

Paolo Traniello (1938-2023)

Editorial board

Contact: info@jlis.it

JLIS.it ricorda con affetto il prof. Paolo Traniello, già docente di Bibliografia e Biblioteconomia nelle Università della Calabria, dell'Aquila e di Roma Tre. I suoi libri, i suoi articoli, le sue relazioni congressuali sono stati e sono tuttora un riferimento per tutti coloro che desiderino avere una visione storica delle biblioteche italiane inserite nel contesto internazionale. I volumi *La biblioteca pubblica: storia di un istituto nell'Europa contemporanea* (1997), *Storia delle biblioteche in Italia dall'Unità a oggi* (2002), *Biblioteche e società* (2005) pubblicati in più edizioni da Il Mulino delineano magistralmente l'evoluzione del sistema bibliotecario italiano dall'Unità a oggi nei suoi aspetti legislativi, amministrativi, organizzativi, culturali, all'interno della contraddittoria politica istituzionale. Un docente che, come pochi, è stato un vero accademico, un ricercatore capace di inaugurare strade nuove e sfatare luoghi comuni. In questo senso, sono fondamentali i suoi studi sulla *public library* anglosassone, istituto assai diverso dalla biblioteca pubblica italiana e in nessun modo sovrapponibile a essa. Il suo ultimo libro, *Le opere e i libri*, descrive con puntualità la figura dell'editore ottocentesco nel suo rapporto con gli autori, analizzando i casi di tre protagonisti della letteratura italiana: Ugo Foscolo, Giacomo Leopardi e Alessandro Manzoni; con ciò Paolo testimonia ancora una volta la sua curiosità intellettuale in varie direzioni del mondo del libro. I suoi studi rappresentano un apporto originale e basilare per comprendere il complesso mondo delle biblioteche; su di essi si sono formate generazioni di studenti, di aspiranti bibliotecari, di bibliotecari.

Il prof. Paolo Traniello è stato un grande studioso, una persona gentile, affabile, garbata, un Maestro, un signore, un amico carissimo.

Librarianship as Civic Engagement

Mauro Guerrini^(a)

a) Università degli Studi di Firenze, Dipartimento di Storia, Archeologia, Geografia, Arte e Spettacolo (SAGAS)

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

Received: 05 September 2023; **Accepted:** 14 September 2023; **First Published:** 15 January 2024

ABSTRACT

The themes discussed in the contribution recur in the author's theoretical reflection in recent years: definition of LIS: the discipline that reflects on the construction, organization, management and use, as well as on the languages and services to the public of the library; relationship between librarian and LIS teacher: both figures have the same identity, both insist on the same cultural and technical background; open access: the theme is fundamental as a substrate for the free sharing of research products; metadata creation: the process of recording data functional to the identification and retrieval of any type of resource; terminology: new concepts and new types of resources have imposed new terms; authority control: by assembling different data, it allows a greater identification of the resources and of the entities associated with them. The reflection ends with the importance of the international dimension of research.

KEYWORDS

LIS; open access; metadata creation; authority control; LIS terminology

Biblioteconomia come impegno civile

ABSTRACT

I temi trattati ricorrono nella riflessione teorica dell'autore degli ultimi anni: definizione di Biblioteconomia: la disciplina che riflette sulla costruzione, l'organizzazione, la gestione e l'uso, nonché sui linguaggi e sui servizi al pubblico della biblioteca; rapporto tra bibliotecario e docente di Biblioteconomia: entrambe le figure hanno una medesima identità, insistono nello stesso alveo culturale e tecnico; open access: il tema risulta fondamentale come substrato per la condivisione libera e gratuita dei prodotti della ricerca; la metadattazione: il processo di registrazione dei dati funzionali all'identificazione e al reperimento di qualsiasi tipo di risorsa; la terminologia: nuovi concetti e nuove tipologie di risorsa hanno imposto nuovi termini; authority control: consentire, assemblando dati diversi, una maggiore identificazione delle risorse e delle entità a esse associate. La lezione chiude con il richiamo alla dimensione internazionale della ricerca.

PAROLE CHIAVE

Biblioteconomia e Scienza dell'Informazione; LIS; open access; Metadattazione; Controllo di autorità; Terminologia LIS

Text of the last lecture held by full professor Mauro Guerrini – University of Florence, SAGAS Department, Aula Magna, October 27, 2023.

Le biblioteche, sempre di più, non sono solo luoghi di prestito e transito di libri, ma sono spazi di dialogo dove si costruisce e ci si sente parte della comunità, dove si fa attività di Information Literacy, dove si studia, dove si coltivano le memorie, dove si legge, dove si cataloga, dove si lavora in e con il digitale, dove si creano cataloghi ragionati e in rete, dove si praticano inclusione e welfare culturale, dove si progettano attività, dove si svolgono laboratori e dove si pratica la libertà in tutte le sue sfaccettature. La biblioteca è tutto questo e non solo, la biblioteca è la bellezza del bene comune, che va tutelato e valorizzato.
Comunicato AIB, 11 agosto 2023

Biblioteconomia: valori e confini

La Biblioteconomia è la disciplina che riflette sulla costruzione, l'organizzazione, la gestione e l'uso, nonché sui linguaggi e sui servizi al pubblico della biblioteca; essa definisce i criteri e le modalità per il governo di una struttura complessa qual è la biblioteca in quanto spazio fisico e virtuale e in quanto sistema che seleziona, conserva, tutela, descrive e trasmette – tramite bibliotecari professionali – quella porzione di universo bibliografico che possiede e a cui dà accesso per la lettura, la ricerca e lo svago: manoscritti, libri e periodici a stampa e digitali, carte geografiche, musica, fotografie, film e altre risorse. La sintetica definizione tenta di rendere merito a un dominio disciplinare ricco ed eterogeneo e, dunque, complesso da inquadrare. Per quanti tentativi si possano fare, permane la sensazione di non aver colto tutte le sfumature, di non aver evidenziato temi importanti, di non aver fatto emergere dettagli insignificanti solo in apparenza, i quali, invece, delineano una disciplina composita, profondamente inserita nel contesto culturale e intrinsecamente democratica.

Le biblioteche di ogni tipologia sono componenti essenziali di una società democratica; esse, che vivono di cultura, sono una risorsa per l'intera organizzazione sociale, culturale ed economica di un Paese. Non casualmente Virginia Carini Dainotti le ha definite “istituti della democrazia”, ovvero enti che aiutano la formazione e la crescita di cittadini consapevoli e responsabili sostenendo lo sviluppo culturale, offrendo le risorse bibliografiche per ampliare le conoscenze e gli spazi per momenti di socialità. Assicurare che tutte le procedure e le politiche siano animate dall'etica del servizio verso il singolo, le comunità, la società nel suo complesso e gli utenti futuri significa, altresì, garantire la libertà intellettuale intesa come libertà d'espressione e di formazione di un pensiero critico. L'equità d'accesso alla conoscenza e all'informazione mira a garantire la disponibilità universale ai programmi e alle risorse bibliografiche superando ogni barriera tecnologica ed economica. L'uguaglianza non è il valore ultimo, bensì l'accesso: uguaglianza, infatti, significa dare alle persone le stesse cose, equità significa dare alle persone le stesse possibilità, come insegna la lezione di don Lorenzo Milani in *Lettera a una professoressa* (Scuola di Barbiana 1967).

Si può, dunque, comprendere come designare finalità e confini della Biblioteconomia sia un obiettivo molto più complesso di quanto si possa credere.

Carlo Revelli ha sottolineato come la Biblioteconomia sia riflessione sulla professione bibliotecaria e sul ruolo della biblioteca all'interno della società. Pertanto, si può affermare che nell'attività del bibliotecario la dimensione tecnica, essenziale per operare con competenza, non può prescindere o separarsi dall'attenzione ai diritti civili e al modo in cui questi vengono vissuti e praticati nell'ambito della comunità d'appartenenza. Senza impegno civile la competenza professionale diventa una dimensione tecnicistica, incapace d'incidere sulla crescita della società. Grazie all'attività dei bibliotecari, la trasmissione della conoscenza registrata contribuisce sempre più alla libertà, ai diritti, al benessere di tutti per le generazioni presenti e future, in un'ottica di sostenibilità come definita dall'ONU nell'Agenda 2030, sottoscritta il 25 settembre 2015 dai governi dei 193 Paesi membri delle Nazioni Unite. L'attuazione dell'Agenda 2030 richiede un coinvolgimento di tutte le componenti della società, dalle imprese private al settore pubblico, dalla società civile agli operatori dell'informazione e cultura, comprese le biblioteche.

Delimitare i confini attuali di un terreno disciplinare così composito come quello della Biblioteconomia non è semplice, confini, peraltro, che sono stati e sono differenti in epoche e in aree geografiche diverse. La stessa valenza semantica del lemma *biblioteconomia* differisce a seconda dell'area geografica, linguistica e dal contesto socioculturale in cui si pone e di cui è espressione. La storia della Biblioteconomia, come per tutte le discipline, è la storia delle opere e delle attività dei suoi principali protagonisti. Essa abbraccia la progettualità e l'innovazione, nel rispetto della tradizione culturale e professionale. Luigi Crocetti – come emerge dalla raccolta *Le biblioteche di Luigi Crocetti: saggi, recensioni, paperoles (1963-2007)* del 2014 (Desideri e Petrucciani 2014) – offre una visione disincantata della storia della Biblioteconomia italiana, che possiede più passato che tradizione. La caratteristica peculiare e positiva della tradizione culturale italiana è il policentrismo. “Il policentrismo sarà dunque da assumere a canone precipuo della nostra cultura. [...] Il valore delle tessere di un mosaico è uniforme. Il valore della tradizione italiana in questo secolo è qui”¹. La tradizione delle biblioteche italiane è principalmente una memoria di bibliotecari-studiosi.

La Biblioteconomia non frapponne rigidi confini, anzi, ha relazioni strette con discipline consorelle quali la Bibliografia e la Documentazione e con l'Archivistica e la Museologia; con la Codicologia, la Paleografia, la Diplomatica che dovrebbero costituire, nel solco di una nobile quanto purtroppo spesso ignorata tradizione, parte del contesto culturale e formativo del bibliotecario; infine, in tempi contemporanei, con l'Informatica e le Digital humanities; e, più in generale, con le discipline psicopedagogiche e sociologiche (biblioteca pubblica) fino alla storia della cultura.

Dai primi anni del 2000, complice l'era digitale e l'interoperabilità potenziale tra bacini digitali d'informazioni, lo spartiacque tra trattamento bibliografico e archivistico delle risorse bibliografiche e documentarie sembra sbiadirsi, alla luce di modellizzazioni pur non del tutto formalizzate e accettate dalle rispettive comunità. Sintetizzare tecniche descrittive di dominio entro nuovi paradigmi significa promuovere l'integrazione disciplinare e una politica di collaborazione che, sorretta da ragionevoli premesse metodologiche, rispetti le peculiarità di ambiti specifici e, al contempo, consenta la condivisione e la diffusione delle conoscenze.

¹ <http://tinyurl.com/3nfmzdpf>.

Biblioteconomia e professione bibliotecaria

Gli studi biblioteconomici, ricorda Michael Gorman, sono il campo d'indagine di quei professionisti che: raccolgono e danno accesso a sottoinsiemi della *testimonianza umana* elaborati in ogni epoca e in ogni luogo, ossia le collezioni; organizzano e catalogano quei sottoinsiemi per renderli reperibili; si dedicano alla conservazione, disseminazione e futura trasmissione della *testimonianza umana*. Con *testimonianza umana* s'intende quel vasto insieme di conoscenze testuali, visive e simboliche, in ogni formato, registrate dall'uomo nel corso del tempo.

Il bibliotecario ha un ruolo decisivo; è un intellettuale con una solida formazione culturale e con un'altrettanta solida competenza tecnica. Egli, quale attore protagonista della disciplina e soggetto attivo nella società, rende efficace il valore della democrazia in ogni momento della professione.

La Biblioteconomia ha due volti: il bibliotecario, ovvero la professione, e il docente, ovvero l'accademia. Entrambi hanno una medesima identità, insistono nello stesso alveo culturale e tecnico e sono convergenti su un terreno che si nutre di proficui interscambi. Non sussiste, infatti, una dicotomia tra le coppie biblioteca-bibliotecari e biblioteconomia-professori, bensì un intreccio di competenze e finalità complementari.

Al bibliotecario, quale persona di cultura, viene chiesta la padronanza di competenze tecniche che lo rendano in grado di rispondere efficacemente alle sfide poste dalla realtà culturale e sociale circostante.

Al docente, quale accademico, viene chiesto di riflettere su un ampio e articolato oggetto di studio e avere la capacità didattica di presentare il contesto internazionale e nazionale, selezionando il grano dalla crusca, ovvero ciò che fa progredire gli studi da ciò che è effimero.

A entrambi, come testimonia l'attività svolta in ambito IFLA (International Federation of Library Associations and Institutions), viene chiesta la competenza necessaria per poter lavorare sulle metodologie e sulle pratiche finalizzate alla costruzione delle raccolte, ai processi e ai linguaggi di mediazione catalografica, di creazione di metadati e di predisposizione dei servizi per i lettori. Questa è la concezione di Biblioteconomia di Diego Maltese basata sulla biblioteca come sistema da organizzare in modo rigoroso in tutte le sue parti, concezione che abbraccia sia il bibliotecario, sia il docente di Biblioteconomia.

Alla base della Biblioteconomia e, conseguentemente, della professione bibliotecaria, Jesse Hauk Shera pone il connubio tra i processi individuali della conoscenza e le azioni e i comportamenti umani, in una compenetrazione di psicologia e sociologia. Scrive: "Lo scopo della Biblioteconomia, a qualunque livello intellettuale operi, è massimizzare l'utilità sociale delle registrazioni grafiche, sia che l'utente servito sia un bambino illetterato immerso nel suo primo libro illustrato, o il più progredito studioso impegnato in qualche misteriosa indagine. Senza dubbio, per di più, ma fondamentalmente, la Biblioteconomia è la gestione della conoscenza" (Shera 1965, 16).

Serrai in *Biblioteconomia come scienza. Introduzione ai problemi e alla metodologia* (Serrai 1973) sostiene che la Biblioteconomia ha come oggetto di studio la struttura e il funzionamento dei sistemi ai quali è affidato il compito di mettere in relazione i prodotti intellettuali e informativi di alcuni uomini con le necessità intellettuali e informative di altri uomini, di solito distanti dai primi nel tempo e nello spazio. Problema cardinale della Biblioteconomia, soprattutto in riferimento alle responsabilità culturali e sociali che a essa spettano, è fornire una giustificazione, in termini logici ed economici, del modo in cui le biblioteche assolvono le funzioni a loro precipue.

L'autore pone l'attenzione sugli scopi che la disciplina deve perseguire per rispondere alle esigenze informative dei lettori; la biblioteca è intesa come un sistema che connette i prodotti intellettuali e informativi con gli individui. Il divario tra la raccolta e i lettori è superato tramite il *catalogo*. Seppur a fasi alterne e nonostante nel tempo siano cambiati i singoli temi in discussione, gli equilibri e le parentele disciplinari, l'attenzione degli studi si è focalizzata tradizionalmente sulla messa a punto del catalogo e, soprattutto dal secolo scorso, a esaminare il rapporto tra la biblioteca, i bibliotecari e la società; sempre più, in epoca contemporanea, la Biblioteconomia ha cercato di trovare elementi di contatto con le discipline sociali per elaborare criteri efficienti di gestione e valutazione delle raccolte e per l'organizzazione di servizi propri dell'era digitale e sempre più mirati alle specifiche necessità dei lettori.

L'open access e la disseminazione della conoscenza in era digitale

Biblioteche e bibliotecari offrono libero accesso alle informazioni nel rispetto dei principi etici professionali e hanno la responsabilità di conservare e permettere la libera e gratuita fruizione della memoria registrata. Un valore centrale nella professione è la fede nella libertà intellettuale che è fondamento della possibilità di raggiungere la saggezza. Con la creazione nel 1997 del Comitato FAIFE (Free Access and International Freedom of Expression), l'IFLA ha compreso tra i temi d'interesse della Biblioteconomia l'analisi e la valutazione del contesto sociale e politico, invitando i bibliotecari a tutelare la libertà intellettuale e a promuovere e difendere i diritti umani in relazione al libero accesso all'informazione. Concetti ribaditi dal *Manifesto IFLA/Unesco sulle biblioteche pubbliche* pubblicato il 27 luglio 2022, che aggiorna l'edizione del 1994 e precedenti (1949, 1972). Il *Manifesto* privilegia alcuni temi specifici della biblioteca pubblica, ma i valori chiave sono validi per tutte le tipologie di biblioteca, soprattutto quando afferma, in esordio, che la "biblioteca pubblica, porta d'accesso locale alla conoscenza, crea i presupposti di base per l'apprendimento permanente, l'autonomia nel processo decisionale e lo sviluppo culturale dell'individuo e dei gruppi sociali" (IFLA e Unesco 2022).

Alex Byrne, in *"Ma per seguir virtute e canoscenza": ethics in library and information science* (Byrne 2017), Lectio magistralis all'Università di Firenze nel 2017, affronta temi come l'etica bibliotecaria, l'accesso aperto all'informazione e le responsabilità delle biblioteche verso gli individui e la società. Le biblioteche svolgono un ruolo fondamentale per la formazione di un'identità culturale e per la crescita intellettuale dei cittadini. Secondo l'autore, la nostra particolare e speciale responsabilità consiste nella dedizione all'umanità, alla conoscenza e alla virtù.

Il tema dell'open access, ovvero il movimento composito nato per iniziativa di studiosi americani e promosso inizialmente da alcune società e fondazioni² alla fine degli anni Novanta del XX secolo con lo scopo di favorire la disseminazione della conoscenza nell'era digitale, risulta fondamentale come substrato per la condivisione libera e gratuita dei prodotti della ricerca. L'aggettivo *open* è declinato in vari modi e si unisce a molti sostantivi che esprimono concetti diversi, dall'architettura (open space), alle modalità di discussione e presentazione di iniziative (open section, open seminar, open day, open conference), all'educazione (open education) alla letteratura scientifica

² <http://tinyurl.com/5n92a2fj>.

(open access), alla scienza (open science), all'istruzione (open university), alle biblioteche (open library) all'intera società (open society), a testimonianza di una filosofia che sottende una visione aperta, dinamica e positiva del rapporto individuo-società. *Open* è affiancato, inoltre, a sostantivi che esprimono nuove visioni della tecnologia improntate alla partecipazione, condivisione e trasparenza, quali *open data* e *open source*. Nel caso dell'applicazione del concetto *open* ai prodotti della ricerca si tratta di un reale cambiamento del paradigma delle modalità di fruizione e trasmissione della conoscenza.

Il concetto di *open*, in Biblioteconomia, proviene da lontano e potremmo individuarne le radici nello statuto stesso della Bibliografia che mira a costruire, e allo stesso tempo a descrivere e a rendere accessibili, le opere dell'universo bibliografico che rappresentano e preservano la conoscenza umana. Radici che affondano, inoltre, nel concepimento del concetto di biblioteca pubblica: da collezioni librerie riservate al solo possessore esse divengono nel tempo aperte a un pubblico sempre più vasto di lettori, che la percepiscono come una struttura di servizio bibliografico. La public library è la prima realizzazione della filosofia open. 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.

I padri della Biblioteconomia moderna, tra cui Melvil Dewey e S.R. Ranganathan, hanno concepito la biblioteca a scaffale aperto, elaborando una classificazione e una disposizione dei libri grazie alla quale il lettore può muoversi liberamente tra le collezioni e arrivare direttamente a insiemi di contenuti omogenei semanticamente. La disposizione classificata suggerisce, inoltre, letture di libri sconosciuti, che possono permettere l'estensione delle conoscenze del lettore, fino ad arrivare alla scoperta casuale, inattesa o fortuita, secondo le dinamiche della serendipità. In particolare, grazie allo scaffale aperto, l'incontro tra libro e lettore avviene senza la mediazione palese del bibliotecario (il quale, però, è sempre presente come l'artefice dell'organizzazione bibliografica). La biblioteca accentua così il suo ruolo di servizio destinato all'uso e alla soddisfazione del lettore. Si realizza lentamente, ma inesorabilmente, l'assunto della prima legge di Ranganathan *Books are for use*, in base alla quale la biblioteca è sempre più concepita come un servizio al cittadino tramite la promozione dell'incontro tra libro e lettore (seconda e terza legge di Ranganathan).

Accesso aperto, dunque, a tutte le fasi della produzione scientifica, a partire dalla produzione dei dati grezzi e sperimentali della ricerca, organizzati in database, a quella dei risultati finali, per il loro riuso e la loro modifica ed espansione. Soprattutto perché in questo modo la scienza diventa funzionale all'innovazione e più utile all'intera società. Sono, infatti, favoriti l'interscambio e le collaborazioni nella ricerca e nella risoluzione dei problemi; più in generale sono stimolati il progresso scientifico, l'innovazione, l'avanzamento dei paesi in via di sviluppo. La filosofia open access deve, pertanto, essere sempre più percepita favorevolmente dai ricercatori sotto il profilo etico, oltretutto bibliometrico, perché, come afferma lo storico Guy Geltner, "l'accesso aperto è un diritto umano, non un modello di business" (Geltner 2015). La realizzazione della filosofia open corrobora il paradigma del servizio bibliotecario contemporaneo.

Nella stessa prospettiva nasce nel 2009 «JLIS.it» (*Italian Journal of Library [poi anche Archive] and Information Science*) che rappresenta oggi un esempio di *diamond road*, ovvero della strada che realizza l'open access mediante la pubblicazione immediata dei contributi in accesso aperto, senza embargo e senza il pagamento di APC (Article Processing Charge), ma col contributo finanziario dell'Università.

La metadattazione, ovvero la trasformazione bibliografica

Stiamo vivendo una fase di trasformazione bibliografica che coinvolge tutto il mondo bibliotecario e la riflessione biblioteconomica. Dall'era convenzionale siamo passati all'era digitale, un percorso appena iniziato, le cui conseguenze sono ancora da verificare. La Biblioteconomia in era digitale mantiene un senso di continuità con i valori tradizionali e, nel contempo, costituisce parte attiva della transizione verso un nuovo paradigma. Il linguaggio di comunicazione tra la biblioteca e i suoi lettori cambia in continuazione e la biblioteca, in quanto istituzione dinamica e *growing organism*, mai statica nella sua prospettiva, non può che essere paladina di nuove forme e di nuove modalità di trasmissione efficace della conoscenza registrata.

Il web ha modificato il modo di *scoprire, accedere, integrare e utilizzare* i dati per le sue caratteristiche di *ubiquità* (virtuale) e per la sua *natura distribuita e scalabile* (modulare). In particolare, è cambiato il contesto tecnologico entro cui si pongono la descrizione delle risorse bibliografiche e la creazione dei suoi accessi: il contesto digitale necessita, infatti, di un linguaggio proprio affinché il prodotto della descrizione e della creazione degli accessi sia del web e non solo nel web, come ha efficacemente affermato Karen Coyle. Il web ha determinato un'accelerazione nella quantità di dati prodotti e fruiti dalle persone e una loro sempre maggiore rapidità e facilità di circolazione tramite strumenti quali computer, tablet e smartphone. L'iniziale web dei documenti si è evoluto negli ultimi due decenni nel web dei dati, in cui le relazioni, cioè i collegamenti qualificati, s'instaurano tra dati e non più tra documenti. L'arricchimento semantico del dato, apportando ulteriore significato rispetto a quello di partenza, consente di generare reti di valore nelle quali i metadati, ovvero i dati sui dati, costituiscono i diversi nodi. Da qui discende una visione estesa dell'ambito d'applicazione della metadattazione, intesa come la catalogazione in era digitale, che abbraccia tutti i contesti, i *cross-domain* necessari ad accompagnare l'esperienza della ricerca.

