dc/terms/title> ?title_3.
11  ?this rdarelationshps:expressionManifested/dcterms:language <http://id.loc.gov/vocabulary/iso639-2/eng>.
12 }
13 LIMIT 1000
```

Table: 8 results in 0.122 seconds. Simple view Ellipse Filter query results Page size: 50

Could not render results with the table plugin, the results currently are rendered with the Table plugin.

this	Title_3
<http://data.bnf.fr/ark:/12148/cb32110269j#about>	"Madame Bovary"@fr
<http://data.bnf.fr/ark:/12148/cb41944218k#about>	"Madame Bovary"@fr
<http://data.bnf.fr/ark:/12148/cb321102686#about>	"Madame Bovary"@fr
<http://data.bnf.fr/ark:/12148/cb34869667n#about>	"Madame Bovary"@fr
<http://data.bnf.fr/ark:/12148/cb32110270r#about>	"Madame Bovary"@fr
<http://data.bnf.fr/ark:/12148/cb352162358#about>	"L'éducation sentimentale"@fr
<http://data.bnf.fr/ark:/12148/cb304404093#about>	"Salammô"@fr
<http://data.bnf.fr/ark:/12148/cb30440408#about>	"Salammô"@fr

Showing 1 to 8 of 8 entries

Figura 7, Figura 8 e Figura 9. Query effettuata su Sparnatural il 29/03/2024 (<https://data.bnf.fr/sparnatural/?lang=en>)

Sparnatural è stato sviluppato nel 2018 dalla società Sparna nell'ambito di OpenArcheo<sup>37</sup>, progetto nato allo scopo di rendere interoperabili i dati della comunità archeologica e diffondere i diversi dataset. Si può collegare a qualsiasi SPARQL endpoint esistente e la sua natura agnostica lo rende adattabile e configurabile per qualsiasi modello. La versione appena mostrata di Sparnatural è stata sviluppata all'interno di un progetto per la diffusione dei dati culturali, finanziato dal Ministero della cultura in partenariato con la BnF e gli Archivi nazionali di Francia, che ha condotto alla creazione di due diversi editor<sup>38</sup>.

Quale potrebbe dunque essere il futuro di Data.bnf? In occasione del primo anniversario del progetto, Romain Wenz<sup>39</sup> e Agnès Simon si pongono già la questione:

Più a lungo termine, data.bnf.fr solleva la questione del divenire di un progetto di “ricerca e sviluppo”. Deve conservare un aspetto sperimentale, per definizione incerto? Essendo il web versatile per natura, il sito è spinto a evolversi continuamente? (Wenz e Simon 2012, 43).

Per rispondere a questa domanda, bisognerebbe innanzitutto riflettere sulla natura di Data.bnf. È un catalogo? Un discovery tool? Un'enciclopedia? Non è semplice dare una definizione univoca, perché Data.bnf è in parte tutto questo, ma non totalmente e non esclusivamente.

Il catalogo della BnF e Data.bnf non coincidono esattamente: progressivamente gran parte dei contenuti del catalogo sono stati riversati in Data.bnf, che però indicizza molte altre risorse, interne ed esterne. Data.bnf non vuole sostituire il catalogo (e neanche gli altri strumenti di accesso ai dati della BnF), bensì continuare a fare da “snodo” tra l'utente, il catalogo, altrimenti invisibile, e gli altri applicativi (Levoïn 2021, 28). Il progetto è nato *anche* come uno spazio di sperimentazione sul catalogo, sperimentazione che però non è effettuata direttamente *nel* catalogo e che per il momento non lo ha modificato. In questo risiede forse la forza più grande del progetto: Data.bnf non è un catalogo ma “lascia percepire come potrebbe essere il catalogo del futuro” (Guerrini 2022, 66), ovvero uno spazio basato non più sui record ma su insiemi di dati dove navigare tra opere, autori, soggetti e altre entità collegate tra loro da relazioni bibliografiche e non dai criteri di Google e dei motori di ricerca (Lapôtre 2017). Il catalogo e Data.bnf rimangono e rimarranno probabilmente due strumenti separati, in attesa del nuovo assetto che vedrà la produzione di dataset strutturati nativamente per il web semantico e nativamente basati su IFLA-LRM e la conversione retrospettiva del catalogo. In questo sarà di grande utilità l'esperienza di Data.bnf che metterà a disposizione tutti gli strumenti sviluppati per la creazione e la gestione dei metadati.

Esistono dei punti in comune tra Data.bnf e i discovery tool: l'obiettivo primario del portale è in effetti fornire un punto d'accesso unico alle risorse della BnF, senza dover ripetere le ricerche nei vari silos. A differenza dei discovery tool che danno accesso a ciò che la biblioteca ha acquisito (libri, riviste, materiali audiovisivi ecc.) o a cui dà l'accesso (abbonamenti a periodici elettronici, banche dati, piattaforme di editori ecc.), Data.bnf dà accesso anche a dataset prodotti da altre istituzioni o provenienti da progetti collaborativi. Ad esempio, nella sezione “Vedi anche” della

<sup>37</sup> <http://openarchaeo.huma-num.fr/explorateur/home>.

<sup>38</sup> <https://sparna-git.github.io/sparnatural-demonstrateur-an/presentation-en.html>.

<sup>39</sup> Romain Wenz è stato a capo del progetto Data.bnf.fr dal 2009 al 2013.

pagina dedicata all'autore Victor Hugo<sup>40</sup>, sono presenti una decina di link a record equivalenti, quali l'authority record della Library of Congress, Wikidata, ISNI, VIAF e MusicBrainz.

Data.bnf potrebbe apparire come una grande enciclopedia digitale, simile nell'aspetto a Wikipedia o ad altri progetti collaborativi della famiglia Wikimedia. Data.bnf ha peraltro in comune con Wikipedia lo sfruttamento delle tecnologie del web semantico, che hanno la forza di far emergere entrambi dal *mare magnum* delle ricerche online. Senza negare la rilevanza di questi progetti, che vedono tra l'altro una frequente e fruttuosa collaborazione con biblioteche, musei e altre istituzioni culturali, è d'obbligo precisare almeno due differenze. La prima concerne l'eshaustività, alla quale ci si aspetta che tenda un'enciclopedia e alla quale Data.bnf non aspira: il portale è alimentato dai cataloghi e depositi della BnF e questa è l'unica copertura che persegue. Potremmo quindi non trovare in Data.bnf una pagina dedicata a un determinato autore, se la BnF non possiede un libro, un manoscritto o un qualsiasi dato a suo riguardo. O ancora, in una pagina esistente non troviamo tutto ciò che riguarda quell'autore o quell'opera, se non integriamo le informazioni con altre fonti, segnalate alla fine della pagina e tra le quali è presente anche la sopracitata Wikipedia. Per la seconda differenza bisogna addentrarsi nello spinoso tema dell'autorevolezza. Non bisogna infatti dimenticare che Wikipedia è un'enciclopedia collaborativa, basata sul contributo di volontari da tutto il mondo. Quindi, se da una parte le pagine possono crescere e arricchirsi in maniera rapida grazie al lavoro di innumerevoli persone, dall'altra il controllo sull'autorevolezza diventa più complesso. Data.bnf può invece contare sulla qualità dei dati di partenza e fornire contenuti affidabili e riutilizzabili. Queste differenze di certo non dividono i due progetti che si completano a vicenda: l'utente può navigare dalla pagina Wikipedia a quella equivalente di Data.bnf e viceversa, e scoprire grazie a entrambi nuove informazioni.

Ribadite dunque le sue funzioni di accrescimento della visibilità dei dati della BnF e di ambiente di prova per l'adattamento dei cataloghi ai modelli bibliografici, la risposta alla domanda di Wenz e Simon è: sì, Data.bnf non può che continuare a evolversi e a cambiare, nell'aspetto e nelle funzionalità. È quello che ha fatto fin dal suo esordio, integrando anno dopo anno nuove tipologie di pagine, aggiungendo filtri, sezioni, allineandosi a nuovi repository esterni e modificando la sua interfaccia che è diventata sempre più intuitiva ed ergonomica.

---

<sup>40</sup> [https://data.bnf.fr/11907966/victor\\_hugo/](https://data.bnf.fr/11907966/victor_hugo/).

## Riferimenti bibliografici

Barré, Maxime. 2013. "Data.bnf.fr. Une application open data à la BnF." *Les Cahiers du numérique* 9 (1): 147-51. <https://www.cairn.info/revue-les-cahiers-du-numerique-2013-1-page-147.htm>.

Bermès, Emmanuelle, Vincent Boulet, e Céline Leclaire. 2016. "Améliorer l'accès aux données des bibliothèques sur le web: l'exemple de data.bnf.fr." IFLA WLIC 2016 – Columbus, OH – Connections. Collaboration. Community in Session 201 – National Libraries, International Federation of Library Associations and Institutions (IFLA). <https://library.ifla.org/id/eprint/1447/1/081-bermes-fr.pdf>.

Bianchini, Carlo. 2019. "Dal web semantico all'indicizzazione per soggetto: un caso di studio su data.bnf.fr e Colon classification." In *Viaggi a bordo di una parola: scritti sull'indicizzazione semantica in onore di Alberto Cheti*, a cura di Anna Lucarelli, Alberto Petrucciani, e Elisabetta Viti, presentazione di Rosa Maiello 15-31. Roma: Associazione Italiana Biblioteche.

Borges, Jorge Luis. 2003. "La biblioteca di Babele." In *Finzioni*, 67-76. Milano: Adelphi.

Boulet, Vincent. 2022. "How to build an Identifiers' policy: the BnF use case." *JLIS.it* 13 (1): 177-184. <https://doi.org/10.4403/jlis.it-12768>.

Boulet, Vincent. 2013. "S'appuyer sur la structure des données et les liens: le format MARC comme tremplin pour le Web de données: l'exemple de data.bnf.fr." IFLA WLIC 2013 – Singapore – Future Libraries: Infinite Possibilities in Session 222 – UNIMARC Core Activity (UCA). <https://library.ifla.org/id/eprint/250/>.