Il *Manifesto sul web semantico* dell'AIB emanato il 12 ottobre 2020 inizia con questa frase: "Il Web è diventato il più grande insieme di significato condiviso nella storia umana, eppure questo significato risulta frammentato a causa di differenze linguistiche, concettuali e normative" (AIB 2020). C'è tutto: il contesto digitale e il paradigma del lavoro intellettuale, il *digital divide* tra chi può e chi non può approfittare della grande opportunità della ricerca e del lavoro in rete, la responsabilità e la condivisione. Subito dopo il *Manifesto* parla del passaggio dal web dei documenti al web dei dati, di dati auto-esplicativi: è un passaggio sostanziale che introduce la tecnologia dei linked data, il cui uso, come afferma il punto finale del primo paragrafo, non deve ingannare: "Dietro alla semplicità dei Linked Data è presente una descrizione complessa dei dati con i quali si formano le *triple*, e tale complessità aumenta se l'obiettivo è quello di collegare i dati fra domini e set di dati differenti" (Ibid.). Prosegue il *Manifesto*: "La maggior parte delle iniziative di gestione dell'informazione stanno adottando strategie aperte, decentralizzate e globali, incentrate sui dati" (Ibid.), ovvero il *Manifesto* introduce la dimensione dell'interoperabilità e dell'accessibilità e apertura dei dati, ovvero dell'open data, filosofia che si inserisce nella strategia più ampia dell'open science. Si tratta di concetti qualificanti e decisivi.

Il termine *metadattazione* dagli inizi del XXI secolo sostituisce il termine *catalogazione* quando il contesto di applicazione si amplia, sia in termini di domini e comunità di utenti, sia in termini di funzioni e scopi dei metadati. La metadattazione è il processo di registrazione dei metadati,

ovvero dei dati funzionali all'identificazione e al reperimento di qualsiasi tipo di risorsa (una persona, un'opera, un oggetto d'arte), *in primis* le risorse bibliografiche. L'intero processo della metadattazione è controllato dai catalogatori e gestori dei metadati; è il catalogatore o creatore di metadati che, con le sue capacità di gestire sistemi di metadati, mette in relazione risorse e lettori, come storicamente ha sempre fatto. La novità consiste nel passaggio dalla centralità del *record* alla centralità del *dato* e, pertanto, dalla gestione dei *record* alla gestione dei *metadati*. In era digitale gli elementi del record tradizionale vengono resi dati granulari; ciò che conta è l'identificazione delle entità e la loro registrazione. La diversità dei due processi, la catalogazione prima e la metadattazione oggi, è spiegata da Carlo Bianchini in un saggio su *Metadattazione* del 2022: "Il prodotto materiale che risulta dai due processi della catalogazione e della metadattazione è diverso. Nel primo caso il risultato è un catalogo, cioè un elenco ordinato secondo un preciso principio, il principio in base al quale il catalogo viene costruito: per autore, per titolo, per classe, per soggetto, topografico e che presenta informazioni strutturate in una sequenza predefinita, secondo gli scopi della descrizione stessa. [...] Nella metadattazione, il prodotto sono i metadati sui medesimi oggetti descritti dal catalogo, ma le modalità d'accesso alle descrizioni e la struttura della presentazione dei dati sono aperte, possono essere definite a posteriori. Sicché dai metadati possiamo ottenere i medesimi ordinamenti e la medesima presentazione strutturata delle informazioni che si avevano nel catalogo a stampa, a schede o elettronico, ma anche molto, molto di più" (Bianchini 2022, 109-110). Con il paradigma della metadattazione, i dati possono essere assemblati e visualizzati secondo molteplici modalità; non è più necessaria la coincidenza tra l'insieme dei dati selezionati e la loro presentazione, ovvero la loro visualizzazione. I dati sono liberi di poter essere aggregati in modalità diverse, purché, ovviamente, sempre secondo una logica e non in modo arbitrario. La destrutturazione del record genera una trasformazione profonda a livello concettuale, con ovvie ripercussioni sugli strumenti di mediazione tra universo bibliografico e lettore. La metadattazione aumenta il potenziale informativo del catalogo, il quale, come storico strumento di mediazione tra le risorse e i lettori, si pone l'obiettivo di esprimere la ricchezza delle relazioni che legano i metadati. Con l'introduzione del web semantico e del paradigma dei linked open data, l'intervento umano sarà sempre necessario ma la tecnologia ci stupirà, anche nella metadattazione. I processi di metadattazione, infatti, sono supportati dalle macchine; il web semantico, i linked data e, dunque, la metadattazione nascono per "erudire" le macchine. In questo senso è interessante il dibattito in corso sull'intelligenza artificiale (AI), prospettiva che introdurrà, come in parte ha già introdotto, un nuovo paradigma del lavoro.

L'indicizzazione compiuta con l'ausilio della macchina non deve far pensare a una deresponsabilizzazione del catalogatore; è l'essere umano che "insegna" alla macchina il significato dei singoli dati. La metadattazione mantiene e rafforza la caratteristica distintiva di essere *in primis* un'attività culturale che implica la conoscenza della risorsa (che cos'è, di chi è) e, quindi, un'attività tecnica (che implica la conoscenza dei linguaggi professionali contemporanei). In definitiva, mutano ancora una volta le modalità delle operazioni da compiere (quante volte sono cambiate dai tempi di Callimaco a oggi?), ma non cambia il significato profondo della catalogazione in quanto operazione di mediazione consapevole tra la collezione e il lettore. Il passaggio alla metadattazione consente alle biblioteche di superare il concetto tradizionale di catalogo in quanto elenco ordinato secondo un preciso principio (per autore, per titolo, per classe, per soggetto ecc.) e sulla base di una struttura predeterminata (le aree ISBD) per arrivare a una modalità

più aperta e flessibile di presentazione e accesso alle informazioni. Esplicative sono le parole di Diego Maltese nella *Prefazione* al volume *Metadattazione* già citato: “La metadattazione è la catalogazione nell’era in cui tutto è connesso. La tradizione catalografica viene confermata nella sua essenza proprio nel momento in cui aggiorna e rafforza le sue tecniche per permettere che il catalogo non sia solo *nel* web, ma che diventi *del* web, ovvero impieghi il linguaggio della contemporaneità, come ha sempre fatto nella storia” (Maltese 2022, 11). Queste parole contengono la filosofia della catalogazione contemporanea.

L’evoluzione della terminologia

Le novità dell’universo bibliografico hanno comportato conseguenze sul piano terminologico; alcune locuzioni hanno acquisito una valenza semantica diversa (come succede in ogni lingua viva) e sono stati coniati neologismi. Si usano sempre meno termini quali *documento*, *catalogazione*, *record* e sempre più parole come *agente*, *creatore*, *dato*, *dataset*, *discovery*, *elemento*, *entità*, *formato*, *istanza*, *item*, *metadati*, *risorsa*, *tag*, *trippla*. Il termine *catalogazione* continua a essere impiegato in molte realtà con significato più ampio rispetto al passato; sono comparse nuove locuzioni quali *metadattazione*, *data manager*, *data librarian*, *data curation librarian*, *data reference librarian*, *data collection building librarian*, *data reference librarian*, *chief data officer*, *metadata strategy* (e il relativo *metadata strategist*). La nuova terminologia è mutuata perlopiù dall’ambito informatico e dai principi del management.

In Europa e in altre parti del mondo, soprattutto oltreoceano, i metadati sono già parte della realtà quotidiana da decenni; il cambiamento concettuale e terminologico ha conseguenze sul lato gestionale. Nelle biblioteche non si trovano più uffici denominati Dipartimento catalogazione, bensì, per esempio, alla Bibliothèque nationale de France (BnF), *Département des Métadonnées*; alla British Library, inizialmente *Collection and Metadata Processing* e adesso, più semplicemente, *Metadata*; alla Stanford University Libraries, *Metadata Department* e in numerose biblioteche americane e canadesi *Metadata Services*. Molti catalogatori, soprattutto americani, amano definirsi *metadata creation manager*, come al Dipartimento della Harvard Library Information and Technical Services. *Metadata editor* è il nome dello strumento con il quale vengono visualizzati e modificati i metadati bibliografici e d’autorità in Alma, il sistema bibliografico di Harvard. Dal 2015 le tre sezioni IFLA Bibliography, Cataloguing e Subject Analysis and Access pubblicano *IFLA metadata newsletter*. Gli editor della newsletter si rivolgono tra loro usando l’espressione *Metadata colleagues*. Alla Bodleian Library di Oxford è stato aperto il servizio di reference *Ask a data librarian* all’interno della biblioteca digitale Bodleian Data Library.

Nell’era digitale si assiste alla disintermediazione nei processi di trasmissione della conoscenza registrata e vengono ridefiniti i confini delle biblioteche, le competenze del bibliotecario e la terminologia utilizzata. La metadattazione è il risultato di una rivoluzione che negli ultimi due decenni ha investito per intero il mondo delle biblioteche: è cambiato l’universo bibliografico, sono cambiate le tecnologie a disposizione e, di conseguenza, è mutata la terminologia per esprimere nuovi concetti e nuove tipologie di risorsa. La nuova terminologia evidenzia il cambiamento avvenuto nell’approccio alla rappresentazione delle collezioni delle diverse istituzioni culturali, biblioteche in particolare. Per esempio, in una fase di nuove formulazioni teoriche e di avvio di una

delicata, difficile e non lineare transizione verso nuove tipologie di catalogo, i modelli concettuali bibliografici della famiglia FR (*Functional Requirements*), dal primo del 1998 FRBR (*Functional Requirements for Bibliographic Records*), FRAD (*Functional Requirements for Authority Data*) e FR-SAD (*Functional Requirements for Subject Authority Data*) all'ultimo del 2017 IFLA LRM, *Library Reference Model* – tutti emanati dall'IFLA – hanno rappresentato e rappresentano un punto di riferimento basilare per l'analisi dei metadati delle risorse bibliografiche e per la terminologia usata o introdotta (Pisanski e Žumer 2010). Emblematico il passaggio dal titolo *Functional Requirements for Bibliographic Records* a *Functional Requirements for Authority Data*, un mutamento terminologico (da *record* nel titolo di FRBR si passa a *data* in quello di FRAD) che denota il cambiamento concettuale avvenuto a cavallo tra XX e XXI secolo; i modelli bibliografici non si riferiscono più al record testuale ma ai singoli dati. Il dato è l'elemento autonomo, in sé autoconsistente, riutilizzabile in contesti diversi; apparentemente il record sembra lo stesso ma la sua struttura è diversa: testuale la prima, dataset la seconda.

La *Dichiarazione di Principi internazionali di catalogazione* (ICP) ha deliberatamente escluso enunciati usati per molti decenni, come *unità bibliografica*, *intestazione*, *rinvio* e *titolo uniforme*, ritenuti inadeguati al nuovo contesto e perciò sostituiti rispettivamente da *manifestazione*; *punto d'accesso autorizzato*; *punto d'accesso controllato*; *forma variante del nome*; *forma autorizzata del nome*; *nome*. Nell'elenco dei termini non più in uso indicati in ICP2016: *Authority record*, inizialmente tradotto come *registrazione d'autorità*,³ viene sostituito da *Authority data*. *Bibliographic record*, reso in italiano sia con *registrazione bibliografica* sia con *scheda*, viene sostituito da *Bibliographic data*. La pubblicazione di *ISBD Preliminary* del 2007 (ICCU, 2009) e di *Consolidated Edition* del 2011 (ICCU, 2012) ha portato a un aggiornamento terminologico legato alla necessità di fondere le precedenti ISBD; per esempio, la formula *non-book material* non è stata mantenuta in quanto ritenuta generica.

I lemmi *documento* e *risorsa* costituiscono un esempio di evoluzione terminologica. *Documento* è uno dei termini maggiormente polisemici e possiede un'essenza poliedrica e transdisciplinare, dato che per anni è stata una parola ampiamente usata in numerosi contesti disciplinari, fra cui quello biblioteconomico; il suo ambito d'impiego, infatti, è pressoché illimitato. Il termine *documento* è stato usato in un'accezione estesa a designare tutti i veicoli d'informazione e conoscenza a cui dà accesso una biblioteca o un centro di documentazione. Il suo uso, tuttavia, tende a essere ristretto al contesto archivistico e il termine è sostituito in Biblioteconomia da *risorsa bibliografica* o semplicemente *risorsa*. Il lemma *risorsa* denota un'entità, analogica o digitale, che comprende un contenuto intellettuale o artistico (o entrambi), concepita, prodotta e pubblicata come un'unità distinta; designa tutte le possibili forme e tutti i possibili veicoli di conoscenza, ovvero tutti i tipi di risorsa conservati nelle biblioteche e nelle altre istituzioni culturali della memoria registrata: manoscritti, libri antichi e moderni, e-book, periodici, musica notata ed eseguita, immagini fisse e in movimento (per esempio, un videodisco, un videogioco, un film), carte geografiche, fotografie, registrazioni sonore, documenti d'archivio, oggetti d'arte e reperti di ogni tipo nella dimensione analogica e digitale. L'utilizzo del lemma *risorsa*, adottato a livello internazionale dai primi anni del Duemila, consente un dialogo interdisciplinare poiché s'inserisce perfettamente nel contesto

³ In italiano il termine record convive con registrazione.

del web semantico, nel quale i dati costituiscono una ricchezza, che siano creati da una biblioteca, da un museo, da un archivio, da un'amministrazione ecc. Il termine *risorsa* implica una visione più ampia e inclusiva della funzione dei dati di uno strumento di reperimento, ed è utile e utilizzabile perché trasversale; esso consente di travalicare il confine di una specifica istituzione culturale (Bianchini 2022, 131). Il lemma *risorsa* delinea tutte le entità possibili dentro il contesto del web dei dati. In RDA (*Resource Description and Access*), lo standard internazionale concepito per l'era digitale, implementato dal 2013 in numerose biblioteche europee e del mondo, *risorsa* compare già nel nome e rappresenta tutte le tipologie di mezzi informativi utili a soddisfare le necessità dei lettori. Il lemma, inoltre, ha avuto ampio impiego nei principi, negli standard, nei modelli concettuali e anche negli acronimi, per esempio, in URI, *Uniform Resource Identifier*; URL, *Uniform Resource Locator* e RDF, *Resource Description Framework*. RDA è il primo standard emanato per la metadatozione e la scoperta delle risorse e dialoga con il web anche sul piano terminologico; in linea con l'ambiente digitale entro il quale opera, RDA elimina termini tradizionali quali *heading* (intestazione) e *main entry* (intestazione principale) a favore di *access points* (punti d'accesso), come stabilito da ICP nel 2009.

L'universo bibliografico, formulazione entrata nel linguaggio biblioteconomico alla fine del XX secolo e codificata da FRBR, è ben definito da Carlo Bianchini (Bianchini 2005): e può essere delineato come l'insieme delle conoscenze registrate su qualsiasi supporto trasmesse nel corso del tempo dalla comunità umana e accessibili nelle biblioteche, negli archivi, nei musei e via web. L'universo bibliografico comprende qualsiasi tipologia di risorsa nella sua dimensione analogica e digitale, nonché le persone e le istituzioni che ne sono responsabili o che svolgono un ruolo rispetto a esse. La comparsa di nuove varietà di risorse tipiche del mondo digitale e di nuovi modi di leggere i testi, di vedere le immagini e di ascoltare la musica produce costanti cambiamenti: le tecnologie editoriali progettate negli ultimi decenni, per esempio, danno la possibilità di fruire la medesima opera in molteplici versioni, formati o presentazioni: come volume a stampa, audiolibro, e-book (pdf, html, epub, mobi) o *libro digitale*, una forma di libro che potremmo definire *bidimensionale*, in quanto costituito dal contenuto e dai suoi metadati che risultano indissolubilmente associati come due facce della stessa medaglia. Il rapporto che si è instaurato tra il contenuto di un'opera e i suoi metadati richiama l'immagine di un iceberg, come ha evocato Fulvio Guatelli: il contenuto è costituito dalla parte emergente del blocco di ghiaccio, mentre i metadati sono rappresentati dalla parte sommersa che consente il galleggiamento della parte emersa (Guatelli 2020).

L'authority control, importanza ed evoluzione

Il controllo d'autorità è il processo che consente la creazione di record d'autorità che includono i dati necessari all'identificazione univoca di un'entità all'interno di un catalogo o di una banca dati. Grazie all'impiego di identificatori univoci e persistenti e alla funzione di *clustering*, qualsiasi forma di un nome può costituire un punto d'accesso all'entità. Questi nomi, così come gli altri metadati associati a una risorsa, diventano punti d'accesso per arrivare all'entità descritta, che può essere un agente, un'opera, un concetto ecc. La scelta di una forma preferita da un'agenzia bibliografica è legata al contesto culturale e linguistico; non esiste, infatti, una forma valida a livello internazionale. In questo senso il concetto di titolo uniforme cade a favore di titolo preferito,

preferito in un contesto linguistico e culturale. Il controllo d'autorità supporta, pertanto, i compiti del catalogo o della banca dati ed è applicato a tutte le entità: agenti, opere, luoghi e soggetti.

Il nome di un'entità può essere conosciuto in diverse lingue e scritture; l'insieme di queste forme varianti è collegato a un identificatore, ossia una stringa numerica o alfanumerica di caratteri associati in modo univoco e permanente a una risorsa per garantire l'accesso stabile a essa nell'ambito di un medesimo dataset. La macchina utilizza quell'identificatore per arrivare alla descrizione della risorsa, consentendo così al lettore la libertà d'impiegare, nel corso della ricerca, una tra le forme incluse nel medesimo *cluster*.

Dichiarare una delle forme di nome come punto d'accesso autorizzato per un'entità semplifica i sistemi, ma potrebbe rivelarsi inutile con i sistemi futuri. La presenza di un'unica forma di nome preferita era un requisito delle vecchie regole di catalogazione e dei primi sistemi per garantire il raggruppamento dei record bibliografici relativi a un'entità. I primi sistemi non erano ancora abbastanza sofisticati da consentire all'utente di preferire una forma piuttosto che un'altra per soddisfare le sue esigenze. Alcuni non lo sono tuttora. Tuttavia, oggi le varie forme di nome delle entità possono essere registrate nei sistemi e utilizzate per visualizzare una singola forma di nome preferita dalla biblioteca (il suo punto d'accesso autorizzato per quell'entità) o una forma preferita dall'utente. Registrando in un record d'autorità le diverse forme del nome, insieme con gli altri attributi che identificano un'entità, aumenta il potere descrittivo che consente di identificare in modo inequivocabile la risorsa. Il risultato del nuovo corso che l'authority services sta prendendo non è collegare più record bibliografici, bensì consentire, assemblando dati diversi, spesso derivanti da fonti diverse, una maggiore identificazione delle risorse e delle entità a esse associate. La creazione di un accesso preferito innesca un ciclo virtuoso per la produzione di dati di qualità. La qualità dei punti d'accesso è determinata dall'autorevolezza di chi stabilisce le relazioni, i singoli collegamenti (Danskin 2013).

Le biblioteche si sono mostrate consapevoli dell'importanza degli identificatori fin dalla loro comparsa. Essi svolgono un ruolo fondamentale nei processi di controllo della qualità dei dati. Inoltre, a seguito del progetto del Controllo bibliografico universale (UBC), è apparsa evidente l'impossibilità per le agenzie bibliografiche nazionali di adottare una forma comune del punto d'accesso per la medesima entità; l'obiettivo della condivisione internazionale dei dati d'autorità e del controllo bibliografico universale può essere perseguito oggi collegando le forme varianti in un authority file virtuale internazionale e associandole a un identificatore univoco che consenta la condivisione dei dati d'autorità provenienti da fonti differenti e la visualizzazione della forma preferita dal lettore. Il concetto di authority control, pertanto, è mutato radicalmente nel corso degli ultimi due decenni, acquisendo sempre più importanza nella gestione dei cataloghi. Nel microcosmo della meta-database, la trasformazione è stata totale, producendo una maggiore complessità della gestione dei dati bibliografici granulari, come richiesto dal web semantico, con iniziative di aggregazione e scambio di dati a livello globale (come VIAF) e non solo all'interno di dati bibliografici (come *Wikidata*).⁴

⁴ <http://tinyurl.com/b5tb6rds>; Pierfranco Minsenti, *Condividere l'authority control con Wikipedia e Wikidata. La nuova vita dei dati prodotti dalle biblioteche. La nuova vita dei dati prodotti dalle biblioteche*, corso per l'AIB Toscana, Firenze, 2-3 marzo 2015, <http://tinyurl.com/wahazuz8>.

Per chiudere

In una visione sempre più globale, la partecipazione propositiva alle attività IFLA è decisiva in quanto ente normativo, le cui decisioni ricadono su tutte le biblioteche del mondo; una cosa è, difatti, interpretare uno standard o il testo di un *Manifesto*, altra cosa è contribuire a elaborarlo e a comprendere le motivazioni di ciascuna scelta compiuta.

La dimensione internazionale, l'aggiornamento continuo e il confronto fra colleghi sono requisiti tradizionali di ogni disciplina. La Biblioteconomia sta vivendo una fase di profonda trasformazione, come ricorda Michael Gorman ne *I nostri valori, rivisti* (Gorman 2018) e, pertanto, occorre che le biblioteche e i bibliotecari si riposizionino, com'è avvenuto innumerevoli volte nel corso della loro storia, dalla Biblioteca d'Alessandra e da Callimaco in poi.

Riferimenti bibliografici

- AIB (Associazione italiana biblioteche). Gruppo di studio catalogazione ed indicizzazione. 2020. *Manifesto per il web semantico: la comunità dei dati*. <http://tinyurl.com/ynjy7tzx>.
- Bianchini, Carlo. 2005. *Riflessioni sull'universo bibliografico: funzioni, oggetti e modelli della catalogazione per autore e per titolo*. Prefazione di Mauro Guerrini. Milano: Sylvestre Bonnard.
- Bianchini, Carlo. 2022. "Intervista." In *Metadattazione: la dimensione contemporanea della catalogazione*. A cura di Mauro Guerrini. Milano: Editrice Bibliografica.
- Danskin, Alan. 2013. "Linked and open data: RDA and bibliographic control." *JLIS.it*, 4 (1): [147]-159. <http://tinyurl.com/mrxu2wts>.
- Desideri, Laura, e Alberto Petrucciani, a c. di. 2014. *Le biblioteche di Luigi Crocetti: saggi, recensioni, paperoles*. Roma: Associazione italiana biblioteche.
- Geltner, Guy. 2015. "Upon leasing Academia.edu." *Hypotheses* (blog), December 7, 2015, <http://tinyurl.com/3jtn6xmd>.
- Gorman, Michael. 2018. *I nostri valori, rivisti. La biblioteconomia in un mondo in trasformazione*. Firenze: Firenze University Press.
- Guatelli, Fulvio. 2020. "FUP Scientific Cloud e l'editoria fatta da studiosi." *Società e storia*, 167 (1): 155-164. <http://tinyurl.com/3bbarjhv>.
- IFLA (International Federation of Library Associations and Institutions), e Unesco (United Nations Educational, Scientific and Cultural Organization). 2022. *Manifesto IFLA/Unesco sulle biblioteche pubbliche*. <http://tinyurl.com/2a7m98ce>.
- Maltese, Diego. 2022. "Prefazione." In *Metadattazione: la dimensione contemporanea della catalogazione*. A cura di Mauro Guerrini. Milano: Editrice Bibliografica.
- Pisanski, Jan, e Maja Žumer. 2010. "Mental models of the bibliographic universe. Part 1: mental models of descriptions; Part 2: Comparison task and conclusions." *Journal of documentation*, 66 (5): 643-680.
- Scuola di Barbiana. 1967. *Lettera a una professoressa*. Firenze: Libreria Editrice Fiorentina.
- Serrai, Alfredo. 1973. *Biblioteconomia come scienza: introduzione ai problemi e alla metodologia*. Firenze: Olschki.
- Shera, Jesse Hauk. 1965. *Libraries and the organization of knowledge*. London: Crosby Lockwood and Son Ltd.

The convergence of Artificial Intelligence and Digital Skills: a necessary space for Digital Education and Education 4.0

Miguel-Ángel Marzal^(a), Maurizio Vivarelli^(b)

a) Carlos III University, Department of Documentation, <https://orcid.org/0000-0003-2039-234X>

b) University of Turin, <https://orcid.org/0000-0002-9328-094X>

Contact: Miguel-Ángel Marzal, mmarzal@bib.uc3m.es; Maurizio Vivarelli, maurizio.vivarelli@unito.it

Received: 30 May 2023; **Accepted:** 13 November 2023; **First Published:** 15 January 2024

ABSTRACT

An analysis study of the state of the art on the convergence process first, then confluence, final symbiosis between Artificial Intelligence (AI) and Digital Competences (CD) for Digital Education and 4.0 is presented, following a descriptive and diachronic method to be able to analyze from different points of view the complexity, problems and opportunities that this process implies, trying to present, in each of the phases of study of this phenomenon, the models, experiences and lines of research that illustrate the educational impact of this symbiosis. The study proceeds to analyze:

- 1) first, the characteristics of AI and the drivers that have led to its impact on Education (Educational Artificial Intelligence),
- 2) the challenges that this impact on Education has brought about,
- 3) and the initiatives of political, social, and educational agents to assimilating the effects of AI in educational innovation.
- 4) Then the analysis stops at the definition, characters, properties and initiatives of the DC in Education. Analyzed in pairs, a detailed analysis is made of the symbiosis process that can promote Education 4.0: the changes in the behavior of the agents in the educational process, the tools and good practices that lead to an effective use of this symbiosis, the progress of intelligent technologies in Education, the enunciation of own educational objectives, which are generating new educational models and also the proposal of new evaluation systems.

KEYWORDS

Artificial Intelligence; Digital Education; Multiliteracies; Digital Literacy; Virtuality in Education; Smart Technologies in Education.

La convergencia de la Inteligencia Artificial y las Competencias Digitales: un espacio necesario para la Educación Digital y la Educación 4.0

RESUMEN

Se presenta un estudio de análisis del estado del arte sobre el proceso de convergencia primero, confluencia luego, simbiosis final entre la Inteligencia Artificial (IA) y las Competencias Digitales (CD) para la Educación Digital y 4.0, siguiendo un método descriptivo y diacrónico para poder analizar desde diferentes puntos de vista la complejidad, problemas y oportunidades que implica este proceso, tratando de presentar, en cada una de las fases de estudio de este fenómeno, los modelos, experiencias y líneas de investigación que ilustran el impacto educativo de esta simbiosis. El estudio procede a analizar: primero los caracteres de la IA y los motores que han llevado a su impacto en Educación (Educativa Artificial Intelligence), los desafíos que ha supuesto este impacto en Educación y las iniciativas de los agentes políticos, sociales y educativos para asimilar los efectos de la IA en la innovación educativa. Luego el análisis se detiene en la definición, caracteres, propiedades e iniciativas de las CD en Educación. Analizado en binomio, se hace un análisis pormenorizado sobre el proceso de simbiosis que puede impulsar la Educación 4.0: los cambios en los comportamientos de los agentes en el proceso educativo, las

herramientas y buenas prácticas que conducen a un uso eficaz de esta simbiosis, el progreso de las tecnologías inteligentes en Educación, la enunciación de objetivos educativos propios, que están generando nuevos modelos educativos y también la propuesta de nuevos sistemas de evaluación.

PALABRAS CLAVE

Inteligencia Artificial; Educación Digital; Multialfabetizaciones; Alfabetización Digital; Virtualidad en Educación; Tecnologías Inteligentes en Educación.

1. La convergencia de la Inteligencia Artificial y las Competencias Digitales como factores hacia la Sociedad del Conocimiento: la Educación

La IA y CD comenzaron un proceso de convergencia cuando la IA ha comenzado a percibirse como un factor de indudable potencialidad formativa: si las CD ya eran reconocidas por los actores socioeconómicos como objeto de estudio y aplicación profesional y educativa, la IA iniciaba su andadura hacia la Educación como disciplina académica por su función beneficiosa contra brechas digitales y disfunciones éticas.

1.1. La Inteligencia Artificial y su dimensión educativa: una visión más global

En su informe anual, la empresa McKinsey Global Survey, en 2022, hace un análisis desde 2017 sobre la evolución y las estrategias de aprendizaje por machine learning y deep learning: el “transfer learning” abre un espacio en las capacidades que las empresas desarrollan con la IA, como factor educativo. Desde esta perspectiva, parecía plausible analizar las potencialidades de IA en Educación, como desde 2019 ha hecho UNESCO, en el documento *Recomendación sobre la Inteligencia Artificial* de 2021, entendida como apoyo a nuevos modos de aprender y ejercitar competencias computacionales, para una Educación de calidad, mediante su objetivo de integrar modelos y algoritmos que generan capacidades para aprender y realizar tareas cognitivas. Sin embargo, las potencialidades de la IA, descubiertas según se van implementando diferentes herramientas y aplicaciones, convierte la funcionalidad educativa de la IA en un campo de análisis necesario para desarrollar la *Educación Digital*. Las primeras aplicaciones de la IA se refirieron a procesos útiles en la gestión académica de las instituciones educativas (previsión de inscripciones, cálculo de abandono, motivaciones, diseño de perfiles, tutoración inteligente, etc.), con un predominio de métodos y análisis cuantitativos. Sistemas como Data Mining, Logistic Regression o Random Forest han permitido cualificar mucho la práctica en el proceso educativo.

Sin embargo, el impacto de la IA pronto se ha hecho más sofisticado. Está siendo un aliado relevante en la incorporación de arquitecturas basadas en la nube, implementadas por técnicas de deep learning (Algayres y Triantafyllou 2020) y estilos de aprendizaje (“analítica de aprendizaje” y “aprendizaje adaptativo”) a partir de las funcionalidades del machine learning (Vargas et al. 2022). Las potencialidades de la IA en Educación han suscitado un debate respecto a sus efectos beneficiosos sobre los procesos educativos y de investigación (Cukurova, Luckin, y Kent 2020), los comportamientos de los estudiantes o sobre experiencias cooperativas con las Realidades Mixta, Virtual (RV), Aumentada (RA), y el *gaming*. Tuomi (2018) ve más adecuado analizar el modo de encauzar y pautar el ritmo, dinámica y nivel de profundidad de los efectos y cambios. Por su parte, I. García-Martínez et al. (2023) han hecho una exposición sobre esta literatura científica del fenómeno, compilando desde Roll y Wylie en 2016, en 2019 a Hooshyar y Lim, o Zawacki-Richter, Bond y Gouverneur, en 2020 a Song y Wang, demostrando el crecimiento notorio de la producción científica sobre la EAI (Educational Artificial Intelligence).

1.1.1. Los campos de debate de la IA en Educación

Parece necesario un breve esbozo de los desafíos a los que puede responder la IA, pero también

determinar los desafíos que la propia IA está planteando en Educación:

- a) La *Humanidad mejorada*, una etiqueta tan perturbadora como equívoca. Barrios-Tao, Diaz y Guerra (2021) la presentan con una doble dimensión: cambiar el concepto y significado de humano (condición y naturaleza); fomentar, con la IA como factor, la cohesión social a partir de la autodeterminación individual, según Floridi et al. (2018), mediante el *pensamiento crítico* (uno de cuyos factores de desarrollo son las multialfabetizaciones) y el aprendizaje permanente, pero además repensando la situación del ser humano en los nuevos entornos respecto a sí mismo, a los otros, a la comunidad. La *autodeterminación individual*, debe basarse en el pensamiento crítico y consciencia de la propia identidad e imagen web, por conocimiento del funcionamiento, aplicaciones y posibilidades de los algoritmos y big data, si bien son competencias que exigen una gobernanza y transparencia digitales. Se abre paso, pues, la filosofía transhumanista, tal como se definía en el Manifiesto Transhumanista de 1999.
- b) La *inclusión social*, derivada de los “derechos emergentes”, íntimamente relacionados con los derechos afectados por las tecnologías, a causa de la descentralización de editores de contenidos, la horizontalidad en la capacidad de regular los controles algorítmicos, una nueva gobernanza digital que desprotege al individuo, una amenaza cierta sobre los recursos que la IA absorbe de la Naturaleza y un nuevo concepto de “pobreza algorítmica”. Todo ello justificó la organización de un Foro por parte de la Unión Internacional de Telecomunicaciones, *La Inteligencia Artificial para el Bien del Mundo*, en 2017, en relación directa con los ODS.
- c) Una *formación integral*, frente a la superficialidad de una información sobreabundante. La IA debe fundamentar un *pensamiento computacional*, una competencia para la que Barr y Stephenson (2011), además de definirlo, proponían unos iniciales indicadores competenciales que midiesen la recolección, análisis y representación de datos, la estructuración de problemas, capacidad de abstracción, y de comprensión de algoritmos. Por ello conviene comprender e interpretar el funcionamiento y comportamiento de los agentes de la IA (Flogie y Aberšek 2021) y su aplicación en el desarrollo educativo de competencias transversales (Crisol-Moya et al. 2020).
- d) *Usos éticos de la IA en contextos educativos*, problema que ya evidencia algunas experiencias con robots. El problema es que las herramientas de la IA no pueden comportarse sólo como un medio e instrumento, sino pueden realizar tareas cognitivas. Selwyn (2019) hace un riguroso examen sobre sus oportunidades o amenazas. El uso experto se asegura mediante una *alfabetización algorítmica*, sobre cuyas competencias y su naturaleza han reflexionado Long y Magerko (2020).
- e) *La autoría y propiedad intelectual a partir de la IA*. El debate paltea si la IA puede ser reconocida como sujeto en la propiedad intelectual. Villalobos (2022) hace una exposición metódica, glosando recientes sentencias en sentidos opuestos. El problema es complicado porque se refiere a un fenómeno universal (IA), pero se tratan de dar respuestas concretas (en cada pleito), teniendo como condicionante las legislaciones nacionales, si en las sentencias la noción de autoría une o separa los conceptos de propiedad intelectual y propiedad industrial, si se trata de autoría para personas físicas o jurídicas, si afecta al soporte de la creación, o a su originalidad. De hecho, en las “Normas de Derecho Civil sobre Robótica”,

el Parlamento Europeo por la Resolución de 16 de febrero de 2017, ya sugirió en su momento la noción de “personalidad electrónica”, luego desarrollada y, por tanto, enunciada de forma más precisa y compleja.

- f) *La interdisciplinariedad*, por cuanto la IA, entre otros diferentes factores que impulsan este fenómeno tan relevante para la organización del conocimiento, pone a disposición nuevas herramientas y métodos que permiten un activo y eficaz trabajo cooperativo. Un espacio ejemplificador puede ser el caso de las *Humanidades Digitales*, en las que se producen diversas iniciativas y estrategias en los métodos y técnicas de investigación y estudio, como pueden sugerir García-Serrano y Menta-Garuz (2022), y las técnicas de Procesamiento de Lenguaje Natural (Mcgillivray, Poibeau, y Ruiz 2020), junto con los análisis de datos.

1.1.2. *El impulso de los agentes socio-económicos y educativos para la dimensión educativa de la IA*

Parece evidente que la IA va a transformar nuestro entorno. Por esta razón la ISO acogía un subcomité, la ISO/IEC JTC/1 SC/42, para desarrollar un programa de normalización en el uso de la IA, así como orientar a los comités JTC 1, IEC e ISO en el desarrollo de aplicaciones de la IA. Sobre este interés aplicativo, la OCDE, asociada con International Business Machines (IBM), apuntaba en 2018 el potencial de uso de la IA por parte de los agentes sociales, en el informe *IA: máquinas inteligentes, políticas inteligentes*, un esfuerzo que también impulsaba la Unión Internacional de Telecomunicaciones (UIT) en dos informes, *Serie de Inteligencia Artificial para el Desarrollo*, en 2017, y *AI for Good Summit*, los años 2017 y 2018. Este evento es anual y, significativamente, el que se ha celebrado en Ginebra entre el 28 de marzo y 3 de abril de 2023, dos de sus cuatro sesiones, una se refiere a “Educational robots providing high-quality and inclusive education for all” y la otra a “How ChatGPT will change the classroom – teachers and students discuss”. Todos los actores políticos en el mundo han comenzado a movilizarse. Tecnología y Educación ya se enunciaban como una simbiosis imprescindible en la publicación de los Objetivos de Desarrollo Sostenible (ODS), dentro de los que el Objetivo 4 se denomina “Educación de Calidad” y que ha impulsado la *Agenda 2030*. Por su parte, UNICEF presentaba en noviembre de 2021 el informe *Policy guidance on AI for children Version 2.0. Recommendations for building AI policies and systems that uphold child rights*, alertando sobre el impacto directo sobre los niños, sin una regulación educativa.

La necesidad de impulsar los apartados 17 a 19 del Preámbulo de la Agenda Digital y el ODS 4 provocó la reunión, entre el 16 y 18 de mayo de 2019, de un foro mundial, que alumbraba el *Consenso de Beijing*, en torno a la IA y Educación. El Consenso se publica en forma de recomendaciones categorizadas: Planificación de la inteligencia artificial en las políticas educativas; La inteligencia artificial para apoyar la docencia y a los docentes; La inteligencia artificial para el aprendizaje y la evaluación del aprendizaje; Desarrollo de valores y competencias para la vida y el trabajo en la era de la inteligencia artificial; La inteligencia artificial para ofrecer oportunidades de aprendizaje permanente para todos; Promoción del uso equitativo e inclusivo de la inteligencia artificial en la educación; Inteligencia artificial con equidad de género e inteligencia artificial para la igualdad de género; Velar por el uso ético, transparente y verificable de los datos y algoritmos educativos; Seguimiento, evaluación e investigación.

La UNESCO publicaba en 2019 su informe *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*, que estructuraba en tres secciones: la Sección I, “Apro-

vechamiento de la IA para mejorar los resultados del aprendizaje”; la Sección II, “Preparar a los estudiantes para prosperar en un futuro saturado de IA”; la Sección III, “Desafíos e implicaciones políticas”. Las *competencias digitales y en IA*, se convierten en un factor nodal, y por esta razón, en diciembre de 2020, UNESCO promovía el International Forum on AI and the Futures of Education Developing Competencies for the AI Era, y propone una alfabetización en IA. El informe de UNESCO *K-12 AI curricula. A mapping of government-endorsed AI curricula* hace relación sobre las “técnicas de la IA” en Educación (Classical AI, Machine learning, Supervised learning, unsupervised learning, Reinforcement learning, Neural networks, Deep learning, General adversarial networks o GANs), “tecnologías IA” (Chatbots, Computer vision, Natural Language Processing, Sensors) y sus retos educativos para llegar a “Alfabetización en IA”.

También la Unión Europea (UE) en 2018 definía en su programa “Estrategia y Política”, una estrategia para la IA: en las “Prioridades” proyectaba desarrollar como una línea de acción “Excelencia y confianza en la IA”. En febrero de 2020 se publicaba el *Libro Blanco sobre la IA*, una senda que ha continuado el esfuerzo programático y normativo en los diferentes campos que pueden verse afectados por la IA.

1.2. Las Competencias Digitales como factor de progreso: una visión más europea

Como factor de progreso, las competencias debían proyectarse en Educación, mediante un *modelo educativo competencial*. El debate se trasladó a qué propiedades debían tener las competencias en Educación, referidas a una mayor cualificación profesional mediante las multialfabetizaciones. Entre ellas cobraba fuerza la *alfabetización digital*, encargada del desarrollo en competencias para conocer y saber por gestión experta de contenidos digitales y virtuales por dominio de las *competencias digitales*. La Unión Europea, desde la Agenda Digital Europea 2010, continuada por la Agenda Digital Europea de 2020, mediante la iniciativa *Brújula digital: el enfoque de Europa para el Decenio Digital*, del 9 de marzo de 2021, uno de sus cuatro objetivos programáticos fundamentales lo constituían las competencias digitales, publicando una dimensión formativa y educativa en el *Plan de Acción de la Educación Digital (2021-2027)* y un plan de comunicación mediante el *Año Europeo de las Competencias 2023*.

A partir de aquí la Comisión ha creado el Joint Research Centre, que desde 2005 impulsaba investigaciones sobre aprendizaje y habilidades en la Era Digital, y mediante el Joint Research Centre’s Publications, que mantiene y gestiona el sitio web JRC Publications Repository, registra publicaciones de carácter teórico, aplicativo, normativo y estudios de caso sobre las alfabetizaciones mediáticas e informacionales, junto con las competencias digitales. La actividad de DigComp ha tenido un mayor impacto en Educación. En 2016, la Comisión editaba a través del European Commission’s Joint Research Centre el *DigComp 2.0: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use*, completado en la versión 2.1 (Carretero et al. 2018), un documento actualizado en la versión 2.2. La preocupación por asegurar el mayor impacto educativo de las competencias digitales ha hecho atender a los educadores, mediante el *Marco DigCompEdu*. El contexto pandémico y la profunda crisis socio-económica consecuente han llevado a un serio replanteamiento de las competencias digitales (y la alfabetización digital) alcanzaban que un impulso poderoso por integrar muy pronto el dominio experto de las tecno-

logías para tratamiento de Big Data, Web Semántica, la Virtualización y la IA. El documento de DigComp se actualiza con otra nueva versión (Vuorikari et al. 2022), para acoger las competencias digitales necesarias referidas en la Web y la IA: *DigComp 2.2: The Digital Competence Framework for Citizens – With new examples of knowledge, skills and attitudes*. Se ha producido la confluencia de IA y Competencias Digitales en Educación.

2. La interacción simbiótica de la IA y las Competencias Digitales en Educación

No cabe duda de que en la actualidad existe una controversia muy importante en torno a los efectos de la IA, debate que busca ofrecer una perspectiva más sobre el impacto educativo de la IA, contemplado como vector en el progreso de unas competencias digitales imprescindibles para un progreso sólido, solidario y cohesivo.

2.1. Los medios y acciones innovadoras de la IA para Educación: unas competencias digitales apropiadas

Las iniciales conexiones educativas de las potencialidades de la IA se habían asentado sobre el objetivo de conceder una representación semántica a los contenidos. Bajo estas premisas las competencias son el nodo gordiano de todo el proceso. Las competencias digitales impactan en un comportamiento sobre los agentes del proceso educativo. Internet de las Cosas y los algoritmos que gestionan los datos masivos están ya alterando el comportamiento y la actitud de las personas. Respecto a los comportamientos de las *autoridades educativas* el informe para el Reino Unido de Baker, Smith y Anissa (2019) es ilustrativo: aconseja una política pública de inversiones I+D para IA, pero mediante implementación de herramientas para el educador; plantea generar un sitio web para instituciones especializadas sobre experiencias para crear un marco adecuado que permita el desarrollo de buenas prácticas; recomienda una política de gobernanza con un programa detallado de acciones e intercambio de datos masivos en Educación y un sistema sólido para medir el impacto de la IA en Educación. Respecto a los *docentes*, el informe analiza los beneficios sobre la actividad del profesor (uso de plataformas de aprendizaje personalizado y diferenciado, retroalimentación de las comunicaciones de los estudiantes, aprendizaje colaborativo, selección de material didáctico, liberación de tareas mecánicas, tener un espectro del progreso competencial del estudiante, innovación educativa). Sin embargo, es entre los docentes donde se encuentran las mayores objeciones en el uso de IA, por la posibilidad de sustituir el ser humano en el proceso educativo, temor basado por técnicas de reconocimiento emocional (Imani y Montazer 2019), emulando la emoción en el procesamiento de lenguaje natural (Wei y Zao 2019) y profesores robots. Los *estudiantes* son el actor con mayor impacto de la IA. Fortalece los aprendizajes múltiples, permite ejercitar el pensamiento computacional y el aprendizaje colaborativo (Yadav et al. 2016), en plena sintonía con al aprendizaje personalizado y adaptativo (Deng, Benckendorff, y Gannaway 2019; Vilkova y Shcheglova 2021). La Declaración PRISMA se convierte en una medición del progreso académico y de aprendizaje, como presentan Page et al. (2021), a modo de un estado del arte.

Calidad y excelencia del rendimiento y eficacia de aprendizaje mediante las aplicaciones de la IA,

se ha convertido en el fundamento de la denominada *Educación 4.0*, un término utilizado para denominar el vínculo entre tecnologías nuevas e integrales y su impacto en Educación Superior, mediante implementaciones y experiencias cooperativas en el uso de datos masivos entre instituciones como FEEDAN, según presentan Guo, Zeng y Dong (2020), una técnica de aprendizaje automático (MLT), que abordan Ciolacu et al., la AI-assisted Higher Education System (AIHEF) para analizar el comportamiento educativo, o el Hybridized Deep Neural Network (HDNN), que busca predecir el rendimiento del estudiante.

La Educación 4.0 está poniendo el foco de atención en el *aprendizaje*, para el que la EAI propone diferentes modelos: a) el *Machine Learning* y *Deep Learning*, que reconoce tres tipos de aprendizaje, dependiendo del fenómeno y la estructura de sus datos, “aprendizaje supervisado”, donde el algoritmo aprende desde las variables explicativas asociadas a una variable respuesta, para luego predecir el valor de esta variable (salidas etiquetadas), “aprendizaje no supervisado”, que predice por identificación de patrones a partir de similitudes, muy útil para una fase de indagación, “aprendizaje de refuerzo”, basado en acierto-error, un panorama mucho más complejo para el Deep Learning, que funciona basando el aprendizaje de los algoritmos en múltiples niveles de representación de datos, una jerarquía en cascada y que sigue modelos de redes neuronales; b) *Interacción aprendiz-interfaz*, por cuanto la IA impulsa la interacción aprendiz (estudiante)-interfaz (máquina), aspecto sobre el que Gunsekera, Bao y Kibelloh (2019) han presentado un necesario estado del arte, si bien un extenso elenco de investigadores argumentan, desde la óptica de la Tecnología Educativa, cómo esta interacción se está haciendo sin las aportaciones de esta disciplina, ni sobre modelos de diseño instruccional desde la Educación. Para ello se ha dibujado una estrategia de integración de la IA, mediante tres líneas: métodos de evaluación de usabilidad e instruccionalidad de los interfaces con finalidad educativa (Nathoo et al. 2020); implementar capacidades específicas de agentes inteligentes educativos para perfeccionar la comunicación interfaz humano-máquina en el dominio de la Educación (Koola, Ramachandran, y Vadakkeveedu 2016); aplicabilidad de datos masivos complejos en Educación para predicción de fenómenos (Baykal, Bulut, y Sahingoz 2018); c) *Los agentes de software conversacionales inteligentes (chatbots)*, con su aplicación más directa y eficiente en la tutoración virtual y recomendaciones sobre repositorios de objetos de aprendizaje, plataformas virtuales de aprendizaje, identificación de perfiles y grupos, chats asistidos, evaluación de código fuente, visualizaciones. Se ha propuesto un método para seleccionar el modelo y sistema más pertinente, como abordan Bajaj y Sharma (2018).

2.2. Tecnologías Inteligentes para Educación: los espacios simbióticos

Parece que el espacio simbiótico de cooperación de la IA y las competencias digitales encuentra su mejor campo en las arquitecturas basadas en la nube (Educación en la nube), que permiten ensayar nuevas técnicas de aprendizaje, impulsando la investigación sobre Entornos Personales de Aprendizaje y Ecologías de Aprendizaje, como “frontera” necesaria en la innovación educativa y el aprendizaje adaptativo. Un panorama sobre el desarrollo de estos espacios simbióticos lo ofrece Jara y Ochoa (2020): los *Sistemas de enseñanza y aprendizaje adaptativos* más potentes si incorporan reconocimiento de voz e imagen para los docentes (Dorca 2015), pero también para los contenidos educativos (Nye 2015); *Sistemas de aprendizaje colaborativo*, para gestionar la dinámica en los

foros y chats para tareas, fomentar la capacidad argumentativa de los estudiantes e incorporar plataformas de juegos y videojuegos educativos (Westera et al. 2020), sistemas que se apoyan en la Analítica de Aprendizaje (Learning Analytics), un campo de investigación cuyos caracteres, propiedades y literatura científica aborda Rojas (2017). El objetivo educativo es un conjunto de competencias digitales muy transversales, para cuyo desarrollo en IA se están utilizando como primeros instrumentos:

a) *Robótica y apps*

Convertida en una estrategia para desarrollar el pensamiento computacional, el aprendizaje activo y las competencias STEM (acrónimo sobre desarrollo de competencias en Ciencias, Tecnología, Ingeniería y Matemáticas en sus siglas en inglés).