Clavel, Thierry. 2017. "Métadonnées en bibliothèques: attention, travaux !" *Bulletin des bibliothèques de France (BBF)* 13. [https://bbf.enssib.fr/tour-d-horizon/metadonnees-en-bibliotheques-attention-travaux\\_68026](https://bbf.enssib.fr/tour-d-horizon/metadonnees-en-bibliotheques-attention-travaux_68026).

Dalbin, Sylvie, Emmanuelle Bermès, Antoine Isaac, Romain Wenz, Yann Nicolas, Tayeb Merabti, Anila Angjeli, Thomas Francart, Lise Rozat, Pierre-Yves Vandebussche, Bernard Vatant, Yves Raimond, e Dominique Cotte. 2011. "Approches documentaires: priorité aux contenus." *Documentaliste-Sciences de l'Information* 48 (4): 42-59. <https://doi.org/10.3917/docsi.484.0042>.

d'Alché-Buc, Florence, Valérie Beaudouin, Emmanuelle Bermès, Philippe Chevallier, Aude Le Moullec-Rieux, Adrien Nouvellet, Christophe Prieur, e François Roueff. 2017. "Analysis of Gallica and Data BnF logs and Modelling of Behaviour Patterns: Presentation of the Main Results." Bibliothèque nationale de France (Paris); Télécom ParisTech. <https://hal.science/hal-01968742v2>.

Guerrini, Mauro. 2019. "La biblioteca integrata: nuovi modelli, nuove tecniche, alcune esperienze europee e italiane." *Biblioteche oggi Trends* 5(1): 59-66. <https://doi.org/10.3302/2421-3810-201901-059-1>.

Guerrini, Mauro. 2022. "Nuovi strumenti di ricerca: data.bnf.fr e i discovery tool." In *Dalla catalogazione alla metadattazione: tracce di un percorso*, prefazione di Barbara B. Tillett, postfazione di Giovanni Bergamin, 66-68. Roma: Associazione italiana biblioteche.

Lapôtre, Raphaëlle. 2016. "Visualiser les données du catalogue." In *Vers de nouveaux catalogues*, 37-47. Paris: Éditions du Cercle de la Librairie. <https://doi.org/10.3917/elec.berme.2016.01.0037>.



- Lapôtre, Raphaëlle. 2017. "Library Metadata on the web: the example of data.bnf.fr." *JLIS.it* 8 (3): 58-70. <https://doi.org/10.4403/jlis.it-12402>.
- Levoïn, Xavier. 2021. "Data.bnf.fr: améliorer la découvrabilité des contenus culturels sur le web." *Archimag* 341: 28-29.
- Manzoni, Laura. 2022. *Identificatori*. Roma: Associazione italiana biblioteche. <https://doi.org/10.53263/9788878123496>.
- Morriello, Rossana. 2023. "Biblioteche, conoscenza, bene comune e lo stretto legame con lo sviluppo sostenibile." *Biblioteche oggi* 41 (2): 5-14. <http://dx.doi.org/10.3302/0392-8586-202302-005-1>.
- Riva, Pat, Patrick Le Boeuf, e Maya Žumer. 2020. *IFLA Library Reference Model: Un modello concettuale per le informazioni bibliografiche*. Roma: ICCU. [https://repository.ifla.org/bitstream/123456789/44/5/ifla-lrm-august-2017\\_rev201712-it.pdf](https://repository.ifla.org/bitstream/123456789/44/5/ifla-lrm-august-2017_rev201712-it.pdf).
- Simon, Agnès, Romain Wenz, Vincent Michel, e Adrien Di Mascio. 2013. "Publishing Bibliographic Records on the Web of Data: Opportunities for the BnF (French National Library)." In *The Semantic Web: Semantics and Big Data*, a cura di Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, e Sebastian Rudolph, 563-577. Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-38288-8\\_38](https://doi.org/10.1007/978-3-642-38288-8_38).
- Wenz, Romain. 2012. "Des outils automatiques pour le signalement en bibliothèque: Expérimentations autour du projet data.bnf.fr." *Bulletin des bibliothèques de France (BBF)* 57 (5): 39-43. <https://bbf.enssib.fr/consulter/bbf-2012-05-0039-008>.
- Wenz, Romain. 2013. "Linked open data for new library services: the example of data.bnf.fr." *JLIS.it* 4 (1): 403. <https://doi.org/10.4403/jlis.it-5509>.

## Reproductions of State Cultural Property: the new Ministerial Decree 108/2024

Paolo Liverani<sup>(a)</sup>

a) University of Florence, <https://orcid.org/0000-0002-7335-6695>

**Contact:** Paolo Liverani, [paolo.liverani@unifi.it](mailto:paolo.liverani@unifi.it)

**Received:** 10 April 2024; **Accepted:** 12 April 2024; **First Published:** 13 April 2024

### ABSTRACT

After a year, the Italian Ministry of Culture has revised the ministerial decree regulating the granting of images of state cultural property. The first decree had been criticised by the entire cultural world for the costs it imposed on publications, including scientific ones. The new decree has improved this aspect by granting broad exemptions, but it has maintained an overly complex system and the ‘concession’ regime, with a prior ‘ethical’ control on the use of the images, which is neither acceptable nor practically feasible.