La Robótica Educativa se comporta como un recurso didáctico idóneo para las competencias STEM, así como las competencias en investigación y experimentación, pero también como centro del proceso educativo por sus efectos sobre el pensamiento computacional. Se ensaya también el desarrollo de estas competencias, con el auxilio de la Robótica Educativa, mediante dispositivos móviles, factor relevante para la Educación 4.0 (Restrepo-Echeverri, Jiménez-Builes and Branch-Bedoya 2022). En todo caso, la integración curricular del pensamiento computacional y las CD de la mano de la Robótica Educativa es controvertida por no estar aún bien estructurada ni definida, sin evaluación apropiada ni formación adecuada entre los docentes.

Aplicaciones se han hecho en todos los niveles educativos, si bien son más numerosas en Educación Superior tanto para plataformas (Ballesteros-Ballesteros et al. 2020), como apps.

b) *Las realidades extendidas y gamificación*

Las implicaciones de estas tecnologías inmersivas son evidentes (EDUCAUSE 2020). El paradigma del proceso educativo transforma completamente la experimentación investigativa, y el conocimiento consecuente, por el tratamiento y procesamiento de objetos físicos, luego transformados en objetos virtuales que compaginan los datos físicos, visuales y virtuales.

Las universidades han visto este fenómeno como una gran oportunidad, por lo que han comenzado a estructurar repositorios de ODE para su uso en los currícula académicos, porque permiten expandir el currículo académico y generar espacios simulados como laboratorios experimentales, muy potentes por su potencial visual, inmersivo e interactivo (Cabero-Almenara, Barroso-Osuna, y Gallego-Pérez 2018). El uso de estas realidades han impulsado también el *m-learning inmersivo*, como muestran Fombona, Pascual-Sevillano y González-Videgaray (2017).

La efectividad educativa de estas realidades debe mucho a su estrecha cooperación con la gamificación, que desarrolla habilidades y competencias útiles en el proceso educativo, tanto para los estudiantes, como para los docentes en su metodología didáctica. Una muestra de estos efectos metodológicos se encuentra en el Mobile Game-Based (Troussas, Krouska, y Sgouropoulou 2020). Existen, de nuevo, experiencias del uso combinado de técnicas de gamificación e IA para todos los niveles educativos. Para Educación Superior, González-Calatayud et al. (2021) han hecho una interesante revisión de literatura científica.

2.3. La confluencia: los objetivos educativos en un modelo para el siglo XXI

Esta evolución de equipamiento técnico y de tecnologías, tuvo en la primera década de este siglo un interesante acompañamiento desde la perspectiva pedagógica, mediante la propuesta de un modelo específico, el *Conectivismo*, que enunciaba G. Siemens en 2004. En realidad, el gran debate en torno a los objetivos educativos de la IA se encuentra en el tipo y métodos necesarios para el desarrollo de unas habilidades digitales idóneas para cuya obtención Scoular et al. (2017) proponen las herramientas de la IA. Son necesarios nuevos modelos educativos, como proponen Bonami et al. (2020):

- a) *Netnografía*, conjunto de estrategias metodológicas sobre el comportamiento comunicacional, cuyos postulados ha innovado Kozinets (2016). El análisis modelo busca comprender quiénes participan en redes y plataformas digitales y cómo se comportan en la producción del conocimiento, prediciendo el proceso educativo y creativo, y desarrollando el aprendizaje personalizado;
- b) *Modelo Gervais* (2016), basado en competencias, con un cierto enfoque “gerencial”, por cuanto la consecución de competencias se mide por resultados mediante la evaluación del esfuerzo;
- c) *Modelo de cuatro dimensiones*, de Fadel y Groff (2019), que busca superar un plan de estudios encorsetado por la estructura de áreas y especialidades de conocimiento propias y algo desconectadas del “mundo real”. Propone un plan multidisciplinario, que busque competencias individuales pero transversales, las cuatro C: comunicación, colaboración, criticismo y creatividad;
- d) *Modelo Brújula de la OECD* (2020), que propone un plan de estudios basado en: fundamentos básicos (alfabetización digital, salud mental y física, habilidades sociales y emocionales) competencias transformadoras (nuevos valores y responsabilidades), agencia y coagencia de estudiantes y ciclo de anticipación-acción-reflexión (aprendizaje continuo y permanente).

Las competencias, por su naturaleza, llevan implícita la necesidad de un sistema de evaluación propio, cualitativo. Por ello, los citados autores, proponen la *Multimodalidad*, referida a las distintas modalidades de canales de comunicación empleables: texto (documentación), voz (reflexiones), vídeos (diseños y planes), mediciones biológicas (comportamiento), clickstream (navegación y actitud digital). La multimodalidad permitiría proponer indicadores evaluativos cualitativos, que garanticen la *calidad*. Se hacen necesarios modelos evaluativos apropiados, sobre los que hacen un análisis metódico Marciniak y Gairin (2018), reflejando dos modelos de indicadores: los que se refieren a la pedagogía (objetivos y estrategias didácticas, actividades, materiales y recursos, evaluación, contenido y estructura del curso), los que se refieren a la infraestructura tecnológica (programas y plataformas digitales, modos de interacción, edición y comunicación).

El objetivo de todos estos modelos y sistemas es implementar una metodología que implemente un modelo educativo computacional con uso de la IA y con un alto contenido virtual y multimedia, para lo que se pretende analizar las aplicaciones más utilizadas y atender a las tendencias de los usos educativos de la IA.

3. Conclusiones y desarrollos futuros

El presente estudio de análisis se planteaba como objetivo declarado mostrar la necesaria simbiosis de la IA y las CD en Educación. Es, sin duda, una realidad poliédrica, que nos permite presentar tres categorías de conclusiones, probables futuras líneas de investigación:

1. La IA ha tenido un inevitable impacto en Educación, planteando efectos muy ambiciosos en los procesos educativos. Sin duda, la IA ha otorgado una primacía indiscutible a las CD, que ya no son sólo el objeto científico de disciplinas académicas (las *multialfabetizaciones*), sino que se han convertido en un factor ineludible para la Sociedad del conocimiento. Se abren paso así las Tecnologías Inteligentes en Educación, cuyo cometido será armonizar e impulsar sinergias entre todas las herramientas, experiencias, investigaciones aplicadas que proporcione la IA, que enriquece los fundamentos de la Educación Digital y sienta las bases para el Desarrollo de una Educación 4.0, para la que ya se están anunciando modelos pedagógicos y didácticos. Es significativo cómo la Educación 4.0 se está convirtiendo en un potente objeto de investigación científica.
2. Si la naturaleza y propiedades de la IA como ciencia y disciplina son claras y consideradas muy beneficiosas, sus aplicaciones son muy comprometidas y objeto de duda y sospecha. Se están solicitando moratorias en sus aplicaciones, se regulan restricciones a algunas de sus herramientas, se reclama una regulación y normativa específicas, actores en el Desarrollo de la IA alertan de sus peligros en muy distintos foros. Nuestro estudio, sin embargo, ha querido presentar una realidad paralela, desde otro prisma, sobre los efectos de la IA en Educación: debe entenderse no como una técnica capaz de generar artefactos con usos polifacéticos (provechosos o peligrosos), sino como una *tecnología* (una “ciencia de la técnica”), que genera modelos con vocación formativa y educativa para la Educación 4.0.
3. El impacto indudable en Educación y sus efectos realmente innovadoras mediante las CD ha planteado el reconocimiento de las competencias de IA dentro del espacio general de las competencias digitales. Las competencias de IA, sin embargo, son y deben ser especiales, por lo que los agentes políticos, sociales y educativos ya están organizando modelos para su aprendizaje (en Europa el Año de las Competencias), como presenta la versión DigComp 2.2, de 2022, con referencias explícitas a la IA. Es preciso comenzar a plantear el diseño de programas educativos específicos para el desarrollo e implementación de estas competencias.

Referencias bibliográficas

- Algayres, Muriel Gaele y Evangelia Triantafyllou. 2020. "Combining the flipped classroom and simulation games in engineering education: A methodological survey." In *SEFI 47th Annual Conference: Varietas Delectat... Complexity Is the New Normality, Proceedings*, 83-92. Budapest: European Association for Engineering Education (SEFI). ISBN 978-287352018-2
- Bajaj, Richa y Vidushi Sharma. 2018. "Smart Education with artificial intelligence based determination of learning styles." *Procedia Computer Science* 132: 834-842. <https://doi.org/10.1016/j.procs.2018.05.095>.
- Baker, Toby, Laurie Smith, y Nandra Anissa. 2019. *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. Londres: Nesta. https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf.
- Ballesteros-Ballesteros, Vladimir Alfonso, Oscar Ivan Rodríguez-Cardoso, Sébastien Lozano-Forero, y Jorge Luis Nisperuza-Toledo. 2020. "El aprendizaje móvil en Educación Superior: una experiencia desde la formación de ingenieros." *Revista científica* 38 (2): 243-257. <https://doi.org/10.14483/23448350.15214>.
- Barrios-Tao, Hernando, Vianney Díaz, y Yolanda Guerra. 2021. "Propósitos de la educación frente a desarrollos de inteligencia artificial." *Cadernos de Pesquisa* 51, <https://doi.org/10.1590/198053147767>.
- Baykal, Saide Isilay, Deniz Bulut, y Ozgur Koray Sahingoz. 2018. "Comparing deep learning performance on BigData by using CPUs and GPUs." In *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, 1-6. Istanbul: IEEE. 10.1109/EBBT.2018.8391429.
- Bonami, Beatrice, Luiz Piazzentini, y André Dala-Possa. 2020. "Educación, Big Data e Inteligencia Artificial: Metodologías mixtas en plataformas digitales." *Comunicar Revista Científica de Educación XXVIII* (65): 43-52. <https://doi.org/10.3916/C65-2020-04>.
- Cabero-Almenara, Julio, Julio Barroso-Osuna, y Óscar Gallego-Pérez. 2018. "La producción de objetos de aprendizaje en realidad aumentada por los estudiantes. Los estudiantes como prosumidores de información." *Revista Tecnología, Ciencia y Educación* 11: 15-46. <https://doi.org/10.51302/tce.2018.221>.
- Carretero Gomez, Stephanie, Yves Punie, Riina Vuorikari, Marcelino Cabrera Giraldez, Stefano Kluzer, Laia Pujol Priego, y William Okeeffe. 2018. *DigComp into Action: Get inspired, make it happen. A user guide to the European Digital Competence Framework*. Luxembourg: Publications Office of the European Union. <https://publications.jrc.ec.europa.eu/repository/handle/JRC110624>.
- Crisol-Moya, Emilio, Liliana Herrera-Nieves, y Rosana Montes-Soldado. 2020. "Virtual Education for All: Systematic Review." *Education in the Knowledge Society* 21: 1-13. <https://doi.org/10.14201/eks.20327>.
- Cui, Wei, Zhen Xue, y Khanh-Phuong Thai. 2019. "Performance comparison of an AI-Based Adaptive Learning System in China." In *Proceedings 2018 Chinese Automation Congress*, 3170-3175. Xian: IEEE. <https://doi.org/10.1109/cac.2018.8623327>.
- Cukurova, Mutlu, Rosemary Luckin, y Carmel Kent. 2020. "Impact of an Artificial Intelligence Research Frame on the Perceived Credibility of Educational Research Evidence." *International*

Journal of Artificial Intelligence in Education 30 (2): 205-235. <https://doi.org/10.1007/s40593-019-00188-w>.

Deng, Ruiqi, Pierre Benckendorff, y Deanne Gannaway. 2019. "Progress and new directions for teaching and learning in MOOCs." *Computers & Education* 129: 48-60. <https://doi.org/10.1016/j.compedu.2018.10.019>.

Dorça, Fabiano. 2015. "Implementation and use of simulated students for test and validation of new adaptive educational systems: A practical insight." *International Journal of Artificial Intelligence in Education* 25: 319-345. <https://doi.org/10.1007/s40593-015-0037-0>.

EDUCAUSE 2020. *2020 EDUCAUSE Horizon Report Teaching and Learning Edition*. <https://library.educause.edu/resources/2020/3/2020-educause-horizon-report-teaching-and-learning-edition>.

Fadel, Charles, y Jennifer S. Groff. 2019. "Four-dimensional education for sustainable societies." *Sustainability, Human Well-Being, and the Future of Education*, editado por Justin W Cook, 269-281. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-78580-6_8.

Flogie, Andreij, y Boris Aberšek. 2021. "Artificial intelligence in education." In *Active Learning-Theory and Practice*, editado por Olena Lutsenko, y Gregory Lutsenko. IntechOpen. <https://doi.org/10.5772/intechopen.96498>.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, y Raja Chatila. 2018. "AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds & Machines* 28: 689–707. <https://doi.org/10.1007/s11023-018-9482-5>.

Floridi, Luciano. 2014. *The 4th revolution: how the infosphere is reshaping human reality*. New York: Oxford University Press.

Fombona, Javier, Maria Angeles Pascual, y Maria Carmen Gonzalez-Videgaray. 2017. "M-learning y realidad aumentada: revision de literatura científica en el repositorio WoS." *Comunicar* 25 (52): 63-72. <https://doi.org/10.3916/C52-2017-06>.

García-Martínez, Inmaculada, José Maria Fernández-Batanero, José Fernández-Cerero, y Samule León. 2023. "Analysing the Impact of Artificial Intelligence and Computational Sciences on Student Performance: Systematic Review and Meta-analysis." *Journal of New Approaches in Educational Research* 12 (1): 171-197. <https://doi.org/10.7821/naer.2023.1.1240>.

Garcia-Serrano, Ana, y Antonio Menta-Garuz. 2022. "La inteligencia artificial en las Humanidades Digitales: dos experiencias con corpus digitales." *Revista de Humanidades Digitales* 7: 19-39. 10.5944/rhd.vol.7.2022.30928.

González-Calatayud, Victor, Paz Prendes-Espinosa, y Rosabel Roig-Vila. 2021. "Artificial Intelligence for student assessment: a systematic review." *Applied Sciences* 11 (6): 5467. 10.3390/app11125467.

Gunesequera, Asela Indunil, Yukun Bao, y Mboni Kibelloh. 2019. "The role of usability on e-learning user interactions and satisfaction: a literature review." *Journal of Systems and Information Technology* 21 (3): 368-394. <https://doi.org/10.1108/JSIT-02-2019-0024>.

- Imani, Maryam y Gholam Ali Montazer. 2019. "A survey of emotion recognition methods with emphasis on E-Learning environments." *Journal of Network and Computer Applications* 147: 102423. <https://doi.org/10.1016/j.jnca.2019.102423>.
- Jara, Ignacio, y Juan Manuel Ochoa. *Usos y efectos de la Inteligencia Artificial en Educación*. Banco Interamericano de Desarrollo: Sector Social. División de Educación, n. IDB-DP 00776. <http://dx.doi.org/10.18235/0002380>.
- Koola, Paul Mario, Satheesh Ramachandran, y Kalyan Vadakkevedu. 2016. "How do we train a stone to think? A review of machine intelligence and its implications." *Theoretical Issues in Ergonomics Science* 17(2): 211-238. <https://doi.org/10.1080/1463922X.2015.1111462>.
- Kozinets, Robert V. (2015). "Netnography." In *The International Encyclopedia of Digital Communication and Society*, editado por Peng Hwa Ang y Robin Mansell. Hoboken, NJ, USA: John Wiley & Sons. ISBN 978-111876777-1.
- Long, Duri, y Brian Magerko. 2020. "What is AI Literacy? Competencies and Design Considerations." In *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, editado por Association for Computing Machinery, 1-16. Honolulu (USA). <https://doi.org/10.1145/3313831.3376727>.
- Marciniak, Renata, y Joaquin Gairín Sallán. 2018. "Dimensiones de evaluación de calidad de educación virtual: revisión de modelos referentes." *RIED. Revista Iberoamericana de Educación a Distancia*, 21 (1): 217-238. <http://dx.doi.org/10.5944/ried.21.1.16182>.
- Mcgillivray, Barbara, Thierry Poibeau, y Pablo Ruiz. 2020. "Digital Humanities and Natural Language Processing: "Je t'aime... Moi non plus." *Digital Humanities Quarterly* 14 (2). Id ffhal-02970302. <http://www.digitalhumanities.org/dhq/vol/14/2/000454/000454.html>.
- Nathoo, Amaan, Girish Bekaroo, Tanveer Gangabissoon, y Aditya Santokhee. 2020. "Using tangible user interfaces for teaching concepts of internet of things." *Interactive Technology and Smart Education* 17 (2): 133-158. <https://doi.org/10.1108/ITSE-09-2019-0061>.
- Nye, Benjamin D. 2015. "Intelligent Tutoring Systems by and for the Developing World: A Review of Trends and Approaches for Educational Technology in a Global Context." *International Journal of Artificial Intelligence in Education* 25 (2): 177-203. <https://doi.org/10.1007/s40593-014-0028-6>.
- Page, Matthew J., Joanne E. McKenzie, Patrick, M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, y Sergio Alonso-Fernández. 2021. "Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas." *Revista Española de Cardiología* 74 (9):790-799. <https://www.revespcardiol.org/es-declaracion-prisma-2020-una-guia-articulo-S0300893221002748>.
- Restrepo-Echeverri, Daniel, Jovani Alberto Jiménez-Builes, y John William Branch-Bedoya. 2022. "Education 4.0: integration of educational robotics and smart mobile devices as a didactic strategy for the training of engineers in STEM." *Revista DYNA* 89 (222): 124-135. <https://doi.org/10.15446/dyna.v89n222.100232>.

- Rojas, Pablo. 2017. "Learning Analytics. Una revisión de la literatura." *Educación y Educadores* 20 (1): 106-128. <https://doi.org/10.5294/edu.2017.20.1.6>.
- Scoular, Claire, Esther Care, y Friedrich W. Hesse. 2017. "Designs for operationalizing collaborative problem solving for automated assessment." *Journal of Educational Measurement* 54 (1): 12-35. <https://doi.org/10.1111/jedm.12130>.
- Selwyn, Neil. 2019. ¿Debería los robots sustituir al profesorado? La IA y el futuro de la educación. Las Rozas (Madrid): Ediciones Morata. <https://bit.ly/3zxyPmO>.
- Troussas, Christos, Akrivi Krouska, y Cleo Sgouropoulou. 2020. "Collaboration and fuzzy-modeled personalization for mobile game-based learning in higher education." *Computers & Education* 144: 103698. <https://doi.org/10.1016/j.compedu.2019.103698>.
- Tuomi, Ilkka. 2018. "The Impact of Artificial Intelligence on Learning, Teaching, and Education." In *Science for Policy report by the Joint Research Centre (JRC)*, editado por Cabrera Giraldez, Marcelino; Riina Vuorikari, and Yves Punie. Luxemburg: Publications Office of the European Union. <https://publications.jrc.ec.europa.eu/repository/handle/JRC113226>.
- Vilkova, Kseniia, y Irina Shcheglova. 2021. "Deconstructing self-regulated learning in MOOCs: In search of help-seeking mechanisms." *Education and Information Technologies* 26: 17-33. <https://doi.org/10.1007/s10639-020-10244-x>.
- Villalobos Portalés, Jorge. 2022. La autoría de la Inteligencia Artificial en el derecho español. *Revista Justicia&Derecho* 5 (1): 1-19. <https://revistas.uautonoma.cl/index.php/rjyd/article/view/1840>.
- Vuorikari, Riina, Stefano Kluzer, Stefano, y Yves Punie. 2022. *DigComp 2.2: The Digital Competence Framework for Citizens – With new examples of knowledge, skills and attitudes*. Luxembourg: Publications Office of the European Union. <https://publications.jrc.ec.europa.eu/repository/handle/JRC128415>.
- Wei, Pengcheng, y Yu Zhao. 2019. "A novel speech emotion recognition algorithm based on wavelet kernel sparse classifier in stacked deep auto-encoder model." *Personal and Ubiquitous Computing* 23: 521-529. <https://doi.org/10.1007/s00779-019-01246-9>.
- Westera, Wim, Rui Prada, Samuel Mascarenhas, Pedro A Santos, João Dias, Manuel Guimarães, Konstantinos Georgiadis, Enkhbold Nyamsuren, Bahreini Kiavash, Zerrin Yumak, Chris Christyowidiasmoro, Mihai Dascalu, Gabriel Gutu-Robu, y Stefan Ruseti. 2020. "Artificial Intelligence Moving Serious Gaming: Presenting Reusable Game AI Components." *Education and Information Technologies* 25: 351-380. <https://doi.org/10.1007/s10639-019-09968-2>.
- Yadav, Aman, Jon Good, Joke Voogt, y Petra Fisser. 2016. "Computational Thinking as an Emerging Competence Domain." In *Competence-Based Vocational and Professional Education: Bridging the Worlds of Work and Education*, editado por Martin Mulder, 1051-1067. Springer International Publishing AG. https://doi.org/10.1007/978-3-319-41713-4_49.
- Xuesong, Zhai, Xiaoyan Chu, Ching Sing Chai, Morris Siu Yung Jong, Andreja Istenic, Michael Spector, Jia-Bao Liu, Jing Yuan, y Yan Li. 2021. "A Review of Artificial Intelligence (AI) in Education from 2010 to 2020." *Complexity*, 2021. <https://doi.org/10.1155/2021/8812542>.

Conceptual model for universal bibliographic control

Sandra Gisela Martin^(a)

a) National University of Cordoba, <https://orcid.org/0000-0003-4045-4915>

Contact: Sandra Gisela Martin, sandragmartin@gmail.com

Received: 25 July 2023; **Accepted:** 28 September 2023; **First Published:** 15 January 2024

ABSTRACT

A proposal is presented for a conceptual model of universal bibliographic control based on general systems theory (GST) and considered as a global system with a decentralized, distributed, collaborative, interconnected and linked model structure. The CBU is reflected on based on the four principles of relevant knowledge proposed by Morin (1999): context, global, multidimensional and complex; and nine dimensions of analysis are proposed: theoretical/conceptual framework, political, legislative, librarian, collaborative, normative, technological, relational, services.

KEYWORDS

Conceptual model; Universal bibliographic control; General systems theory.

Modelo conceptual para el control bibliográfico universal

RESUMEN

Se presenta una propuesta de modelo conceptual de control bibliográfico universal basado en la teoría general de sistemas (TGS) y considerado como un sistema global con una estructura de modelo de descentralizado, distribuido, colaborativo, interconectado y vinculado. Se reflexiona sobre el CBU a partir de los cuatro principios del conocimiento pertinente planteados por Morin (1999): el contexto, lo global, lo multidimensional y lo complejo; y se plantean nueve dimensiones de análisis: teórica/marco conceptual, política, legislativa, bibliotecológica, colaborativa, normativa, tecnológica, relacional, servicios.

PALABRAS CLAVE

Modelo conceptual; Control bibliográfico universal; Teoría general de sistemas.

Introducción

Se parte del concepto de modelo que según la RAE (2020b) se define como un “esquema teórico, generalmente en forma matemática, de un sistema o de una realidad compleja, como la evolución económica de un país, que se elabora para facilitar su comprensión y el estudio de su comportamiento.” En palabras de Mondino (2014) “los modelos son una simplificación de la realidad, la abstracción de complejas estructuras” (p.59).

En el mundo documental, los modelos conceptuales permiten describir y representar la realidad con un alto nivel de abstracción y constituyen un marco de referencia teórico para la elaboración de modelos y estructuras de datos. En otras palabras, se puede decir que los modelos conceptuales pueden implementarse, en muchos casos, mediante los modelos de datos que son prácticos y permiten describir los conjuntos de datos y las operaciones para manipularlos. Estas descripciones permiten establecer así, la estructura física de las bases de datos y las relaciones entre los datos.

Específicamente en el ámbito bibliotecológico, existen varios modelos conceptuales bibliográficos desarrollados por IFLA quien a través del Grupo de Revisión de Modelos Conceptuales Bibliográficos (*Bibliographic Conceptual Models Review Group* (BCM RG)) se encarga de brindar apoyo para el mantenimiento y desarrollo de los modelos y, a su vez, reporta al Comité de Normas de la IFLA.

Desde otra perspectiva, los modelos conceptuales bibliográficos, se implementan a partir de modelos de datos que tienen su desarrollo en los campos de la descripción bibliográfica. Los modelos de datos más conocidos son MARC 21 y los distintos esquemas de metadatos, como por ejemplo Dublin Core (DC), pero el más actual, y aún en etapa piloto, es BIBFRAME (del inglés *Bibliographical Framework*).

Los modelos de datos bibliográficos favorecen el CBU y en ese sentido Garduño Vera (1996) propone un modelo “como apoyo al control bibliográfico universal y al diseño de prototipos de bases de datos bibliográficas” (p. xvii) para lo cual comparó y estableció los campos de mayor uso en tareas de control bibliográfico y su aplicación en actividades de intercambio bibliográfico.

En base a lo anteriormente expuesto y siguiendo a Guerrini (2021a) el desafío del control bibliográfico es definir y compartir modelos conceptuales, para luego comenzar a trabajar colaborativamente en el diseño de protocolos entre las distintas comunidades y grupos de usuarios. La creación de modelos conceptuales conlleva tiempo, esfuerzo y fundamentalmente un debate interdisciplinario. Ningún modelo es acabado en sí mismo, y cada modelo puede complementarse con otros. En la revisión de la literatura se encontraron dos propuestas diferentes que constituyen antecedentes para el análisis de los modelos conceptuales para el CBU. El primer trabajo, de Akidi (2020) examina las prácticas del control bibliográfico en la era digital desde la perspectiva de un marco teórico y un modelo conceptual. Se basa por un lado, en las cinco leyes de Ranganathan como teoría importante para ilustrar cuestiones sobre la disponibilidad, la accesibilidad y la utilización de los recursos de información, ya que son los principios básicos de las prácticas de control bibliográfico; y por el otro, en el modelo Technology Acceptance model (TAM) desarrollado por Davis en 1989 que analiza la aceptación de las tecnologías tanto por parte de las agencias bibliográficas como por los usuarios que necesitan el acceso y uso de los recursos de información.

El segundo documento, fue presentado por Vivarelli (2021) en el VII Encuentro Arte do bibliografía y trata sobre los modelos y formas del pensamiento bibliográfico como un punto de

vista para interpretar la complejidad. El autor señala que las transformaciones actuales en los modelos de producción, organización y gestión del conocimiento registrado, el crecimiento vertiginoso de los datos, cada vez más numerosos, relacionados y conectados, hacen imprescindible una reflexión sobre los principios y modelos con los que se lleva a cabo la representación del conocimiento. Plantea interpretar el pensamiento bibliográfico a partir de tres niveles y a un espacio de integración:

- *nivel ontológico*, con respecto a las “cosas” del mundo real
- *nivel epistemológico*, que se refiere a las formas de conocer las “cosas”
- *nivel de registros* de representaciones de objetos, bibliográficos y no bibliográficos
- *nivel del Multiverso Bibliográfico*, espacio en donde la complejidad bibliográfica no puede ser considerada como un universo multidimensional, sino como un Multiverso Bibliográfico, multidimensional, inter y transdisciplinar, que tiende a integrar el desorden de los datos con los modelos normalizados de registro y ordenación de la información.

Se presenta a continuación, una propuesta de modelo conceptual de control bibliográfico universal basado en la teoría general de sistemas (TGS) que trata de analizar el todo en función de la interacción de sus partes y las partes en función del todo (Bertalanffy 2006). Al respecto, Morin (1999) en su libro *Los siete saberes necesarios para la educación del futuro* sostiene que:

la supremacía de un conocimiento fragmentado según las disciplinas impide, a menudo operar el vínculo entre las partes y las totalidades y, debe dar paso a un modo de conocimiento capaz de aprehender los objetos en sus contextos, sus complejidades y sus conjuntos. (p. 1)

La TGS formulada por Ludwig von Bertalanffy en la década del 30 se postula como una forma de pensar y una manera de mirar el mundo con la finalidad de lograr una unidad en la diversidad (Francia 1984). Justamente el CBU se presenta como un sistema global, una unidad integrada por múltiples componentes que interactúan y que se caracterizan por la diversidad de elementos ya sean documentos, actores, legislaciones, bibliotecas, agencias o procesos, entre otros.

Un sistema es concebido como un conjunto de entidades (elementos) interrelacionadas que interactúan entre sí con un objetivo en común. Los cuatro ejes – conjunto, elementos, interacción y objetivo – brindan a los sistemas, la peculiaridad de la globalidad. La mención de *conjunto* da la idea de una totalidad desde una dimensión holística donde el todo es más que la suma de las partes (concepto conocido como sinergia). Morin (1999) sostiene que “el todo tiene cualidades o propiedades que no se encontrarían en las partes si éstas se separaran las unas de las otras y ciertas cualidades o propiedades de las partes pueden ser inhibidas por las fuerzas que salen del todo” (p. 13). Es decir, que el todo no es la mera suma de las partes, dado que posee propiedades “emergentes” que no se encuentran en sus partes. Luego, la interacción de las partes en un sistema dinámico permite que el sistema logre su finalidad.

Los sistemas están inmersos dentro de una jerarquía. Todo sistema dispone de autonomía y se considera como un todo en sí mismo, pero a su vez, puede estar subordinado a un sistema superior y al mismo tiempo estar compuesto por varios subsistemas. Cuando un componente de un sistema es en sí mismo un sistema, se constituye en un subsistema. Es así que se plantean macrosistemas, sistemas y subsistemas (Ander-Egg 2002; Francia 1984; Johansen Bertoglio 2008; Levaggi 2000; Volpentesta 2004). En todos debería haber interrelación entre sus componentes

y transferencia de información entre ellos, para que cada uno pueda realizar sus tareas y lograr sus objetivos.

El CBU se manifiesta como un macrosistema, compuesto por diversos subsistemas. Mediante la TGS se intenta realizar un análisis sistémico que lleva a reflexionar sobre cómo pensar hoy el CBU de acuerdo a su impronta, aún vigente, de sistema mundial de control e intercambio de información bibliográfica planteada por UNESCO en la década del setenta (UNESCO, 1979).

Morin (1999) se pregunta “¿cómo lograr el acceso a la información sobre el mundo y cómo lograr la posibilidad de articularla y organizarla?” (p.12). El proceso del CBU tradicionalmente se ha caracterizado por poner énfasis en organizar y describir documentos con normativas bibliográficas internacionales, pero aún queda mucho por avanzar en el acceso, articulación, vinculación y conexión de datos y procesos a nivel internacional. Las posibilidades que brindan actualmente las tecnologías podrían cambiar la historia del CBU aunque sin lugar a dudas queda mucho trabajo por delante.

A continuación, se reflexiona sobre el CBU a partir de los cuatro principios del conocimiento pertinente planteados por Morin (1999): el contexto, lo global, lo multidimensional y lo complejo.

El contexto

Morin (1999) sostiene que “hay que ubicar las informaciones y los elementos en su contexto para que adquieran sentido. Para tener sentido la palabra necesita del texto que es su propio contexto y el texto necesita del contexto donde se enuncia” (p. 12). Según la norma ISO 9000:2015 el contexto de una organización es un proceso que determina los factores que influyen en el propósito, objetivos y sostenibilidad de la organización. El contexto interno considera factores tales como valores, cultura, conocimiento y desempeño de la organización, el contexto externo incluye factores como entornos legales, tecnológicos, de competitividad, de mercados, culturales, sociales y económicos (ISO 9000 2015).

El contexto actual del CBU está signado por un ecosistema bibliográfico caracterizado por la biodiversidad y dentro de un entorno mundial volátil, incierto, complejo y ambiguo. Las distintas tipologías de documentos y la variedad de formas de edición hacen que las fronteras de lo nacional y lo internacional sean muy difíciles de determinar, por lo que se puede afirmar que el mundo está en presencia de un contexto bibliográfico sin fronteras, en el cual más allá de las producciones nacionales, el control bibliográfico tiene por sobre todo un impacto a nivel mundial.

Vivarelli (2021) plantea un contexto caracterizado por un multiverso bibliográfico, multidimensional, inter y transdisciplinar. En ese mismo sentido Morin (1999) afirma que la educación del futuro tiene un gran desafío “porque hay una inadecuación cada vez más amplia, profunda y grave por un lado entre nuestros saberes desunidos, divididos, compartimentados y por el otro, realidades o problemas cada vez más poli disciplinarios, transversales, multidimensionales, transnacionales, globales, planetarios.” (p. 12). Pensar en un abordaje actual y en futuro del CBU, tiene mucho paralelismo con lo que está pasando en la educación.

Dentro del contexto el CBU se conforma de partes o componentes. Por un lado, entran en acción las *partes abstractas o conceptos*, tales como reglas, normas, leyes, procedimientos, redes de agencias, control bibliográfico nacional (CBN), bibliotecas nacionales, bibliografías nacionales o

identificadores. Por el otro, los *sujetos* (*personas o instituciones*) dado que para el logro del CBU es necesario que el modelo conecte los distintos actores (entidades) del ecosistema documental para favorecer la interacción entre:

1. Estado
2. Bibliotecas, Archivos, Museos y Galerías
3. Bibliotecas públicas
4. Bibliotecas nacionales
5. Agencias bibliográficas nacionales (no siempre están alojadas en las bibliotecas nacionales)
6. Agencias administradoras de identificadores bibliográficos
7. Autores, creadores, etc.
8. Editores/editoriales/productores
9. Asociaciones profesionales
10. Sociedades de escritores
11. Organismos responsables de la web: responsables de la web de datos
12. Universidades, etc.

Teniendo en cuenta lo analizado en los distintos capítulos de este trabajo, se puede visualizar un contexto del CBU conformado por dos realidades de acuerdo al origen del patrimonio cultural documental (Tabla 1).

Tabla 1: Contexto del CBU

CBU	
Patrimonio cultural documental registrado	Patrimonio cultural documental no registrado
Depósito legal	Sin depósito legal
Registro de identificadores: ISBN, ISSN, ISMN, ISRC, ISWC	Sin registro de identificadores
Sin curaduría de datos	Con curaduría de datos

Fuente: elaboración propia

Por un lado, el CBU del patrimonio cultural documental registrado que proviene del cumplimiento del depósito legal y/o de los registros de las bases de datos de los distintos sistemas de identificadores, si bien los datos suelen estar dispersos, el control es más factible de lograr dado que al estar registrado e identificado, sólo faltaría conectar y vincular los sistemas. En este caso las bibliotecas nacionales no deberían “identificar” el patrimonio, sino conectar el mismo y no deberían realizar procesos de curaduría, sino de mejoras de los metadatos.

Para tener una idea global de la magnitud de este tipo de patrimonio, se muestra en la Tabla 2 la información con las cantidades aproximadas de los registros que poseen los distintos sistemas de identificadores a nivel mundial.

Tabla 2: Cifras aproximadas de los identificadores

ID	Números/registros	Agencias registrantes	Países
<i>Identificadores bibliográficos</i>			
ISBN	150 millones	150	200
ISSN ¹	2.5 millones		93
ISRC	Sin datos		
ISMN	Sin datos		
Handle	Sin datos		
PII	Sin datos		
PURL	Sin datos		
DOI	275 millones	11	
ISAN ²	1.3 millones 14.000 proveedores registrantes	14	81
ISWC ³	52 millones	54	79
<i>Identificadores institucionales y de personas</i>			
ARK	Sin datos		
VIAF	87 millones de entidades 33 millones de registros clúster 172 millones de registros bibliográficos	56	30
ISIL	Sin datos		
ORCID	13 millones	1	51
ISNI	14,5 millones	30	
Ringgold	600.000 registros		
ROR	102.000 organizaciones		

Fuente: elaboración propia

Acaso, la conexión y vinculación de millones y millones de registros, provenientes de tantos países y recolectados por una gran cantidad de agencias administradoras ¿no sería una forma de concretar el CBU al menos de lo que está registrado en estos sistemas?

Por otro lado, el CBU del patrimonio cultural documental no registrado, que no es alcanzado, en la mayoría de los países, por el depósito legal ni por el registro de identificadores. En particular, los documentos que forman parte del patrimonio gráfico, filatélico, cartográfico o del *software*. Este contexto es el que mayor dificultad genera para lograr un CBU, dado que resulta más difícil y más laboriosa la tarea de buscar, identificar, recopilar y registrar el patrimonio que se encuentra disperso y del cual se requiere de un proceso de curación de contenidos y de un arduo trabajo de las bibliotecas nacionales o todas aquellas que estén involucradas en el CBU.

¹ ISSN <https://www.issn.org/es/comprender-el-issn/registro-internacional-del-issn/>

² ISAN https://www.isan.org/docs/ISAN_Overview.pdf

³ ISWC <https://www.iswc.org/es>

Si el objetivo es poder identificar y brindar acceso a todo lo producido en los distintos formatos y tipologías de documentos en el mundo entero, ambos contextos del CBU se deberían tener en cuenta para garantizar que se refleje la totalidad del patrimonio cultural documental a nivel mundial.

Lo global

En palabras de Morin (1999) “lo global es más que el contexto, es el conjunto que contiene partes diversas ligadas de manera inter-retroactiva u organizacional” (p. 13). Como señalan Dunsire y Willer (2014) todo se conecta con todo a nivel global, si no lo es ahora, lo será en el futuro. El CBU se presenta como un gran sistema global de redes interconectadas y relacionadas en donde se aprovechen las infraestructuras locales y los recursos de información globales.

Lo multidimensional y lo complejo

Un modelo conceptual para el control bibliográfico universal, dado el nivel de complejidad, requiere ser desarrollado en forma colaborativa, con múltiples miradas y con el aporte de distintas disciplinas, razón por la cual el presente modelo sólo constituye una propuesta inicial, un punto de partida, un modelo en evolución que deberá ser debatido y enriquecido con una comunidad internacional integrada por los distintos actores del ecosistema documental.

Empleando las palabras de Morin (1999):

complexus significa lo que está tejido junto; en efecto, hay complejidad cuando son inseparables los elementos diferentes que constituyen un todo (como el económico, el político, el sociológico, el psicológico, el afectivo, el mitológico) y que existe un tejido interdependiente, interactivo e inter-retroactivo entre el objeto de conocimiento y su contexto, las partes y el todo, el todo y las partes, las partes entre ellas. (p.13)

Como señala Mondino (2014) los “modelos teóricos (conceptuales) son muy útiles cuando se quieren estudiar fenómenos o sistemas complejos, puesto que representan la materia en estudio de modo más simple, centrándose en los aspectos más destacados del fenómeno, soslayando ciertos detalles” (p.60). Teniendo en cuenta la complejidad del control bibliográfico universal, se propone un modelo conceptual considerado como un sistema global y con una estructura de modelo de descentralizado, distribuido, colaborativo, interconectado y vinculado.

La multidimensionalidad es propuesta por Vivarelli (2021) para analizar, lo que el autor denomina, el multiverso bibliográfico. Desde la posición de Morin (1999) “las unidades complejas, como el ser humano o la sociedad, son multidimensionales; el ser humano es a la vez biológico, psíquico, social, afectivo, racional. La sociedad comporta dimensiones históricas, económicas, sociológicas, religiosas... El conocimiento pertinente debe reconocer esta multidimensionalidad” (p. 13). En ese mismo sentido, el modelo conceptual del CBU constituye una totalidad compleja que está basada en la interconexión de varias dimensiones (Tabla 3) con una mirada sistémica del proceso donde se visualizan las partes y los actores intervinientes.

Tabla 3: Dimensiones del modelo conceptual CBU

Dimensión	Alcance	
Teórica/marco conceptual	CBU	- <i>Debate internacional</i> Patrimonio cultural documental Bibliotecas nacionales: redefinición de visiones/misiones Depósito legal Bibliografía nacional - <i>Directrices/lineamientos internacionales</i> IFLA UNESCO
Política	CBN	Políticas nacionales de información Políticas de gestión del patrimonio cultural documental Políticas de depósito legal Política de depósito legal electrónico Política de elaboración de la bibliografía nacional Plan nacional de patrimonio cultural documental
Legislativa	CBN	Legislación sobre el patrimonio cultural documental Legislación sobre el depósito legal
Bibliotecológica	CBU	Formación de los profesionales de la información Investigación Bibliotecas
Colaborativa	CBN CBU	Agencias bibliográficas regionales Acuerdos con productores Convenios de interconexión con identificadores
Normativa	CBU	Normalización bibliográfica
Tecnológica	CBN CBU	Interconexión Interoperabilidad Datos abiertos enlazados Web semántica
Relacional	CBN CBU	Redes de relaciones dinámicas e interactivas
Servicios	CBN CBU	Buscadores Métricas, indicadores, reportes estadísticos

Fuente: elaboración propia

Dimensión teórica/marco conceptual: debates terminológicos

La presente dimensión tiene que ver, en primer lugar, con la necesidad de profundizar el debate internacional entre todas las comunidades vinculadas al mundo documental en pos de conceptualizar el patrimonio cultural documental. Como propone Vivarelli (2021) un abordaje inter y transdisciplinar para analizar el multiverso bibliográfico. El abordaje transdisciplinar aporta una mirada holística, vinculadora e integradora del CBU que trasciende las divisiones de cada disciplina. Unido a ello, se plantea la posibilidad de reflexionar sobre cuál es hoy el rol y las funciones de una biblioteca nacional a partir de una revisión y redefinición de las visiones y misiones. También, resulta indispensable, debatir sobre la importancia y el alcance del depósito legal y la bibliografía nacional en la identificación del patrimonio documental nacional. En el marco de esta investiga-

ción, se delinearón algunas posturas conceptuales, que ameritan un abordaje profundo, tendientes a cambiar algunas terminologías tales como:

- Agencia bibliográfica nacional por Agencia Documentográfica Nacional
- Bibliografía nacional por documentografía nacional
- Control Bibliográfico Universal (CBU) por Control Documentográfico Universal (CDU)

Un punto crucial de análisis debería ser el alcance tanto de las bibliografías nacionales como del CBU para determinar si se seguirá poniendo énfasis en los documentos publicados y provenientes del depósito legal o se abrirá su relevancia para contemplar otros tipos de documentos que también resultan de interés nacional y que representan su cultura y su identidad.

En segundo lugar, y como resultado de los debates anteriores, se visualiza la importancia de actualizar las directrices/lineamientos internacionales emitidas por las distintas organizaciones tales como IFLA o UNESCO. En general, las directrices sólo focalizan aspectos sobre preservación, protección y digitalización del patrimonio, existiendo una carencia en el abordaje de temas vinculados al acceso, la identificación y la difusión del patrimonio. En relación al depósito legal, las directrices presentan una falta de precisión en muchos puntos, especialmente en lo relativo al depósito legal digital. Una mayor descripción de los ítems a incluir en la legislación, ayudaría a los distintos países a elaborar leyes con mayor grado de profundidad y especificaciones que no den margen a vacíos legales o a malas interpretaciones.

Las directrices sobre bibliografías nacionales de IFLA son las más actualizadas, dado que su última versión corresponde al año 2021 y contienen una serie de indicaciones muy detalladas sobre las distintas prácticas y las diferentes alternativas que existen para la elaboración de las mismas. Finalmente, las directrices sobre CBU necesitan una urgente actualización para adecuarlas al contexto actual tal como fue descrito en el Capítulo 6.

Pero quizás, lo más importante es destacar que estos documentos que no sean elaborados y redactados en forma de silos, sino que sean concebidos como documentos que estén integrados y coordinados de manera que se complementen entre sí.

Dimensión política

El rol del Estado y la fijación de políticas a nivel nacional permiten dar un contexto a la gestión del patrimonio cultural documental. Para conocer la realidad de cada país resulta imprescindible proceder a la identificación y revisión de las políticas públicas y de las legislaciones existentes a nivel nacional, provincial, departamental y municipal en torno a la gestión y protección del patrimonio cultural documental. En base a ese panorama, se puede tener una visión de la situación actual y una ruta hacia dónde avanzar.

En muchos casos, la existencia de una política nacional de información ayuda a establecer planes, objetivos y responsabilidades, lo que permite tener en claro los procesos de gestión que involucran a las distintas bibliotecas ya sean nacionales y/o públicas, provinciales o municipales.

Dentro de este marco, la definición de una política nacional de gestión del patrimonio cultural documental, ya sea mediante una estructura centralizada o descentralizada en coordinación con las distintas organizaciones y niveles del gobierno nacional, provincial, regional o local, permite aunar esfuerzos para concretar las acciones necesarias que garanticen la correcta gestión del patrimonio documental.

En apoyo a la política nacional de gestión del patrimonio documental, una política de depósito legal que incluya además el depósito legal electrónico, aportaría los elementos necesarios para lograr la identificación y registro de los documentos producidos a lo largo y ancho de un país. Es aquí donde cada nación debería establecer su estructura de agencias bibliográficas que trabajen en red. Y en última instancia, también podría establecerse una política de elaboración de la bibliografía nacional en concordancia con las leyes de patrimonio y de depósito legal.

Estas políticas gubernamentales, permiten establecer una serie de acciones mediante las cuales el Estado interviene en estos temas que demanda la sociedad. Pero, aparte del establecimiento de políticas, los planes permiten llevar a cabo mediante acciones concretas las ideas plasmadas en las políticas. De allí la importancia de bajar las políticas a planes nacionales sobre el patrimonio cultural documental, que, por lo general, pueden ser coordinados por las bibliotecas nacionales y que contemplan todo el proceso de la gestión del patrimonio documental desde la identificación, selección, recopilación, organización (registro y descripción), control, custodia, conservación y preservación hasta la difusión.

Dimensión legislativa

Íntimamente vinculado a la dimensión política, se encuentra la dimensión legislativa. Las políticas, suelen ir acompañadas por legislaciones que brindan la obligatoriedad de ciertos procesos y dan las pautas de cómo concretarlos. Como ya se mencionó, la promulgación y/o actualización de las legislaciones sobre el patrimonio cultural documental y el depósito legal constituye un pilar fundamental para avanzar en el CBU.

Si bien muchos países poseen leyes de protección y gestión del patrimonio, la mención a los mecanismos en relación al patrimonio cultural documental no siempre es completa o precisa. Por tal razón, para fomentar el CBN resulta imprescindible contar con leyes nacionales, provinciales y/o municipales sobre el *patrimonio cultural documental*.

Y en la misma línea, una adecuada legislación sobre el depósito legal ayudaría a gestionar el patrimonio documental. Existe una necesidad de actualizar, en algunos casos, o de generar en otros, leyes sobre el depósito legal que contemplen todos los aspectos y que no dejen vacíos o lugar a dudas sobre su interpretación y alcance. Para ello sería conveniente que las leyes incluyan un marco conceptual claro, sobre lo que comprende el depósito legal, sobre el concepto de patrimonio cultural documental y sobre los tipos de patrimonio contemplados por la legislación. También sería aconsejable que queden explícitos los responsables, la obligatoriedad, las entidades encargadas del depósito legal, las instituciones depositarias, el tipo de contenido, las exclusiones, el alcance en relación a la territorialidad y la nacionalidad, la cantidad de ejemplares/copias y el plazo, las formas de difusión, los mecanismos de cumplimiento (beneficios o compensaciones, multas o sanciones), las copias para preservación, y la especificación sobre el depósito legal digital en cuanto a:

- Tipo de publicaciones electrónicas
- Tipo de contenido
- Forma de depósito
- Forma de acceso

- Formatos a depositar
- Preservación digital

Si bien, muchos de estos aspectos se mencionan y desarrollan en la bibliografía y en algunos lineamientos, no siempre se encuentran correctamente especificados en las legislaciones. De allí que las carencias y las confusiones conceptuales hacen que la aplicación de la ley no siempre es completa.

Dimensión bibliotecológica⁴

En esta dimensión se propone fortalecer la formación de los profesionales de la información, mediante el diseño de instancias de capacitación formal y no formal en todos los aspectos relativos a la gestión del patrimonio cultural documental, y sobre los temas relacionados tales como, bibliografías nacionales, depósito legal, CBU, identificadores unívocos, datos abiertos y desarrollos tecnológicos.

La adecuada formación y el compromiso de la comunidad bibliotecaria favorecerá que los profesionales puedan participar de las dimensiones políticas y legislativas, tanto en el diseño y elaboración de políticas públicas como en las propuestas de leyes.

Por otro lado, también se debería incentivar la investigación en la temática especialmente desde las instituciones que brindan carreras de grado y posgrado en ciencias de la información. El conocimiento en cada país sobre el estado del arte del control bibliográfico, como así también el estudio comparado de la situación latinoamericana y mundial permitiría tener un diagnóstico de lo logrado y una perspectiva sobre dónde se puede avanzar.

En tercer lugar, resulta imprescindible conocer las actividades y responsabilidades que tienen las bibliotecas para lograr el control bibliográfico nacional mediante la distribución y coordinación de tareas y servicios.