### KEYWORDS

D.M. 108/2024; Reproductions; State Cultural Property.

## Riproduzioni dei beni culturali statali: il nuovo Decreto Ministeriale 108/2024

### ABSTRACT

Il Ministero della cultura italiano dopo un anno ha rivisto il Decreto Ministeriale che regola la concessione delle immagini dei beni culturali dello stato. Il primo decreto aveva sollevato le critiche di tutto il mondo della cultura per i costi che imponeva alle pubblicazioni, anche a quelle scientifiche. Il nuovo decreto ha migliorato questo aspetto concedendo ampie esenzioni, ma ha mantenuto un impianto troppo complesso e il regime della “concessione”, con un controllo preventivo “etico” sull’uso delle immagini, che non è né ammissibile, né praticamente attuabile.

### PAROLE CHIAVE

D.M. 108/2024; Riproduzioni; Beni culturali.

Un anno fa la pubblicazione da parte del Ministero della Cultura delle *Linee guida per la determinazione degli importi minimi dei canoni e dei corrispettivi per la concessione d'uso dei beni in consegna agli istituti e luoghi della cultura statali* (DM 11 apr. 2023, n. 161)<sup>1</sup> aveva suscitato una reazione compatta e corale fortemente critica da tutto il mondo della cultura in quanto imponeva una procedura assai complessa e pesanti pagamenti perfino per le pubblicazioni a carattere scientifico quando queste avessero dovuto utilizzare immagini dei beni culturali di proprietà statale. Avevano scritto al Ministro l'[Associazione Italiana Biblioteche](#)<sup>2</sup> assieme a una quindicina di altre sigle, le [Consulte universitarie](#) degli archeologi e degli storici dell'arte,<sup>3</sup> il [Consiglio Universitario Nazionale](#),<sup>4</sup> la Conferenza dei Rettori, l'[Accademia dei Lincei](#)<sup>5</sup> e si erano espresse numerose personalità della cultura con vasta eco nella stampa. Perfino la Corte dei Conti ([Deliberazione 20 ottobre 2023, n. 76/2023/G](#))<sup>6</sup> aveva bacchettato il Ministero rilevando come l'*open access* sia un moltiplicatore di ricchezza con positive ricadute sul PIL. La Corte, inoltre, sostiene che "l'introduzione di un 'tariffario' siffatto pare, peraltro, non tener conto né delle peculiarità operative del web, né del potenziale danno alla collettività da misurarsi anche in termini di rinunce e di occasioni perdute; ponendosi, così, in evidente contrasto anche con le chiare indicazioni che provengono dal Piano Nazionale di Digitalizzazione (PND) del patrimonio culturale."

A seguito di questa bocciatura plebiscitaria, che aveva accomunato associazioni e istituzioni diversissime come forse non si era mai visto prima, il Ministero ha aspettato circa un anno senza rispondere ufficialmente, nonostante le numerose sollecitazioni.<sup>7</sup> Alla fine il responsabile dell'ufficio legislativo del Ministero della Cultura è stato spostato, promuovendolo alla Direzione Generale degli Archivi, ed è stato pubblicato un nuovo Decreto Ministeriale ([DM 108 del 21 marzo 2024](#)),<sup>8</sup> che contiene modifiche al precedente DM 161/2023. È indubbio che il nuovo provvedimento segni un progresso notevole: ora finalmente le pubblicazioni sia scientifiche che divulgative, nonché i quotidiani e i periodici di informazione sono esentati dal pagamento di un canone, fatti salvi ovviamente gli eventuali costi vivi per l'esecuzione delle riprese o per la loro fornitura.

Va anche apprezzato il fatto che l'esenzione dal pagamento del canone sia stata estesa ai cataloghi di mostre (entro le 4000 copie di tiratura), che in precedenza erano generalmente compresi tra le iniziative a scopo di lucro. Lo stesso dicasi per il materiale destinato a manifestazioni di valorizzazione del patrimonio culturale con la precisazione secondo cui "il biglietto di ingresso non è di per sé sufficiente a caratterizzare una iniziativa di valorizzazione come a fine di lucro, ma va valutato l'insieme delle circostanze in cui si realizza l'iniziativa stessa". Questa precisazione è un passo

<sup>1</sup> <https://www.benculturali.it/comunicato/dm-161-11042023>.

<sup>2</sup> <https://www.aib.it/notizie/osservazioni-sul-dm-11-aprile-2023-n-161/>.

<sup>3</sup> <https://cunsta.it/articoli-news-cunsta/comunicati-cunsta/548-appello-contro-le-nuove-disposizioni-sull-uso-delle-immagini-dei-beni-culturali-dm-161-2023>.

<sup>4</sup> <https://www.cun.it/provvedimenti/sessione/331/raccomandazione/raccomandazione-del-18-05-2023>.