Dimensión colaborativa

El CBU requiere de un trabajo inter y transdisciplinario, con una responsabilidad distribuida y compartida entre los distintos actores del universo documental. Ya no es suficiente que cada país disponga de una única agencia bibliográfica nacional que asuma la totalidad del gigantesco trabajo que genera el control bibliográfico, sino que se necesita de un trabajo mancomunado de una red de agencias bibliográficas nacionales con la figura de la biblioteca nacional u otra institución como la coordinadora de la amplia red de agencias. Una opción podría ser que las bibliotecas nacionales generen acuerdos colaborativos con las agencias administradoras de identificadores, especialmente con aquellas que están alojadas en otras instituciones por fuera de la biblioteca nacional. Por lo general, los países disponen de agencias nacionales de registro del ISBN, ISSN, ISMN, ISRC e

⁴ De acuerdo al Diccionario de la Real Academia Española (2020) la bibliotecología es una ciencia que estudia las bibliotecas en todos sus aspectos, mientras que la biblioteconomía, es una disciplina encargada de la conservación, organización y administración de las bibliotecas. La bibliotecología proviene de los vocablos griegos: *biblion*, libro; *theke*, caja y *logos*, ciencia, tratado y constituye un léxico de mucho uso en el ámbito hispanoamericano que incluye, según algunos autores, otras disciplinas tales como la bibliografía, la bibliología y también la biblioteconomía, entre otras. Mientras que la biblioteconomía proviene de los vocablos griegos: *biblion*, libro; *theke*, caja y *nomos*, regla, ley y su uso es más generalizado en España.

ISWC que funcionan de manera aislada, independiente y sin conexión entre ellas. La posibilidad de lograr alianzas para compartir y conectar las bases de datos de esos identificadores, de mejorar los metadatos y de desarrollar servicios, permitiría avanzar hacia un verdadero control bibliográfico nacional.

De acuerdo a las características propias de cada país, se podría establecer una estructura descentralizada coordinada por las bibliotecas nacionales y que integren una red de bibliotecas públicas, provinciales junto con una red de agencias nacionales, donde pueda existir una agencia por cada uno de los identificadores. El trabajo en equipo podría aportar un mayor control bibliográfico, incluso con bibliotecas o agencias que se dediquen exclusivamente a gestionar el control bibliográfico de un determinado tipo de patrimonio, que luego se aportará a la correspondiente bibliografía nacional. Como lo manifiesta Parent (2007) al destacar la necesidad de compartir responsabilidades en la elaboración de las bibliografías nacionales para lograr un contexto de equilibrio entre la disminución de los recursos y la creciente necesidad de proporcionar acceso a la información de los fondos, las agencias bibliográficas nacionales deben recurrir a sus redes internas de bibliotecas para que les ayuden a mantener una bibliografía nacional actualizada y oportuna.

Para Parent (2007) el segundo enfoque para abordar el reto de describir enormes cantidades de material digital es trabajar con otros, con socios y da como ejemplo, los acuerdos que se podrían establecer con otros repositorios de material, como bibliotecas, archivos y museos, para que asuman la responsabilidad de proporcionar información descriptiva de partes de la colección nacional, quizás para un formato concreto de material, como el audiovisual, o en una región geográfica específica. Destaca que este tipo de acuerdos ya existen en varios países y que los metadatos también podrían ser suministrados por proveedores de información, como los editores.

Dimensión normativa

La comunidad bibliotecaria tiene una larga trayectoria en el desarrollo y aplicación de normativas para la descripción bibliográfica. Quizás esta sea la dimensión más consolidada dentro de todo el proceso que conlleva el CBU. Bide (2021) justifica la importancia del desarrollo y aplicación de estándares de identificación y metadatos porque satisfacen la necesidad de que la información cruce las fronteras entre diferentes sistemas informáticos conservando la mayor integridad semántica posible. La normalización en los sistemas de identificación es esencial para reducir errores, ambigüedades y para conectar sistemas. Para Bide (2021) con una norma (o conjunto de normas) bien especificada, la información pasa entre sistemas sin necesidad de intervención humana para resolver cuestiones de significado ambiguo. El intercambio de información y datos requiere normas, tanto a nivel nacional como internacional, para la descripción, la identificación y el formato de los datos (Bianchini et al., 2021)

Entre las ventajas que proporciona la aplicación de normativas se puede enumerar la calidad de los datos, la reducción de la complejidad, la transferencia e intercambio de datos, la vinculación de la información o la optimización de los procesos. Si bien se destaca la importancia de la normalización, también es necesario resaltar la necesidad de una mayor simplificación en los procesos. La normalización no necesariamente implica la uniformidad, sino una manera de trabajar para poder interactuar utilizando ciertos criterios comunes pero respetando la diversidad. La normalización

constituye un proceso dinámico y continuo dado que los contextos van cambiando y las normas se deben ir adaptando.

Dimensión tecnológica

La infraestructura tecnológica sostiene todo el proceso y posibilita el desarrollo de los servicios que se pueden brindar a partir de dos premisas: la interoperabilidad y los datos abiertos.

La norma ISO 21127:2014 define interoperabilidad como la capacidad que tienen sistemas de información diferentes para comunicar algunos de sus contenidos. La interoperabilidad puede implicar que – a) dos sistemas pueden intercambiar información, y/o – b) se puede acceder a varios sistemas con un único método. La norma diferencia la *interoperabilidad sintáctica*, cuando la codificación de la información y los protocolos de acceso de los sistemas correspondientes son compatibles, de modo que la información puede procesarse sin errores, de la *interoperabilidad semántica*, dado que muchas veces la interoperabilidad sintáctica no garantiza que cada sistema procese los datos de forma coherente con el significado previsto. A veces se requiere que los datos dispongan de tablas con una equivalencia semántica.

A través de la interoperabilidad se logra la vinculación, del uso compartido y de la reutilización de los datos. Tal como se expresa en el sitio web del BCM RG (2019) al igual que otras instituciones de la memoria, las bibliotecas tendrán que desempeñar un papel importante en la web semántica. En ese contexto, las ontologías y los modelos conceptuales en el campo de la información del patrimonio cultural son trascendentales, y la interoperabilidad entre estas ontologías y modelos quizás sea aún más crucial.

La publicación de los metadatos de los registros bibliográficos y de los identificadores como datos abiertos favorece la interconexión del universo documental en la web semántica, lo que redundaría en un control bibliográfico universal. La tecnología de hoy permite así concebir al CBU como la interconexión de múltiples bases de datos distribuidas, abiertas y enlazables en la red.

Actualmente, las universidades a través de los repositorios institucionales (RI) y los sistemas de información para la investigación, los denominados sistemas CRIS (del inglés, *Current Research Information Systems*) constituyen un verdadero ejemplo de lo que se puede realizar para lograr un CBU, en este caso, de la producción científica y académica a nivel nacional e internacional. Los repositorios institucionales permiten recopilar, organizar, almacenar, preservar y difundir de manera abierta la producción resultante de la actividad académica e investigadora una institución. A su vez, los repositorios institucionales que tienen un alcance local, aportan sus registros a los sistemas nacionales de repositorios institucionales (por ejemplo, en Argentina, al Sistema Nacional de Repositorios Institucionales⁵ (SNRD)).

Por su lado, los CRIS conectan obras (artículos, libros, ponencias, proyectos de investigación), autores (científicos), entidades gubernamentales, organizaciones (universidades, centros de investigación), agencias de financiación y editores, entre otros actores. El eje de la conexión y vinculación de las distintas entidades del sistema CRIS lo constituyen los identificadores de autores, instituciones y obras. Así cada universidad aporta su CRIS local al sistema CRIS nacional.

⁵ Argentina. Sistema Nacional de Repositorios Institucionales <https://repositoriosdigitales.mincyt.gob.ar/>

El patrimonio documental universitario que se encuentra tanto en los RI como en los CRIS no está alcanzado por el depósito legal, y por lo general queda invisibilizado dentro del CBU. Sin embargo, es uno de los patrimonios que actualmente tiene mayor registro, acceso, visibilidad y conexión.

Otro ejemplo, comparable con el CBU es el sistema Redalyc que conecta, a partir de una infraestructura de acceso abierto diamante, a revistas científicas, instituciones, autores, países y artículos. Constituye un verdadero control bibliográfico regional de la producción de artículos en revistas científicas que ofrece una serie de servicios tales como la disponibilidad de datos abiertos, indicadores o conexión con ORCID.

Dimensión relacional

En su concepción de multiverso bibliográfico, Vivarelli (2021) amplía sus fronteras, no sólo utiliza los registros de la historia documental dentro del universo bibliográfico clásico, sino también todas las huellas de las diferencias de la realidad, a partir de las cuales se produce el ciclo del conocimiento futuro, representando internamente tipos de registros diferentes y heterogéneos. El universo bibliográfico, en este sentido, aspira a superar la distinción dualista entre objeto y sujeto, integrándolos en redes de relaciones dinámicas e interactivas. Por eso, la cultura bibliográfica presente y futura puede ser calificada como ciencia de las relaciones (Vivarelli, 2021).

Las relaciones entre los multiversos, permitiría lograr un sistema global interconectado en donde se aprovechen las infraestructuras locales y los recursos de información globales.

Dimensión de servicios

La dimensión de servicios asociados a todo el proceso del CBU, por lo general es la menos desarrollada. No sólo se trata de lograr conectar sistemas, sino determinar los servicios que se ofrecerán. Los buscadores, cobran aquí importancia fundamental para poder buscar, encontrar e identificar los datos necesarios. La capacidad de recuperación de la información, las posibilidades de búsquedas simples y avanzadas, la posibilidad de aplicar filtros, de poder exportar registros, de disponer de datos abiertos, constituyen sólo algunos de los servicios que se esperan como resultados, salidas del sistema del CBU.

También, resulta adecuado pensar en las métricas, los indicadores y los reportes estadísticos que permitan obtener datos sobre tendencias o patrones en el tiempo. Estos servicios resultan útiles para la toma de decisiones o para la investigación tanto para editoriales, investigadores, bibliotecas, proveedores, distribuidores o autores, como para usuarios en general.

La disponibilidad de Informes o reportes estadísticos con datos de documentos por temáticas, por editoriales, por tipología documental, por identificadores, etc. permitiría poder conocer la magnitud del CBU y lograr una retroalimentación del sistema.

Por lo general, las agencias de ISBN son las que suelen tener mayor cantidad de estadísticas a nivel nacional, y en América Latina, a partir de la labor de CERLAC también se disponen de datos regionales que ofrecen un panorama de la edición de libros mediante la consolidación y análisis de los datos provenientes del registro ISBN. Lamentablemente, no sucede lo mismo con el resto de los registros de identificadores, cuyos datos no se conocen.

Por último, el servicio de datos abiertos ya sea para poder ser conectados a través de APIs o para bajar el conjunto de datos con el objetivo de poder reutilizarlos o distribuirlos.

En base a las dimensiones desarrolladas, el control bibliográfico como programa a largo plazo establecido en la década del 70 por la UNESCO e IFLA con el objetivo de crear un sistema mundial de control e intercambio de información bibliográfica, de una forma internacionalmente aceptable, sobre las obras publicadas en todos los países, aún hoy está vigente y se sostiene en base a tres premisas:

- a) Control bibliográfico nacional (CBN)
- b) Control bibliográfico universal (CBU)
- c) Disponibilidad universal de documentos (DUP)

Las acciones y las relaciones entre los procesos del CBN, del CBU y de la DUP permiten afirmar que los límites nacionales trascienden desde lo local a lo global y que para promover el acceso y la disponibilidad de los documentos producidos en todas las tipologías, en todos los formatos y en todos los países, se podría comenzar a fomentar la idea de una Documentografía enlazada universal (DEU) como complemento o reemplazo del conocido concepto de CBU.

Referencias bibliográficas

- Akidi, Juliana. 2020. «Bibliographic Control Practices in the Digital Age: Conceptual and Theoretical Framework Perspective», 16.
- Ander-Egg, Ezequiel. 2002. *Claves para introducirse en el estudio de la teoría general de sistemas*. Hvmantitas 2000. Fundamentos de la acción social. Buenos Aires: Grupo Editorial Lumen: Hvmantitas.
- Arispe, Verónica, Elsie Ramírez, y Erin Freas-Smith. 2021. «Bibframe 100 Pilot at the Library of Congress». En . Buenos Aires: Biblioteca Nacional Mariano Moreno. <https://www.youtube.com/watch?v=dkUNyK9agjI>.
- Bertalanffy, L. Von. 2006. *Teoría general de los sistemas: fundamentos, desarrollo, aplicaciones*. Fondo De Cultura Economica USA.
- Bianchini, Carlo, Stefano Bargioni, y Camillo Carlo Pellizzari di San Girolamo. 2021. «Beyond VIAF: Wikidata as a Complementary Tool for Authority Control in Libraries». *Information Technology and Libraries* 40 (2). <https://doi.org/10.6017/ital.v40i2.12959>.
- Bibliographic Conceptual Models Review Group (BCM RG). 2019. «Related Efforts». 2019. <https://www.ifla.org/node/947>.
- Bide, Mark. 2021. «Identifier and Metadata Standards in the Publishing Industry. A report on current status for IPA and IFRRO». IPA, IFRRO. <https://internationalpublishers.org/images/aa-content/our-industry/state-of-publishing-reports/metadata-standards-in-the-publish-industry.pdf>.
- Dunsire, Gordon, y Mirna Willer. 2014. «The local in the global: universal bibliographic control from the bottom up». En , 14. Lyon, France. <http://library.ifla.org/817/1/086-dunsire-en.pdf>.
- Francia, Alvaro. 1984. *Introducción a la teoría general de los sistemas en torno a una comprensión sistémica de la cultura*. Buenos Aires: Librería Agropecuaria.
- Garduño Vera, Roberto. 1996. *Modelo bibliográfico basado en formatos de intercambio y en normas internacionales orientado al control bibliográfico universal*. Monografías 19. México: UNAM.
- Guerrini, Mauro. 2021. «New perspectives of the Universal Bibliographic Control in the digital era». En *bibcongress*. <https://www.bibcongress.ru/en/Program/Report?id=121&event=18>.
- ISO 9000. 2015. «Sistemas de Gestión de La Calidad. Fundamentos y Vocabularios».
- ISO 21127. 2014. «Information and Documentation — A Reference Ontology for the Interchange of Cultural Heritage Information». <https://www.iso.org/obp/ui/#iso:std:iso:21127:ed-2:v1:en>.
- Johansen Bertoglio, Oscar. 2008. *Introducción a la teoría general de sistemas / Oscar Johansen Bertoglio*. México, D.F.: Limusa : Grupo Noriega.
- Levaggi, Gero. 2000. *Teoría general de los sistemas. Aplicación a la administración de los negocios*. Colección Estudio. Buenos Aires: Ugerman Editor.
- Mondino, M. 2014. «Modelos conceptuales y mentales: Elementos para repensar la enseñanza y el aprendizaje», *Enfoques*, 26 (1): 57-78.

- Morin, Edgar. 1999. *Los siete saberes necesarios para la educación del futuro*. UNESCO. <https://edgarmorinmultiversidad.org/index.php/libros-sin-costo/94-los-7-saberes-necesarios-para-la-educacion-del-futuro-de-edgar-morin.html>.
- Parent, Ingrid. 2007. «The Importance of National Bibliographies in the Digital Age». En, 10. Durban, South Africa. <http://archive.ifla.org/IV/ifla73/papers/089-Parent-en.pdf>.
- Real Academia Española. (2020). Bibliotecología. En *Diccionario de la lengua española. Edición del Tricentenario* (24. ed.). RAE. <https://dle.rae.es/bibliotecología>.
- Real Academia Española. (2020). Biblioteconomía. En *Diccionario de la lengua española. Edición del Tricentenario* (24. ed.). RAE. <https://dle.rae.es/biblioteconomía>.
- Real Academia Española. 2020. Modelo. En *Diccionario de la lengua española. Edición del Tricentenario* (24. ed.). RAE. <https://dle.rae.es/modelo>.
- Riva, Pat, y Maja Žumer. 2018. «FRBRoo, the IFLA Library Reference Model, and Now LRMoo: A Circle of Development». En , 7. Kuala Lumpur, Malaysia.
- Tillett, Barbara. 2004. *¿Qué es FRBR? Un modelo conceptual del universo bibliográfico*. Washington, DC: Library of Congress. <http://www.loc.gov/catdir/cpsd/Que-es-FRBR.pdf>.
- UNESCO. 1979. «Guidelines for the National Bibliographic Agency and the National Bibliography». UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000048658>.
- Vivarelli, Maurizio. 2021. «Modelli e forme del pensiero bibliografico. In cerca di un punto di vista per interpretare la complessità = Modelos e formas do pensamento bibliográfico. Em busca de um ponto de vista para interpretar a complexidade». En . Universidade de Bolonha (Unibo) - Fórum Internacional A Arte da Bibliografia. <https://www.youtube.com/watch?v=L7nU7jOrhjs>.
- Volpentesta, Jorge Roberto. 2004. *Sistemas administrativos y sistemas de información*. 1a. ed. Buenos Aires: Osmar D. Buyatti.

The reconciliation of SBN authority records with Wikidata. Progresses and perspectives after a decade of work (2013-2023)

Camillo Carlo Pellizzari di San Girolamo^(a)

a) Scuola Normale Superiore, <https://orcid.org/0000-0003-2699-1693>

Contact: Camillo Carlo Pellizzari di San Girolamo, camillo.pellizzaridisangirolamo@sns.it
Received: 30 June 2023; **Accepted:** 23 September 2023; **First Published:** 15 January 2024

ABSTRACT

The article retraces diachronically the history of the reconciliation of Opac SBN's authority records with Wikidata, from the perspective of the editing activity on Wikidata; the article deals with the creation of the property used in Wikidata to connect to Opac SBN's name authority file (P396), the progressive growth of the number of values of P396 in Wikidata (compared with the identifiers of other European authority files), the reconciliation strategies that have been adopted and the main difficulties which have slowed down this process, and finally the relevant novelties brought firstly by the adoption of the new Opac SBN and secondly by the recent decisions taken by ICCU in order to promote a greater data openness.

KEYWORDS

Authority control; Entity management; Data reconciliation; Wikidata; Opac SBN.

Riconciliare le voci di autorità in SBN con Wikidata. Progressi e prospettive dopo un decennio di lavoro (2013-2023)

ABSTRACT

L'articolo intende ripercorrere in diacronia la storia della riconciliazione delle voci di autorità dell'Opac SBN con Wikidata, dalla prospettiva del lavoro svolto su Wikidata; vengono quindi trattate la nascita della proprietà usata da Wikidata per connettersi con l'authority file dei nomi dell'Opac SBN (P396), la crescita progressiva del numero di valori di P396 in Wikidata (comparata con quella degli identificativi di altri authority file europei), le strategie di riconciliazione adottate e le principali difficoltà che hanno rallentato questo processo, infine le importanti novità arretrate prima dall'adozione del nuovo Opac SBN e poi dalle recenti decisioni dell'ICCU volte a favorire una maggiore apertura dei dati di autorità.

PAROLE CHIAVE

Controllo di autorità; Entity management; Riconciliazione dei dati; Wikidata; Opac SBN.

1. Introduzione

Questo articolo intende delineare, attraverso i principali snodi, il rapporto che ha legato gli elementi di Wikidata¹ alle voci di autorità dell’Opac SBN² nel corso degli ultimi dieci anni (2013-2023). Tutti i link sono stati consultati per l’ultima volta il 30 giugno 2023 (nota: gli URL di formattazione, cioè i link contenenti il segnaposto \$1, non funzionano in sé, ma solo sostituendo al segnaposto \$1 un identificativo).

2. Storia della riconciliazione con Wikidata

Wikidata esiste dal 29 ottobre 2012 (Vrandečić, Pintscher, e Krötzsch 2023). Il primo elemento (Q1)³ è stato creato il medesimo giorno,⁴ mentre la prima proprietà (P6) è stata creata il 4 febbraio 2013; nei mesi compresi tra queste due date non era quindi possibile aggiungere dichiarazioni agli elementi. Le prime proprietà potevano avere come tipo di dato soltanto “elemento” (la prima fu la succitata P6) o “file multimediale su Commons” (la prima fu P10); la prima proprietà con tipo di dato “stringa” (P212), successivamente convertito in “identificativo esterno” laddove necessario, è stata creata il 6 marzo 2013; lo stesso giorno sono state create altre proprietà fondamentali quali “identificativo ISNI” (P213) e “identificativo VIAF” (P214).

2.1 Nascita e stabilizzazione della P396 (2013-2014)

La proprietà per connettere gli elementi di Wikidata alle voci di autorità di persone ed enti presenti nell’Opac SBN è stata proposta il 30 marzo 2013⁵ e sulla base di tale proposta è stata creata come P396 il 6 aprile 2013. La proprietà conteneva gli identificativi delle voci di autorità secondo il formato “lungo” esemplificabile con “IT\ICCU\CFIV\017574” e formalizzabile in espressione regolare come “IT\ICCU\\D{2}[A-Z0-3]V\\d{6}”; il formato “breve”, esemplificabile con “CFIV017574” e formalizzabile in espressione regolare come “\D{2}[A-Z0-3]V\d{6}”, è menzionato due volte nella pagina di discussione della proprietà prima del 2022,⁶ anno in cui fu poi adottato (cfr. infra 2.3). Si può notare comunque che il VIAF,⁷ al quale ICCU invia una parte delle proprie voci di autorità di persone ed enti dal 2009 (Caffo 2013, 10), ha sempre utilizzato negli URL delle sue schede *processed* il formato “breve”, ponendo l’ID nel formato “breve” nel campo 001 e l’ID nel formato “lungo” nel campo 035.⁸

¹ <https://www.wikidata.org/>.

² <https://opac.sbn.it/>.

³ <https://www.wikidata.org/entity/Q1>. Gli elementi e le proprietà di Wikidata hanno un URI costituito dal loro ID (es. Q1, P6) preceduto da <https://www.wikidata.org/entity/>.

⁴ Le informazioni sulla data di creazione delle pagine di Wikidata sono ricavabili applicando <https://www.wikidata.org/wiki/Special:PageInfo> alla pagina desiderata.

⁵ https://www.wikidata.org/wiki/Wikidata:Property_proposal/Archive/4#P396.

⁶ https://www.wikidata.org/wiki/Property_talk:P396/Archive#4_formats (25 aprile 2013); https://www.wikidata.org/wiki/Property_talk:P396/Archive#Format_question_on_frwiki (2 novembre 2018).

⁷ <https://viaf.org/>.

⁸ Es. <https://viaf.org/processed/ICCU%7CCFIV017574>.

Fino al 6 dicembre 2014⁹ la P396, avente come tipo di dato “stringa”, si presentava come una raccolta di etichette, descrizioni e alias in varie lingue, senza alcuna dichiarazione. Comunque, tale proprietà poteva già essere inserita negli elementi di Wikidata,¹⁰ al fine di indicare una corrispondenza tra un dato elemento e una data voce di autorità. Nel corso del 2013 il numero di valori di P396 presenti in Wikidata supera il migliaio e nel corso del 2014 supera le due migliaia.¹¹

Nelle prime settimane, i valori della P396 aggiunti agli elementi si presentavano come una semplice stringa, senza alcun link alla voce di autorità; il problema di quale link usare è stato posto già il 25 aprile 2013 nella pagina di discussione della proprietà,¹² dove vengono dibattute più ipotesi; il 21 maggio 2013 si decide di associare agli identificativi inseriti negli elementi tramite P396 un link che, tramite un complesso URL di formattazione,¹³ puntava direttamente all’Opac SBN tramite il gadget AuthorityControl.js;¹⁴ tale link è stato poi semplificato con una modifica del 7 giugno 2013 adottando un nuovo URL di formattazione più breve.¹⁵

Parallelamente, il 31 maggio 2013 la P396 viene aggiunta su Wikipedia in italiano al template Controllo di autorità,¹⁶ in modo da far comparire il link alle voci di autorità dell’Opac SBN nelle voci dell’enciclopedia;¹⁷ l’aggiunta viene discussa nei giorni successivi¹⁸ e definitivamente approvata il 13 giugno 2013.¹⁹ Anche altre Wikipedie hanno aggiunto, in momenti diversi, la P396 ai rispettivi template di controllo di autorità.²⁰

Una guida su come aggiungere la P396 a Wikidata²¹ viene scritta su Wikipedia in italiano negli ultimi mesi del 2013²² e poi ampliata a inizio 2015.²³ Sempre a fine 2013 viene creata su Wikipedia in italiano anche una pagina²⁴ volta a raccogliere i casi in cui autori presenti in Wikipedia in italia-

⁹ <https://www.wikidata.org/w/index.php?oldid=162895731> (10 ottobre 2014) è la versione più recente di P396 prima del 6 dicembre 2014.

¹⁰ Es. <https://www.wikidata.org/w/index.php?diff=prev&oldid=88111699>.

¹¹ 1300 nel luglio 2013 (Guerrini 2013, 189); oltre 1600 il 21 novembre 2013 (Martinelli 2014; Vullo 2014, 115); 2530 entro il 15 luglio 2014 (Lucarelli 2014, 246).

¹² https://www.wikidata.org/wiki/Property_talk:P396/Archive#Link_somewhere?

¹³ [http://opac.sbn.it/opacsbn/opaclib?db=solr_auth&resultForward=opac/iccu/full_auth.jsp&do_cmd=search_show_cmd&fname=none&item:1032.Codice_Auth=\\$1](http://opac.sbn.it/opacsbn/opaclib?db=solr_auth&resultForward=opac/iccu/full_auth.jsp&do_cmd=search_show_cmd&fname=none&item:1032.Codice_Auth=$1). Nota: negli URL di formattazione usati in Wikidata, \$1 è il segnaposto che indica l’identificativo.

¹⁴ <https://www.wikidata.org/wiki/MediaWiki:Gadget-AuthorityControl.js> (la modifica è <https://www.wikidata.org/w/index.php?diff=prev&oldid=44787646>).

¹⁵ [http://id.sbn.it/af/\\$1](http://id.sbn.it/af/$1) (la modifica è <https://www.wikidata.org/w/index.php?diff=prev&oldid=49767070>). Questo medesimo URL di formattazione è tuttora usato nelle schede *processed* del VIAF (es. <https://viaf.org/processed/ICCU%7CCFIV017574> > <http://id.sbn.it/af/IT/ICCU/CFIV/017574>). L’URL di formattazione [http://id.sbn.it/af/\\$1](http://id.sbn.it/af/$1) è stato poi esportato all’interno di P396, a seguito della creazione dell’apposita proprietà “URL di formattazione” (P1630), con una modifica del 6 dicembre 2014 (<https://www.wikidata.org/w/index.php?diff=prev&oldid=179251416>).

¹⁶ [https://it.wikipedia.org/wiki/Template:Controllo di autorità](https://it.wikipedia.org/wiki/Template:Controllo_di_autorit%C3%A0).

¹⁷ <https://it.wikipedia.org/w/index.php?diff=prev&oldid=59142049>.

¹⁸ [https://it.wikipedia.org/wiki/Discussioni_template:Controllo di autorità/Archivio/1#ICCU](https://it.wikipedia.org/wiki/Discussioni_template:Controllo_di_autorit%C3%A0/Archivio/1#ICCU).

¹⁹ <https://it.wikipedia.org/w/index.php?diff=prev&oldid=59415709>.

²⁰ Cfr. lista (incompleta) in <https://www.wikidata.org/wiki/Q18174792>.

²¹ <https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteche/VIAF/SBN>.

²² Cfr. versione del 1 dicembre 2013 (<https://it.wikipedia.org/w/index.php?oldid=62776507>).

²³ Cfr. versione del 1 aprile 2015 (<https://it.wikipedia.org/w/index.php?oldid=71707971>).

²⁴ Dal 2023 intitolata [https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteche/VIAF/SBN/Elenco autori privi di identificativo SBN](https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteche/VIAF/SBN/Elenco_autori_privi_di_identificativo_SBN).

no non avevano una voce di autorità nell'Opac SBN linkabile (sul problema delle voci di autorità invisibili, cfr. infra 2.2).

2.2 Evoluzione della riconciliazione (2015-2021)

Nel 2015 è stata siglata la prima convenzione tra ICCU e Wikimedia Italia (ICCU e Wikimedia Italia 2015). Nello stesso periodo ICCU si è dotata di un wikipediano in residenza, Luca Martinelli, che è stato attivo in tale ruolo dal 7 aprile 2015 al 31 dicembre 2019, usando sui progetti Wikimedia l'utenza "Sannita (ICCU)";²⁵ nel corso del 2015, Martinelli si è dedicato sia a "operare direttamente sui progetti Wikimedia" sia a "formare il personale dell'Istituto all'uso dei progetti" e la sua attività nei primi mesi si è concentrata principalmente sull'Anagrafe delle Biblioteche e CulturaItalia, ma ha riguardato anche "il lavoro di sincronizzazione delle schede di autorità di SBN di livello AUF con il progetto Wikidata".²⁶

Al 16 novembre 2015 erano 13471 i valori di P396 presenti in Wikidata, a fronte di 64502 voci di livello 97 (AUF) rese disponibili fino a quel momento da ICCU (Aste, Mataloni, e Martinelli 2015, 72). Entro il 19 dicembre 2015 il numero di valori ha raggiunto quota 16266 (Martinelli 2016, 82). Nello stesso anno 2015, tuttavia, ICCU ha correttamente individuato due rilevanti problemi che ostacolavano la conversione dei dati delle sue banche dati in linked open data (LOD) e che quindi impedivano anche un efficace lavoro di riconciliazione delle sue voci di autorità con un qualunque database esterno (es. Wikidata).

Innanzitutto, gli URI resi disponibili dall'Opac SBN (sia per i record bibliografici sia per le voci di autorità), formati dal dominio *id.sbn.it* e dall'identificativo univoco del record bibliografico²⁷ o dal dominio *id.sbn.it/af* e dall'identificativo univoco della voce di autorità,²⁸ cessavano di funzionare nel caso in cui l'ID fosse "cancellato o fuso con un altro", e ciò comportava l'"impossibilità di riferirsi a identificativi persistenti (ad esempio, BID e VID) per la natura stessa della catalogazione partecipata che prevede fusioni e cancellazioni dei record da parte delle biblioteche"; ICCU sottolineava quindi la necessità di "prevedere un sistema di re-indirizzamento che permetta, nel caso di fusione fra due record, di essere automaticamente reindirizzati al nuovo URI, tramite il vecchio collegamento, mantenendo così il riferimento all'oggetto originario" (Aste, Mataloni, e Martinelli 2015, 69, 71, 72 nota 22).²⁹ Tale sistema è stato attuato nel giugno 2023.

In secondo luogo, "l'archivio d'autorità Autori presenta un numero limitato di schede di autorità rispetto al numero complessivo di autori presenti nell'Indice SBN. Basti pensare che, a fronte di circa 4 milioni di record autori in Indice sono presenti soltanto 204.000 schede di autorità controllate" (*ibid.* 70). Ciò significa che solo le voci di autorità di livello 95 o superiore, che ICCU dal

²⁵ [https://meta.wikimedia.org/wiki/User:Sannita_\(ICCU\)](https://meta.wikimedia.org/wiki/User:Sannita_(ICCU)).

²⁶ Riguardo ai livelli di autorità, cfr. https://norme.iccu.sbn.it/index.php?title=Guida_musica/Appendici/Appendice_XV da cui si può ricavare l'equivalenza tra i codici numerici e i codici di tre lettere: 05 = REC, 51 = MIN, 71 = MED, 90 = MAX, 95 = SUP, 96 = LAV, 97 = AUF.

²⁷ Es. <http://id.sbn.it/BVE0391786>.

²⁸ Es. <http://id.sbn.it/af/IT\ICCU\CFIV\002049>.

²⁹ La necessità di garantire "la permanenza degli identificativi, anche dopo la loro cancellazione e/o sostituzione con altri", per le voci di autorità è stata successivamente sottolineata nel 2019 dal Gruppo di lavoro sulle biblioteche digitali dell'AIB (AIB - Gruppo di lavoro sulle biblioteche digitali 2019, 5).

2014 invia anche al VIAF, erano visibili nell'Opac SBN e raggiungibili tramite l'URI costruito col dominio *id.sbn.it/af*;³⁰ circa il 95% delle voci di autorità di persone ed enti risultava quindi invisibile nell'Opac SBN e non raggiungibile tramite un URI e quindi, di fatto, non riconciliabile con un altro database (es. Wikidata). Anche questo problema è stato superato, nel luglio 2023, con l'apertura delle voci di autorità fino al livello 51 compreso.

Nel maggio 2015 la P396 è uno dei primi identificativi esterni a essere dotato di un catalogo³¹ su Mix'n'match,³² un *tool* creato due anni prima³³ al fine di facilitare la riconciliazione tra database di vario genere e Wikidata. Tale catalogo conteneva 82035 voci di autorità di persone, corrispondenti all'insieme delle voci aventi allora livello 97; è arrivato fino a circa 36mila abbinamenti con Wikidata e ha continuato ad essere utilizzato fino al 10 agosto 2021, quando è stato sostituito da un nuovo catalogo più ampio,³⁴ comprendente 184874 voci di autorità di persone (l'insieme delle voci aventi allora livello pari o superiore a 95), che ha raggiunto a sua volta oltre 45mila abbinamenti nel corso di pochi mesi, fino alla sua dismissione il 7 febbraio 2022 nell'ambito del cambio di formato dei valori di P396 (cfr. infra 2.3).³⁵

Dall'11 agosto 2016 esistono, grazie al template Property uses,³⁶ dati quasi giornalieri sull'uso delle proprietà di Wikidata (anche se tali dati mischiano l'uso delle proprietà come dichiarazioni, come qualificatori e come riferimenti): in quella data³⁷ risultano 22489 valori di P396, che si possono confrontare ad esempio con i valori di due proprietà quali P269 (IDREF, l'authority file del SUDOC, il catalogo delle biblioteche universitarie francesi),³⁸ a quota 324750, e P691 (NKC, l'authority file del catalogo della Biblioteca nazionale della Repubblica ceca),³⁹ a quota 35011. Lo stesso confronto, cinque anni dopo (l'11 agosto 2021),⁴⁰ vede P396 a quota 50711, P269 a quota 538706, P691 a quota 728459. Tuttavia, ai fini di valutare la riconciliazione, i dati più utili sono quelli che considerano soltanto l'uso delle proprietà come dichiarazioni, disponibili grazie al template

³⁰ L'invio al VIAF delle voci di autorità di livello 95, oltre a quelle di livello 97, è annunciato come imminente in (Caffo 2013, 10) ed è confermato dai dati della tabella 3 di (Bianchini, Bargioni, e Pellizzari di San Girolamo 2021) (consultabile in https://catalogo.pusc.it/beyond_viaf/#tb3), che mostra come nel 2014 ICCU abbia mandato al VIAF 108207 nuove voci di autorità.

³¹ <https://mix-n-match.toolforge.org/#/catalog/58>.

³² <https://mix-n-match.toolforge.org/>.

³³ <http://magnusmanske.de/wordpress/?p=114>.

³⁴ <https://mix-n-match.toolforge.org/#/catalog/4629>.

³⁵ Mix'n'match in quanto strumento di riconciliazione collaborativa applicato alle voci di autorità dell'Opac SBN è il modello del progetto Coming AUTH proposto in (Galeffi et al. 2016), dove sono forniti anche dati precisi sulla distribuzione per livello delle voci di autorità, all'epoca 4366920 di cui 61078 a livello 97 e 145998 a livello 95, per un totale di 207076 voci di autorità visibili (il 4,7% del totale).

³⁶ https://www.wikidata.org/wiki/Template:Property_uses.

³⁷ <https://www.wikidata.org/w/index.php?oldid=363250448>.

³⁸ Le voci di autorità di persone del SUDOC inviate al VIAF erano 3278189 nel settembre 2020 (cfr. https://catalogo.pusc.it/beyond_viaf/#tb2), quindi si può considerare il SUDOC come un authority file di dimensioni comparabili a quello di SBN.

³⁹ Le voci di autorità di persone del NKC inviate al VIAF erano 791187 nel settembre 2020 (cfr. https://catalogo.pusc.it/beyond_viaf/#tb2), quindi si può considerare il NKC come un authority file di dimensioni circa 5 volte inferiori a quello di SBN.

⁴⁰ <https://www.wikidata.org/w/index.php?oldid=1478274358>.

Number of main statements by property⁴¹ dal 27 maggio 2020:⁴² P396 conta 34258 usi, a fronte di 447059 usi per P269 e 376123 usi per P691; al 16 dicembre 2021,⁴³ P396 è salita a 46039 usi, a fronte di 527335 usi per P269 e 447377 usi per P691. La disparità è evidente: la riconciliazione delle voci di autorità di SBN, essendo necessariamente limitata a circa 200mila voci visibili, ha numeri decisamente inferiori rispetto ad authority file sia di dimensioni comparabili (es. SUDOC) sia di dimensioni nettamente inferiori (es. NKC).

2.3 Nuovo Opac SBN e nuovi progetti (2022-2023)

Dal dicembre 2021, con la pubblicazione del nuovo Opac SBN e del portale Alfabetica, si è assistito a una ripresa dell'interesse reciproco di ICCU e della comunità di Wikidata per la riconciliazione tra le voci di autorità dell'Opac SBN e Wikidata, che ha portato a significativi passi avanti nel corso del 2022 e della prima metà del 2023.

L'innovazione più importante è stata il nuovo Opac SBN,⁴⁴ che ha compiuto due rilevanti progressi, in un'ottica di apertura dei dati, rispetto al vecchio Opac SBN. Un primo progresso è stato l'apertura completa (cioè indipendente dai livelli di autorità) delle voci di autorità delle opere⁴⁵ e dei luoghi,⁴⁶ precedentemente invisibili (Mataloni e Incelli 2022, 60). Un secondo progresso è stato l'apertura delle voci di autorità di persone ed enti aventi livello 90,⁴⁷ oltre alle già note voci di livello 95 o superiore, che ha quasi raddoppiato il numero di voci di autorità visibili.⁴⁸

Entrambe queste novità sono state velocemente recepite in Wikidata. In corrispondenza con la definitiva sostituzione del vecchio Opac SBN col nuovo Opac SBN (28 febbraio 2022)⁴⁹ sono state create la P10396 ("identificativo SBN di un'opera") e la P10397 ("identificativo SBN di un luogo"), già proposte il 3 febbraio 2022.⁵⁰ Inoltre, poiché il nuovo Opac SBN non supportava alcun link alle voci di autorità che utilizzasse gli ID nel formato "lungo" (cfr. supra 2.1), a seguito di una discussione si è deciso di modificare tutti i valori di P396 dal formato "lungo" (utilizzato fin dal 2013) al formato "breve"; tale transizione si è compiuta nel corso del mese di febbraio 2022.⁵¹ Il

⁴¹ https://www.wikidata.org/wiki/Template:Number_of_main_statements_by_property.

⁴² <https://www.wikidata.org/w/index.php?oldid=1191554377>.

⁴³ <https://www.wikidata.org/w/index.php?oldid=1544843287>.

⁴⁴ <https://nuovo-opac.sbn.it/>.

⁴⁵ <https://opac.sbn.it/opere>.

⁴⁶ <https://opac.sbn.it/luoghi>.

⁴⁷ L'8 febbraio 2021 questa novità era stata preannunciata dalla direttrice dell'ICCU, Simonetta Buttò, durante la International Conference Bibliographic Control in the Digital Ecosystem. Cfr. (Buttò 2022, XXIV): "We are now planning to turn our policy, which was to date limited to the sharing of authority records of a very high level, into a new approach with the help of the Wiki community thanks to our longstanding collaboration with Wikidata, which dates back to an agreement with Wikimedia Italy signed in March 2015".

⁴⁸ Nel dicembre 2021 erano disponibili 185132 voci di autorità di SBN nel VIAF (livello 95 o superiore) e 326849 nel nuovo Opac SBN (livello 90 o superiore), stante (Bianchini, Bargioni, e Pellizzari di San Girolamo 2022, 255); per un'analisi della qualità complessiva di questi due insiemi di voci di autorità, cfr. (*ivi*, 259-272 e tabelle in 296-307).

⁴⁹ Cfr. annuncio ufficiale di ICCU in <https://web.archive.org/web/20230423124440/https://m.facebook.com/iccu2/posts/2295078783965795> e conferma in <https://www.wikidata.org/w/index.php?diff=prev&oldid=1584551713>.

⁵⁰ https://www.wikidata.org/wiki/Wikidata:Property_proposal/SBN_new_authority_IDs.

⁵¹ Cfr. https://www.wikidata.org/wiki/Property_talk:P396/Archive#Nuovo_OPAC_SBN_e_formato_degli_ID e https://www.wikidata.org/wiki/Wikidata:Requests_for_permissions/Bot/Dexbot_15.

28 febbraio 2022 è stato adottato un nuovo URL di formattazione⁵² e il 6 marzo è stato caricato un nuovo catalogo di Mix'n'match⁵³ comprendente 328460 voci di autorità di persone (l'insieme delle voci aventi allora livello pari o superiore a 90), seguito il 30 marzo da un altro catalogo⁵⁴ comprendente 30180 voci di autorità di enti (similmente, l'insieme delle voci aventi allora livello pari o superiore a 90).

Un breve capitolo a parte meritano le voci di autorità dei luoghi: la nuova P10397 è stata dotata già il 1 marzo 2022 di un catalogo di Mix'n'match⁵⁵ comprendente tutte le 2625 voci allora esistenti. Tali voci sono state quasi tutte riconciliate con Wikidata già nella prima metà di marzo (i valori sono 2554 il 9 marzo)⁵⁶, in gran parte da Camillo Carlo Pellizzari di San Girolamo,⁵⁷ che ha parallelamente compilato una lista di 371 probabili errori contenuti in tali voci; la gran parte di questi errori (218) è stata risolta entro il maggio 2022 in collaborazione con ICCU.⁵⁸ Allo stato attuale, quindi, Wikidata è stata riconciliata con oltre il 90% delle voci di autorità dei luoghi presenti nell'Opac SBN.

Grazie ai nuovi cataloghi Mix'n'match n° 5076 e 5137, che hanno quasi raddoppiato il numero di voci di autorità disponibili nel precedente catalogo n° 4629 sfruttando l'apertura delle voci di autorità di livello 90, la riconciliazione ha subito una significativa accelerazione: il 6 marzo 2022 i valori di P396 erano 47179,⁵⁹ in crescita di soli 3101 ID (circa 17 al giorno) rispetto ai 44078 di sei mesi prima,⁶⁰ mentre solo sei mesi dopo i valori hanno toccato quota 71132,⁶¹ in crescita di 23953 ID (circa 133 al giorno). Una parte rilevante di questa crescita (circa 9mila ID), come si può dedurre dall'analisi dei valori giorno per giorno,⁶² è dovuta all'attività dell'Automatic name/date matcher, un meccanismo interno a Mix'n'match capace di abbinare con sicurezza un ID a un elemento di Wikidata nel caso in cui essi abbiano in comune il nome (o una parte rilevante di esso), l'anno di nascita e l'anno di morte; tuttavia, anche non considerando questi abbinamenti, resta evidente come l'attività umana di riconciliazione sia molto più intensa dopo l'importazione dei nuovi cataloghi rispetto a prima.⁶³

Un ulteriore momento di svolta si ha il 15 novembre 2022, quando l'URL di formattazione viene modificato;⁶⁴ come evidenziato nel qualificatore esplicativo, questo “non è l'URL di formattazione standard, ma permette di accedere ai dati di qualunque voce di autorità”.⁶⁵ La questione, in realtà, era più complessa: l'Opac SBN permetteva di raggiungere i VID tramite 5 distinti URL, raggrup-

⁵² [https://opac.sbn.it/nome/\\$1](https://opac.sbn.it/nome/$1) (la modifica è <https://www.wikidata.org/w/index.php?diff=prev&oldid=1584553691>).

⁵³ <https://mix-n-match.toolforge.org/#/catalog/5076>.

⁵⁴ <https://mix-n-match.toolforge.org/#/catalog/5137>.

⁵⁵ <https://mix-n-match.toolforge.org/#/catalog/5063>.

⁵⁶ <https://www.wikidata.org/w/index.php?oldid=1590240743>.

⁵⁷ <https://bambots.brucemyers.com/NavelGazer.php?property=P10397>.

⁵⁸ https://www.wikidata.org/wiki/Property_talk:P10397#Segnalazioni.

⁵⁹ <https://www.wikidata.org/w/index.php?oldid=1846018466>.

⁶⁰ <https://www.wikidata.org/w/index.php?oldid=1493643828> (6 settembre 2021).

⁶¹ <https://www.wikidata.org/w/index.php?oldid=1722609194> (6 settembre 2022).

⁶² In particolare da <https://www.wikidata.org/w/index.php?diff=prev&oldid=1589186603>.

⁶³ Cfr. i contatori degli abbinamenti (per mese e per nome utente) presenti in ciascun catalogo Mix'n'match.

⁶⁴ Il nuovo URL di formattazione è [https://opac.sbn.it/risultati-autori/-/opac-autori/detail/\\$1?core=autori](https://opac.sbn.it/risultati-autori/-/opac-autori/detail/$1?core=autori).

⁶⁵ <https://www.wikidata.org/w/index.php?diff=1772061684&oldid=1742211216>.

pabili secondo l'afferenza a due distinti *core*.⁶⁶ I due *core* permettevano di accedere a due distinti gruppi di voci di autorità, secondo le seguenti regole:

- *core* autori: tutte le voci di autorità di livello di autorità 90 o superiore;
- *core* autoriall: tutte le voci di autorità, indipendentemente dal livello, a cui sia collegato almeno un record bibliografico presente nell'Opac SBN.

Il *core* autoriall si configurava *quasi* come un soprainsieme del *core* autori (poiché esistono alcune voci di autorità di livello 90 o superiore prive di record bibliografici nell'Opac SBN) e permetteva di accedere a *quasi* tutte le voci di autorità (eccetto appunto quelle prive di record bibliografici nell'Opac SBN).

Nonostante ricavare gli ID delle voci visibili solo nel *core* autoriall e non nel *core* autori fosse un'operazione macchinosa,⁶⁷ l'esistenza del *core* autoriall ha consentito di superare *in gran parte* uno dei due fondamentali problemi che avevano fortemente rallentato la riconciliazione delle voci di autorità di persone ed enti con Wikidata, vale a dire l'invisibilità del 90-95% delle medesime. Nel 2023 nei mesi di febbraio⁶⁸ e marzo⁶⁹ sono stati organizzati dal Gruppo Wikidata per Musei, Archivi e Biblioteche (GWMAB)⁷⁰ due concorsi a premi al fine di "migliorare l'interconnessione tra Wikidata e le voci di autorità dell'Opac SBN, in particolare per gli autori aventi una voce nella Wikipedia in italiano". Secondo le statistiche finali complessive, 21 utenti hanno partecipato attivamente ad uno o entrambi i concorsi, sono stati aggiunti in totale 11563 valori di P396 nel corso dei due mesi (in media 196 al giorno) e sono state fatte 399 segnalazioni valide di errori in SBN (in media 6,8 al giorno), quasi tutte risolte in meno di un mese in collaborazione con ICCU; tutto ciò a fronte di una spesa totale di 245 € (su 300 € preventivati), concessi da Wikimedia Italia all'organizzatore dei concorsi, Camillo Carlo Pellizzari di San Girolamo, tramite un microgrant.⁷¹ Alla fine dei concorsi, il 2 aprile 2023, i valori di P396 hanno toccato quota 96919, più del doppio rispetto al 6 marzo 2022; nel medesimo giorno, riprendendo il confronto già svolto in precedenza (cfr. supra 2.2), i valori di P269 erano 949642 e i valori di P691 erano 706941 (risulta evidente come, nonostante i progressi fatti, il ritardo accumulato sia ancora ampiamente da colmare).⁷²

⁶⁶ Esemplicando con il VID CFIV103849:

- *core* autori: 1) <https://opac.sbn.it/risultati-autori/-/opac-autori/detail/CFIV103849?core=autori> = 2) <https://opac.sbn.it/risultati-autori/-/opac-autori/detail/CFIV103849> = 3) <http://id.sbn.it/af/IT/ICCU/CFIV/103849> (reindirizza al n°2, usato dal VIAF) = 4) <https://opac.sbn.it/nome/CFIV103849> (reindirizza al n°2; permalink);
- *core* autoriall: 5) <https://opac.sbn.it/risultati-autori/-/opac-autori/detail/CFIV103849?core=autoriall>.