<sup>5</sup> Raimondi, Eugenio. «*Immagini e beni culturali, Lincei: 'Norme da rifare'*», *Avvenire*, 22 luglio 2023, <https://www.avvenire.it/agora/pagine/immagini-e-beni-culturali-lincei-rifare-norme>.

<sup>6</sup> <https://www.corteconti.it/HOME/Documenti/DettaglioDocumenti?Id=250a9d21-c914-43f9-8165-3c60a197b824>.

<sup>7</sup> Si sono avuti solo alcuni interventi sui quotidiani da parte del responsabile dell'ufficio legislativo del Ministero, Antonio Tarasco, e dello stesso ministro Sangiuliano, che ovviamente non hanno alcun peso normativo anche quando portino chiarificazioni sulla *ratio* del decreto.

<sup>8</sup> <https://www.benculturali.it/comunicato/26075#:~:text=in%20formato%20PDF-,D.M.,e%20luoghi%20della%20cultura%20statali%E2%80%9D>.

avanti molto significativo rispetto all'approccio enormemente più restrittivo delle precedenti linee guida, che consideravano ogni operazione che comportasse un pagamento anche minimo come finalizzata al lucro.

Restano tuttavia forti perplessità di carattere generale sul provvedimento e una serie di criticità di dettaglio. Il Ministero, infatti, non ha avuto il coraggio di abbandonare la vecchia formulazione e di riprendere da capo la materia aprendo una consultazione per una revisione della norma, che tenesse maggior conto delle istanze degli addetti ai lavori.

Innanzitutto si è mantenuta la strutturazione delle linee guida per coefficienti e tabelle a cascata, che rende estremamente macchinoso il calcolo degli importi dovuti, creando inevitabilmente notevoli difformità tra le varie amministrazioni chiamate ad applicarli e con un aggravio di lavoro per gli istituti di cultura.

In secondo luogo resta purtroppo valido il regime di concessione stabilito dall'art. 2, comma 2 del precedente DM 161/2023:

*Indipendentemente dal canone o dal corrispettivo individuato, la concessione per l'uso e la riproduzione dei beni culturali è comunque subordinata alla previa verifica di compatibilità della destinazione d'uso della riproduzione con il carattere storico-artistico dei medesimi beni culturali, ai sensi dell'articolo 20 del Codice dei beni culturali e del paesaggio, di cui al decreto legislativo 22 gennaio 2004, n. 42.*

Qui – come tutti avevano rilevato – c'è una grave confusione tra la compatibilità di uso del bene (sancita dal [Codice dei Beni Culturali](#):<sup>9</sup> “I beni culturali non possono essere [...] adibiti ad usi non compatibili con il loro carattere storico o artistico”) e quella delle sue immagini. Si tratta di due cose ben diverse tra loro, che non possono in alcun modo essere accomunate se non con una forzatura giuridica inaccettabile.

Il regime di concessione, poi, costringe il richiedente a fare domanda con marca da bollo da 16 euro, a cui va aggiunta una seconda marca da bollo di pari valore per il decreto di concessione. Questo anche nel caso in cui non sia dovuto alcun canone e l'immagine sia già stata ottenuta in precedenza dall'istituto di cultura oppure nel caso in cui i costi vivi per la riproduzione siano di pochi euro.<sup>10</sup> A dire il vero sia il vecchio che il nuovo DM definiscono le tariffe, ma non le procedure, che avrebbero bisogno di un provvedimento apposito. Finora, sulla base del precedente DM 161/2023, le prassi amministrative degli istituti di cultura hanno seguito orientamenti differenti. Pur in mancanza di dati complessivi, l'impressione è che gli archivi siano stati più ligi alla lettera richiedendo le marche da bollo, mentre i musei e le soprintendenze territoriali non le abbiano sempre richieste, forse anche perché ricevono una gran quantità di domande dall'estero, il che complicherebbe ulteriormente una situazione già sufficientemente ingarbugliata.

Per spiegare meglio il problema è utile fare un esempio concreto: ipotizzando che l'applicazione della norma sulle marche da bollo sia rigorosamente rispettata da tutti gli istituti, se uno studioso avesse bisogno di pubblicare un articolo con dieci immagini provenienti da dieci istituti diversi

<sup>9</sup> <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2004-01-22;42!vig=>.

<sup>10</sup> Cfr. Modolo, Mirco. 2023. «Il d.m. 11 aprile 2023, n. 161: osservazioni e proposte.» *Aedon. Rivista di Arti e Diritto online*, 2.2023. <https://aedon.mulino.it/archivio/2023/2/modolo.htm>; Brugnoli, Andrea. 2023. «Il d.m. 11 aprile 2023, n. 161 e il suo impatto sulla ricerca e sull'editoria: brevi note a margine di un caso studio di pubblicazione in Open Access.» *ibid.* <https://aedon.mulino.it/archivio/2023/2/brugnoli.htm>.

dovrebbe pagare presumibilmente dai 55 ai 70 euro per i costi vivi (a seconda del tipo di riproduzione richiesta, se foto b/n di buona qualità o file di immagine a colori) e ben 320 euro di marche da bollo, una tassa evidentemente sproporzionata che quintuplica i costi. Se poi si trattasse di una monografia con cento e più immagini dovremmo moltiplicare il tutto per dieci, un costo pressoché insostenibile. Si può sperare che di fronte a questa patente assurda si utilizzi un'interpretazione più liberale esentando dalla marca da bollo le richieste che non comportino il versamento di un canone? Così, per esempio, ha fatto l'Archivio di Stato di Venezia recependo prontamente il nuovo DM.<sup>11</sup> In fondo, dal 2017 fino al DM 161/2023, era questa la procedura adottata dagli Archivi di Stato sulla base delle circolari 33/2017 e 39/2017 della Direzione generale archivi, dando seguito a una raccomandazione del Consiglio Superiore dei Beni Culturali del 16/05/2016. Una simile linea interpretativa si fonda in ultima istanza sul comma 3 bis dell'art. 108 del Codice dei Beni Culturali, in base al quale "sono in ogni caso libere le seguenti attività, svolte senza scopo di lucro, per finalità di studio, ricerca, libera manifestazione del pensiero o espressione creativa, promozione della conoscenza del patrimonio culturale: 1) la riproduzione di beni culturali diversi dai beni archivistici [...]; 2) la divulgazione con qualsiasi mezzo delle immagini di beni culturali, legittimamente acquisite, in modo da non poter essere ulteriormente riprodotte a scopo di lucro". Se sono libere è evidente che non sono sottoposte a concessione.

Non è chiaro però se e in quali casi rimanga valida la "verifica di compatibilità", che è indipendente dalla corresponsione di un canone. Almeno per le immagini "libere" da concessione non dovrebbe essere richiesta, ma in ogni caso c'è da domandarsi quale necessità esista di un controllo "etico" da parte dello Stato per le pubblicazioni scientifiche o anche solo divulgative e didattiche, nonché per giornali e periodici "nell'esercizio del diritto-dovere di cronaca". Si tratterebbe infatti solo di ingolfare gli uffici con richieste puramente formali. Veramente il Ministero crede che (per esempio) uno storico dell'arte della Columbia University di New York o un giornalista della *Frankfurter Allgemeine Zeitung* vadano a cercarsi una normativa di difficile reperimento e comprensione, sulla quale non si raccapezzano bene nemmeno gli Italiani? E varrebbe la pena di istituire una *task force* che monitori migliaia di pubblicazioni internazionali, che in ogni caso sono esentate dal pagamento dei diritti? E i quotidiani come concilierebbero il diritto di cronaca, che richiede tempi immediati, con le lungaggini burocratiche delle concessioni? Infine, qualora l'occhiuto censore identificasse una pubblicazione con un'immagine di un bene statale di cui non risultasse chiara la relativa concessione, come distinguerebbe le foto fornite dagli istituti di cultura dalle "riproduzioni già disponibili on-line che restano liberamente scaricabili" o dalle "riproduzioni eseguite direttamente da privati" per le quali "nessun rimborso spese è dovuto" (punto A.2.1 delle *Linee Guida*)? Senza contare che la recente sentenza del tribunale di Stoccarda<sup>12</sup> ha mostrato come non sia perseguibile legalmente il cittadino di un altro stato che non segua la norma italiana

<sup>11</sup> <https://www.archiviodistatovenezia.it/it/servizi-al-pubblico/servizio-di-fotoriproduzione.html?view=article&i-d=34:pubblicazione-di-immagini-di-documenti&catid=17:servizio-di-fotoriproduzione>.

<sup>12</sup> Stella, Gian Antonio, «Via alla vendita del puzzle dell'Homo Vitruvianus: nessun diritto di immagine all'Accademia di Venezia.» *Corriere della Sera* 24 aprile 2024: [https://www.corriere.it/cronache/24\\_aprile\\_04/disfida-puzzle-dell-homo-vitruvianus-via-libera-teseschi-6621c1dc-f2bf-11ee-ab87-79667834e629.shtml?s=08](https://www.corriere.it/cronache/24_aprile_04/disfida-puzzle-dell-homo-vitruvianus-via-libera-teseschi-6621c1dc-f2bf-11ee-ab87-79667834e629.shtml?s=08). Il Ministero ha annunciato un ricorso (Bison, Leonardo, «Uomo vitruviano 'libero'. Il Mic: 'Atto abnorme, faremo ricorso'.» *Il Fatto Quotidiano* 7 aprile 2024, <https://www.ilfattoquotidiano.it/in-edicola/articoli/2024/04/07/uomo-vitruviano-libero-il-mic-atto-abnorme-faremo-ricorso/7504132/>), ma è assai dubbio che ottenga soddisfazione.



e non paghi il canone per l'utilizzo di immagini di beni culturali italiani, il che stabilisce una sorta di privilegio al contrario: solo i cittadini italiani sono tenuti a pagare.

Se si volesse tenere traccia delle pubblicazioni scientifiche che riguardano opere di proprietà statale per arricchire le schede inventariali, la soluzione sarebbe elementare: si potrebbe richiedere una semplice comunicazione (e magari un estratto in pdf della pubblicazione) evitando così la procedura della concessione, assai onerosa per l'impegno amministrativo degli uffici competenti e costosa per i richiedenti.

Un'ulteriore assurdità derivante dal regime di concessione consegue dall'art. 3 del nuovo DM 108/2024, che stabilisce che "i provvedimenti di concessione sono tempestivamente pubblicati nella sezione 'Trasparenza' del Ministero, ivi inclusi gli importi dei canoni e corrispettivi determinati". Dunque non solo le concessioni a titolo oneroso e con introiti significativi, ma in linea di principio anche quelle a titolo gratuito (per un articolo scientifico o per un quotidiano) andrebbero pubblicate in quella sede: potenzialmente migliaia di richieste per una o poche foto di nessuna rilevanza e senza corrispettivo economico dovranno intasare il sito ministeriale. Sembra che l'estensore della norma non si accorga o non si curi degli straordinari costi amministrativi e dei carichi di lavoro che in questa maniera vanno a gravare sulla pubblica amministrazione, senza che se ne ricavi alcun utile, né morale né materiale.

Veniamo ora alle osservazioni di dettaglio. Nel nuovo decreto i "macroprodotti" (denominazione di cui sfugge il senso preciso) per le riproduzioni sono passati dai nove del vecchio decreto ai sei dell'attuale, ma non si capisce come inquadrare alcune categorie che sembrano sfuggire alle maglie della norma: rispetto alla prima versione sono stati eliminati (giustamente) i videoclip e i fotocolor. Questi ultimi possono essere accomunati probabilmente alle diapositive, benché con l'avvento del digitale resti difficile capire quanta richiesta esista di questo tipo di riproduzioni, che ormai nessuno più utilizza. Sono sparite però le scansioni, il che solleva dei problemi non dappoco: come si fa a calcolare il costo della riproduzione di un manoscritto? Ormai, infatti, la scansione ha sostituito il microfilm, che anche in questo caso nessuno usa più e che invece è rimasto compreso tra i macroprodotti.<sup>13</sup> I costi saranno da equiparare a quelli delle fotocopie o a quelli dei microfilm? In questi ultimi si prevede il costo della "digitalizzazione" (un euro a fotogramma), ma sembra che in questo caso ci si riferisca piuttosto a una vecchia prassi, ormai superata in quanto la digitalizzazione di un microfilm dà un'immagine di cattiva qualità e non è più economica della nuova scansione del manoscritto originale, certamente preferibile se non ci siano problemi conservativi.

Una terza ipotesi sarebbe quella di comprendere le scansioni tra le "immagini digitali", se non fosse che non si capisce bene in quale sottospecie andrebbero poi inquadrare: quelle per il web? O quelle per la stampa? Per studiare un codice o trascriverne delle parti, però, basta esaminarlo a schermo: non è necessaria né la pubblicazione su web, né la stampa. In ogni caso il costo (rispettivamente cinque e sette euro a scatto) sarebbe assolutamente esorbitante e fuori mercato, soprattutto per la riproduzione a bassa risoluzione nella versione per il web. La scansione di un codice di 500 pagine verrebbe così a costare dai 2500 ai 3500 euro, costo paragonabile a quello di una riproduzione fotografica con stampe 24x30, che certamente è molto più complessa e onerosa per l'amministrazione oltre al fatto che, ancora una volta, l'avvento della foto digitale l'ha resa del

---

<sup>13</sup> Ci si potrebbe anche chiedere come si fa a etichettare i microfilm come macroprodotti, con un evidente ossimoro.

tutto obsoleta, mentre la riproduzione come microfilm si fermerebbe a 500 euro. Appare evidente come l'estensore del provvedimento non abbia chiare né le differenze tra i vari procedimenti, né l'evoluzione delle tecniche fotografiche, né – soprattutto – le esigenze degli studiosi. Occorrerebbe distinguere invece tra scansione e riproduzione fotografica digitale, come fa per esempio l'Archivio Centrale dello Stato con prezzi equi. L'impressione complessiva è che l'estensore delle norme sia rimasto fermo alla situazione tecnologica di una generazione fa, senza alcuna dimestichezza con il lavoro quotidiano di un archivio fotografico. La conseguenza sarà che l'interpretazione della norma sarà a carico degli istituti di cultura, cosicché per digitalizzare un codice una biblioteca potrebbe scegliere di equiparare le scansioni ai microfilm, un'altra alle immagini digitali, con una differenza di prezzo tra le due opzioni anche del 600%.

Vengono inoltre considerate libere “le riproduzioni di immagini di beni culturali contenute in pubblicazioni liberamente accessibili da chiunque (c.d. *open access*) in quanto prive di un prezzo di copertina”. Anche in questo caso sembra che l'estensore del testo non abbia dimestichezza né con la terminologia, né con le licenze di questo tipo, né infine con le prassi editoriali. I criteri dell'*open access* sono stati definiti con l'iniziativa di Budapest e, nel 2003, con la [Dichiarazione di Berlino](#).<sup>14</sup> In seguito, sono stati adottati dalle università italiane con la [Dichiarazione di Messina \(2004\)](#), a cui hanno aderito finora 71 atenei italiani.<sup>15</sup> Tali criteri concedono agli utenti il riutilizzo e l'adattamento del materiale senza restrizioni commerciali (licenza CC BY). L'*open access* nella sua forma pura, però, non sarebbe ammesso dal Codice dei Beni Culturali, che vieta la possibilità di riuso di immagini a finalità commerciali (art. 108, c. 3 bis, n. 2). Altrimenti il DM dovrebbe specificare il tipo di licenza tenendo conto almeno delle principali: quelle Creative Commons Zero (CC 0), Creative Commons Attribuzione (CC BY), Creative Commons Attribuzione – Condividi allo stesso modo (CC BY-SA), o infine quella con restrizioni per gli usi commerciali (CC BY-NC): basterebbe andare su Wikipedia alla voce [open access](#)<sup>16</sup> per avere un'idea di queste varietà.

Non si solleva qui il problema dell'evidente conflitto tra il Codice dei Beni Culturali e le normative europee e italiane relative alle pubblicazioni frutto di progetti finanziati con fondi pubblici (progetti ERC o PRIN, per esempio ma non solo), che prevedono nel bando appunto l'*open access*. Il problema andrebbe finalmente risolto con una liberalizzazione più coraggiosa, semplicemente cassando una riga dell'articolo 108 del Codice dei Beni Culturali.

A margine andrebbe considerato anche che molti editori in *open access* utilizzano la prassi del *print on demand*, per cui esiste anche un'edizione a stampa dell'opera *online*, che viene realizzata su richiesta e può essere acquistata da chi preferisce il cartaceo, ovviamente pagando i costi di stampa. Tale pratica mira, inoltre, alla conservazione anche in formato cartaceo attraverso il deposito legale e alla fruizione nelle principali biblioteche per chi non avesse dimestichezza con gli strumenti informatici. Questa fattispecie sembra sfuggire ancora una volta al decreto in esame, pur essendo assai diffusa.

C'è infine qualche ingenuità dovuta alla scarsa dimestichezza con l'ambiente accademico e la sua normativa. Alludo alla previsione di gratuità per “le riproduzioni di beni culturali e il loro riuso per le riviste scientifiche e di Classe A di cui agli elenchi” dell'ANVUR. Poiché le riviste di

<sup>14</sup> [https://openaccess.mpg.de/67682/BerlinDeclaration\\_it.pdf](https://openaccess.mpg.de/67682/BerlinDeclaration_it.pdf).

<sup>15</sup> <https://www.cruis.it/open-access.html>.

<sup>16</sup> [https://it.wikipedia.org/wiki/Open\\_access](https://it.wikipedia.org/wiki/Open_access).

Classe A sono necessariamente già comprese tra gli elenchi di quelle scientifiche, sarebbe bastato limitarsi a ricordare queste ultime. Sarebbe come dire (per esempio) che sono esentati tutti gli Italiani e inoltre anche i Toscani. Errore tra tutti il meno grave, ma ulteriore indizio di ingenuità e approssimazione.

In ultimo va rilevata un'interessante spia che rivela come negli uffici ministeriali si abbia una qualche coscienza dell'inadeguatezza della normativa, benché rinnovata dopo un solo anno di applicazione. Infatti, all'art. 3 del nuovo DM si specifica che:

1. *L'organo amministrativo di vertice del Ministero potrà adottare direttive specifiche di maggior dettaglio.*
2. *L'applicazione delle Linee guida allegate al presente decreto sarà oggetto di monitoraggio da parte dell'organo amministrativo di vertice del Ministero, anche in vista di una possibile revisione delle stesse.*

A parte il fatto che resta poco chiaro quale sia l'organo amministrativo di vertice (l'Ufficio Legislativo? Il Gabinetto del Ministro? Uno dei nuovi dipartimenti in corso di istituzione?), è notevole l'ammissione che non si esclude di poterci tornare su (almeno) una terza volta. Ciò, a dire la verità, sarebbe fortemente auspicabile sulla base delle contraddizioni e dei limiti sopra evidenziati, a patto però che l'intervento non sia solo cosmetico, ma sostanziale, cosa non difficile se solo il Ministero abbandonasse l'orgogliosa impermeabilità alla consultazione degli esperti che, a vario titolo, hanno competenza per dare un parere tecnico autorevole. In fondo era questa la procedura virtuosa meritoriamente adottata in passato dallo stesso Ministero per scrivere il *Piano Nazionale di Digitalizzazione (PND)*, che andrebbe recuperato e fortemente valorizzato, ma che invece sembra essere stato chiuso in un cassetto.

È speranza vana quella di una radicale semplificazione e di una decisa liberalizzazione delle immagini dei beni culturali statali, così come auspicato sia dal *PND* che dalla Corte dei conti e da tutto il mondo della cultura?