⁶⁷ Cfr. la guida all'aggiunta di P396 in Wikidata, aggiornata al 31 gennaio 2023 ([https://it.wikipedia.org/w/index.php?oldid=131829704#Quali tipi di voci di autorità SBN inserire](https://it.wikipedia.org/w/index.php?oldid=131829704#Quali_tipi_di_voci_di_autorit%C3%A0_SBN_inserire)).

⁶⁸ https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei_Archivi_e_Biblioteche/Concorso/2023/02.

⁶⁹ https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei_Archivi_e_Biblioteche/Concorso/2023/03.

⁷⁰ https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei_Archivi_e_Biblioteche.

⁷¹ https://wiki.wikimedia.it/wiki/Microgrant/2023/Premi_per_due_nuovi_concorsi_su_Wikidata/Relazione.

⁷² <https://www.wikidata.org/w/index.php?oldid=1866303459>.

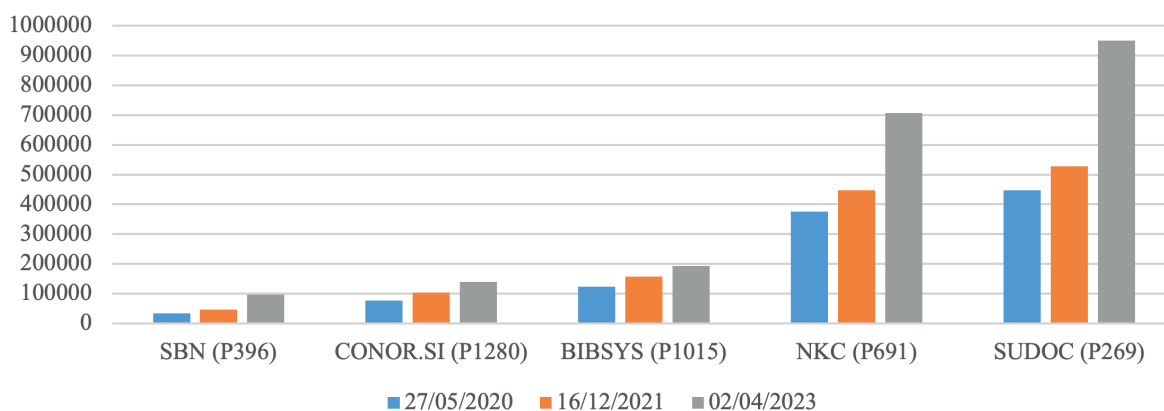


Tabella 1: numero di valori presenti in Wikidata (come dichiarazioni) per cinque proprietà di authority file europei membri del VIAF (fonte: template Number of main statements by property)

3. Conclusioni

La riconciliazione tra Wikidata e le voci di autorità dell'Opac SBN, ormai in corso da più di un decennio, ha accumulato, com'è evidente, un ampio ritardo rispetto a quella di altri authority file. La principale ragione, come detto, è stata senza dubbio l'invisibilità di oltre il 90% delle voci di autorità, che impediva un qualunque lavoro sistematico di riconciliazione (da parte di Wikidata, ma potenzialmente anche da parte di altri database). Tuttavia, la soluzione dei principali problemi che hanno causato tale ritardo (cfr. supra 2.2) apre prospettive molto positive per il futuro: tutti i VID (esclusi quelli di livello 05) sono ora normalmente visibili nell'Opac SBN e i VID soggetti a fusione ora reindirizzano ai VID con cui sono stati fusi. Inoltre, è prevista in Indice 3 l'introduzione del campo Unimarc/A 017, che permetterà ai catalogatori di strutturare l'ID Wikidata all'interno delle voci di autorità.

Finora purtroppo solo poche decine di persone hanno contribuito in modo rilevante alla riconciliazione delle voci di autorità di SBN con Wikidata,⁷³ ma i recenti concorsi di inizio 2023 hanno dimostrato come, anche con incentivi minimi, ci siano potenzialmente decine di persone pronte a dare un efficace contributo alla riconciliazione, e anche alla bonifica, delle voci di autorità di SBN. Nei prossimi mesi e anni la collaborazione tra ICCU e la comunità di Wikidata permetterà, soprattutto tramite l'inserimento di lezioni sull'uso di Wikidata nei corsi di formazione rivolti ai catalogatori della rete SBN, di allargare significativamente il numero di persone coinvolte in questa operazione, che recherà un significativo beneficio sia a Wikidata sia all'authority file di SBN, permettendo non solo di aumentare considerevolmente la quantità dei dati presenti in entrambi i database, ma anche di individuare e correggere errori e imprecisioni.

Per i prossimi anni sarebbe inoltre interessante imbastire un confronto periodico tra i dati presenti in Wikidata e quelli presenti nell'Opac SBN usando le API e gli SPARQL endpoint resi disponibili da entrambe le piattaforme. Grazie a una tale procedura sarebbe possibile sia valutare quali dati

⁷³ Solo 51 utenti, compresi 4 bot, hanno aggiunto almeno 100 valori di P396 a Wikidata (cfr. <https://web.archive.org/web/20230406193540/https://bambots.bruceymyers.com/NavelGazer.php?property=P396>).

siano presenti in uno solo dei due database e possano essere importati nell'altro sia individuare sistematicamente i dati discrepanti da sottoporre a una verifica umana per stabilire se uno dei due database necessita di essere modificato. Al fine della realizzazione di un *workflow* del genere sarà ovviamente fondamentale la già citata collaborazione tra ICCU e la comunità di Wikidata.

Riferimenti bibliografici

AIB - Gruppo di lavoro sulle biblioteche digitali. 2019. *Cosa SBN dovrebbe fare nel breve-medio periodo*. https://www.aib.it/wp-content/uploads/2019/07/Cosa_sbn_dovrebbe_fare.pdf.

Aste, Margherita, Maria Cristina Mataloni, e Luca Martinelli. 2015. «Linked data: il mondo di internet e il ruolo delle biblioteche, degli archivi e dei musei.» *DigItalia* 10 (1/2): 65–73.

Bianchini, Carlo, Stefano Bargioni, e Camillo Carlo Pellizzari di San Girolamo. 2021. «Beyond VIAF.» *Information Technology and Libraries* 40 (2). <https://doi.org/10.6017/ital.v40i2.12959>.

Bianchini, Carlo, Stefano Bargioni, e Camillo Carlo Pellizzari di San Girolamo. 2022. «Le voci di autorità dei nomi di persona in SBN e Alphabetica: problemi e prospettive.» *Bibliotheca.it* 11 (1): 247–314. <https://doi.org/10.6092/ISSN.2283-9364/15078>.

Buttò, Simonetta. 2022. «Welcome by the Director of the Central Institute for the Union Catalogue of Italian Libraries.» In *Bibliographic Control in the Digital Ecosystem*. A cura di Giovanni Bergamin e Mauro Guerrini, XXIII–XXV. Firenze: Firenze University Press. <https://www.torrossa.com/en/resources/an/5256281>.

Caffo, Rossella. 2013. «SBN tra presente e futuro.» *Biblioteche oggi* 31 (7): 8–13. <https://doi.org/10.3302/0392-8586-201307-008-1>.

Galeffi, Agnese, Andrea Marchitelli, Patrizia Martini, e Lucia Sardo. 2016. *Coming AUTH: A Project for Authority Data Enrichment and Deduplication in SBN*. <https://aibstudi.aib.it/article/download/11569/10888/20268>.

Guerrini, Mauro. 2013. «Per un nuovo catalogo SBN e per una nuova Bibliografia nazionale italiana.» *DigItalia* 8 (2): 185–90.

ICCU (Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche). 2020. *Norme per il trattamento di informazioni e dati comuni a tutte le tipologie di materiale. Authority file. Nomi. Nomi di persona*. https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file/Nomi.

ICCU, e Wikimedia Italia. 2015. «Convenzione tra ICCU e Associazione Wikimedia Italia - Associazione per la diffusione della conoscenza libera.» https://www.iccu.sbn.it/export/sites/iccu/documenti/2016/convenzione_Wikimedia.pdf.

Lucarelli, Anna. 2014. «“Wikipedia loves libraries”: in Italia è un amore corrisposto...» *AIB studi* 54 (2/3). <https://doi.org/10.2426/aibstudi-10108>.

Martinelli, Luca. 2014. «Wikidata: A New Way to Disseminate Structured Data.» In *Faster, Smarter and Richer. Reshaping the library catalogue*, Roma (Italy), February 27-28 February. <http://eprints.rclis.org/22754/>.

Martinelli, Luca. 2016. «Wikidata: la soluzione wikimediana ai linked open data.» *AIB studi* 56 (1): 75–85. <https://doi.org/10.2426/aibstudi-11434>.

Mataloni, Maria Cristina, e Egidio Incelli. 2022. «OPAC SBN: innovare i servizi nel solco di una solida tradizione.» *DigItalia* 17 (1): 58–65.

Vrandečić, Denny, Lydia Pintscher, e Markus Krötzsch. 2023. «Wikidata: The Making Of.» In *Companion Proceedings of the ACM Web Conference 2023*, 615–24. Austin TX USA: ACM. <https://doi.org/10.1145/3543873.3585579>.

Vullo, Giuseppina. 2014. «“Faster, smarter and richer. Reshaping the library catalogue”. La conferenza internazionale FSR in Vaticano.» *DigItalia* 9 (1): 111–20.

The evolution of authority work in SBN. From origins to Alfabetica and future prospects

Valentina Atturo^(a), Elena Ravelli^(b)

a) Archivio storico della Presidenza della Repubblica, <https://orcid.org/0000-0001-5614-2444>

b) ICCU, <https://orcid.org/0000-0001-6402-2039>

Contact: Valentina Atturo, v.atturo@quirinale.it; Elena Ravelli, elena.ravelli@cultura.gov.it

Received: 30 June 2023; Accepted: 18 October 2023; First Published: 15 January 2024

ABSTRACT

The article focuses on the main innovations and lines of action by ICCU on the authority headings in the SBN Opac, which open new perspectives on data integration. This work gives an account of the qualitative review activities carried out by the Working Group for the management and maintenance of the Authority File of SBN, the updates of SBN's authority headings rules among with their recent opening the Opac, the creation of redirects for merged headings, the implementation of LOD and API, and the innovations brought by the Alfabetica portal. These novelties will be followed closely by further advances that will be characterizing Index 3.

KEYWORDS

SBN; Alfabetica; Linked Open Data (LOD); Authority control.

L'evoluzione dell'*authority work* in SBN. Dalle origini ad Alfabetica e prospettive future

ABSTRACT

L'articolo mette a fuoco le principali innovazioni e le linee di intervento da parte di ICCU sulle voci di autorità dell'Opac SBN, che aprono nuove prospettive sulla integrazione dei dati. Si dà qui conto delle attività di revisione qualitativa compiute dal Gruppo di lavoro tecnico per la gestione e manutenzione dell'Authority file di SBN, degli aggiornamenti della normativa sulle voci di autorità di SBN, della recente apertura delle medesime nell'Opac, della creazione di reindirizzamenti per le voci accorpate, dell'implementazione di LOD e API, delle novità recate dal portale Alfabetica. Queste novità saranno seguite a stretto giro da ulteriori progressi che caratterizzeranno l'Indice 3.

PAROLE CHIAVE

SBN; Alfabetica; Linked Open Data (LOD); Authority control.

Nell'ambito di un lavoro condiviso, il paragrafo 2 è ascrivibile a Valentina Atturo e il 3 a Elena Ravelli. L'introduzione e le conclusioni sono il frutto di una riflessione comune delle due autrici.

1. Introduzione

Spetta a Rosella Caffo, precedente Direttrice dell'Istituto centrale per il catalogo unico (d'ora in avanti ICCU), il merito di aver per prima avviato una collaborazione con l'associazione Wikimedia Italia (d'ora in avanti WMI), corrispondente italiana della Wikimedia Foundation, la fondazione no-profit statunitense che gestisce i progetti Wikimedia, dando risposta all'esigenza di aprire maggiormente il Servizio bibliotecario nazionale (d'ora in avanti SBN) alla cooperazione e all'inter-scambio dei dati con il mondo dell'editoria, degli archivi, dei musei e, non ultima, di Wikipedia. In particolare, già nel 2014 era stato attivato un link che da Wikipedia rimandava alle registrazioni di autorità presenti nell'Opac SBN,¹ sfruttando la presenza di un permalink per permettere un più agevole puntamento alla voce di autorità e ai documenti ad essa collegati (Caffo 2013, 12). Del resto, già nel verbale del Comitato tecnico scientifico di SBN datato 11 luglio 2013² figurava tra i punti in discussione quello relativo agli sviluppi dei Linked Open Data e di Wikidata: Caffo comunicava in quella sede l'avvio di una collaborazione con la comunità italiana dei progetti Wikimedia e con Wikidata³ per realizzare un'ampia gamma di progetti, tra i quali, in prima istanza, l'aggancio attraverso il VIAF (Virtual International Authority File)⁴ dei dati tra l'authority file di SBN e Wikidata al fine di produrre una maggiore visibilità dei dati di SBN agli utenti.

Il 9 marzo 2015 fu siglato un accordo-quadro di collaborazione⁵ (ICCU e Wikimedia Italia 2015) tra ICCU e WMI: nell'ottica generale di "stabilire una collaborazione volta all'utilizzo e all'integrazione dei dati e dei materiali dei progetti ICCU con i progetti Wikimedia" (art. 1), le parti individuavano il primo progetto su cui sviluppare la propria collaborazione nell'Anagrafe delle biblioteche italiane,⁶ gestita dall'ICCU e la cui base dati era stata rilasciata con licenza Creative Commons Zero 1.0 Universal (CC0). Nello specifico, WMI si impegnava a integrare la base dati con i progetti Wikidata e OpenStreetMap, garantendo un flusso di aggiornamenti stabile e biunivoco (art. 2). Ancora in sede di Convenzione, l'ICCU si impegnava, in linea con la propria missione istituzionale, "a perseguire una politica progressiva di apertura e pubblicazione dei propri contenuti sotto licenze libere, coerentemente con quanto previsto dal presente accordo-quadro e dagli eventuali accordi integrativi, così come a promuovere tale approccio presso i propri Partner" (art. 3). L'accordo stabiliva inoltre di portare avanti ulteriori iniziative in futuro, volte a diffondere i principi dell'Open access e dell'Open culture.

¹ <https://opac.sbn.it/>. Ultimo controllo della sitografia nel mese di agosto 2023.

² https://www.iccu.sbn.it/export/sites/iccu/documenti/2014/CTS_SBN_Verb11-7-13/VERBALE_Cts_11072013.pdf.

³ <https://www.wikidata.org/>.

⁴ <https://viaf.org/>. Il VIAF combina le forme assunte dalle voci di autorità in sistemi diversi all'interno di un unico archivio di riferimento. Lo scopo del servizio è di abbassare i costi per le biblioteche e aumentare l'utilità delle voci di autorità creando associazioni e collegando gli archivi di autorità internazionali più autorevoli e rendendo tali informazioni disponibili online. Tutte le descrizioni di una determinata entità sono raccolte in un grappolo (*cluster*) che riunisce le differenti denominazioni dell'entità. Questo servizio consente agli utenti di identificare nomi, luoghi, opere ed espressioni pur mantenendo le forme preferite a livello locale per quanto riguarda forma, lingua e alfabeto. Il VIAF, implementato e gestito da OCLC (Online Computer Library Center), è un progetto che vede la collaborazione di decine di biblioteche nazionali (tra le altre, la Library of Congress, la Deutsche Nationalbibliothek e la Bibliothèque Nationale de France) e un numero crescente di altre biblioteche, agenzie catalografiche e altre istituzioni in tutto il mondo.

⁵ I risultati della collaborazione ottenuti sui progetti Wikimedia sono elencati, almeno in parte, in <https://it.wikipedia.org/wiki/Progetto:GLAM/ICCU>.

⁶ ABI: <https://anagrafe.iccu.sbn.it/>.

Questa convenzione, valida per tre anni, è stata rinnovata nel 2018 per altri tre anni, con l'impegno a "proseguire la collaborazione negli ambiti finora individuati" e a "valutare eventuali ampliamenti anche ad altri ambiti di collaborazione" (art. 2) (ICCU e Wikimedia Italia 2018); un secondo rinnovo è stato siglato nel 2022 (ICCU e Wikimedia Italia 2022), quando l'ICCU e WMI, nel ribadire il proprio formale impegno a continuare nel solco dei precedenti accordi volti al riutilizzo e all'integrazione dei dati e dei materiali dei progetti dell'Istituto con i progetti Wikimedia e con OpenStreetMap, convenivano (art. 2) di proseguire la collaborazione negli ambiti già individuati, fra cui, a titolo esemplificativo:

- il lavoro di collegamento e sincronizzazione con l'ABI e con le voci di autorità di SBN e lo sviluppo della piattaforma wiki relativa alle Normative catalografiche;⁷
- la valutazione di eventuali ampliamenti anche ad altri ambiti di collaborazione, sia relativi alle basi dati gestite dall'ICCU, sia relativi a progetti in essere o futuri compatibili con lo spirito dell'accordo;
- l'organizzazione e la promozione di iniziative volte a diffondere i principi dell'Open access e dell'Open culture, anche nel quadro della collaborazione nel progetto Europeana, la piattaforma digitale dell'UE che dà accesso al patrimonio culturale proveniente da oltre 3.500 istituzioni tra musei, biblioteche e archivi di tutta Europa.⁸

2. Brevi premesse: dalle origini del lavoro di autorità (authority work) in SBN agli scenari attuali

Scopo delle attività di *authority* è, come noto, quello di "consentire la disambiguazione tra entità (persone, enti, opere) distinte aventi denominazione identica o molto simile", "raggruppare le pubblicazioni che abbiano caratteristiche logiche comuni", nonché "assicurare l'omogeneità formale di ciascuna voce" (Weston 2015, 281-282). Del resto, "il controllo della forma del punto di accesso permette la struttura sindetica (relazionale) del catalogo agevolando i collegamenti tra le risorse che la biblioteca seleziona, raccoglie e rende fruibili per i propri utenti" (Guerrini 2003, 13).

Presso l'ICCU, il lavoro di autorità (*authority work*)⁹ si inserisce nella più ampia cornice istituzionale volta all'elaborazione e alla promozione delle normative catalografiche per le biblioteche italiane in accordo con gli standard internazionali e con l'evoluzione dello "scenario evolutivo dell'interoperabilità e dell'accessibilità nel web dei dati catalografici e dei contenuti digitali" (Caffo 2013, 10).

2.1 La collaborazione con il VIAF e i "livelli" delle registrazioni

In particolare, già dal 2009 l'Istituto collabora con il VIAF inviando, nel formato UNIMARC/A (d'ora in avanti U/A), le voci di autorità con livello 97 (*Authority*) e 95 (*Super*), vale a dire record

⁷ https://norme.iccu.sbn.it/index.php?title=Normative_catalografiche.

⁸ <https://www.europeana.eu/it>.

⁹ Come noto, la bibliografia sul controllo di autorità è copiosa. Per una ricognizione sul panorama italiano e sul contesto internazionale, si veda (Guerrini 2003). Facciamo qui riferimento al lavoro di autorità come alla "pratica di sottoporre la creazione e la gestione dei punti di accesso alle descrizioni bibliografiche a una serie di accertamenti volti ad assicurare l'uniformità e l'inequivocità", secondo la definizione di (Weston 2015, 281).

corrispondenti a voci controllate ed esaustive, mentre da gennaio 2023 l'Istituto fornisce al VIAF anche le voci di autorità di livello 90 (*Massimo*) per accrescere le visualizzazioni dei record italiani¹⁰. Come noto, nell'ambito di SBN con il concetto di *livello di autorità* – rispetto al quale è in corso un'approfondita riflessione presso l'ICCU volta a un complessivo ripensamento e aggiornamento di criteri ormai non pienamente efficaci¹¹ – si intende il dato che “esprime con un codice numerico il livello di affidabilità e di completezza della catalogazione fornita e si applica a qualsiasi documento, titolo di accesso o voce di autorità”¹². Il “livello di autorità attribuito al record in fase di creazione o di correzione è oggetto di specifici controlli da parte del sistema, in quanto:

1. non può superare il livello massimo di autorità attribuito al Polo;
2. se è minore o maggiore di 90, il record può essere modificato soltanto da chi ha un livello di autorità pari o superiore;
3. se è uguale a 90, il record può essere modificato soltanto da chi ha un livello di autorità superiore oppure dal Polo stesso che gli ha attribuito il livello 90” (Ibid.).

La corrispondenza del codice rispetto alla compilazione della voce di autorità si inserisce all'interno di un *range* numerico che si snoda tra i due estremi del livello minimo 05 e del livello massimo 97. Si faccia riferimento allo schema che segue:

- 05: il livello 05 è predisposto per le retroconversioni da cataloghi o da repertori, i dati previsti sono quelli del trattamento del nome a livello minimo (51), se presenti e desumibili dalla scheda o dal repertorio di riferimento, in particolare i dati da inserire, se possibile, sono quelli relativi ai codici di qualificazione e i dati necessari alla disambiguazione del nome;
- 97: il livello 97 è previsto per le notizie per le quali sono state effettuate tutte le attività di controllo sui repertori e di validazione dei dati e per le quali è stata controllata la congruenza dei titoli collegati. In particolare, per le voci di autorità, devono essere implementati, possibilmente, tutti i codici di qualificazione, le datazioni, le forme varianti necessarie alla ricerca della voce nel catalogo, le fonti, l'*International Standard Name Identifier*¹³ e la nota informativa.

Il livello 96 corrisponde a una fase di “lavorazione” e si applica a quelle registrazioni per le quali si stanno predisponendo i dati per innalzare il record a livello di *authority*. Il codice 95, invece, “indica un livello di autorità che non si differenzia dal livello 90 per contenuti, ma che, essendo attribuito soltanto ai Poli delle due biblioteche nazionali centrali, di Roma e Firenze, consente di tutelare la qualità dei dati destinati a essere pubblicati sulla Bibliografia nazionale italiana”.¹⁴ Da un punto di vista qualitativo della voce di autorità, il codice 90 non si discosta dal livello 97 per i dati

¹⁰ Sono naturalmente esclusi dall'invio al VIAF i VID qualificati come <omonimi non identificati>.

¹¹ Come infatti scrivono (Bianchini, Bargioni, Pellizzari da San Girolamo, 2022): “I livelli gerarchici delle voci di autorità e le limitazioni alla loro modifica basate sui livelli di autorità dei catalogatori sono una pratica utile e sostenibile nello scenario bibliografico attuale?”. Trattasi, evidentemente, di una domanda non banale la cui risposta, in termini di riflessione condivisa tra catalogatori e utenti del Servizio bibliografico nazionale, non può eludere le premesse storiche e le ricadute in termini di prassi e di adeguamento tecnologico per l'Indice SBN. Anche, e soprattutto, in termini di gestione del progresso e di sedimentazione catalografica.

¹² https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Codici/Informazioni_di_servizio/Livelli_di_autorit%C3%A0.

¹³ ISNI: <https://isni.org/>.

¹⁴ https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Codici/Informazioni_di_servizio/Livelli_di_autorit%C3%A0. Da notare però che l'Istituto centrale per i beni sonori ed audiovisivi (ICBSA), esclusivamente per il materiale audiovisivo, attribuisce il livello 95 alle risorse catalogate e lavorate in authority.

da inserire nella registrazione e/o per i flussi di lavorazione complessivi (controllo e identificazione dell'entità a partire da fonti repertoriali, gestione delle fusioni rispetto ai duplicati, spostamento dei titoli incongruenti, creazione delle forme varianti del nome quali ulteriori punti di accesso), ma presuppone un grado, seppur minimale, di incompletezza rispetto al codice 97.

Interessa notare che i livelli 96 e 97, “riservati ai record di autorità in fase di lavorazione o già lavorati, possono essere attribuiti ai record solo da utenze personali che intervengono sulle voci di autorità non attraverso l'aggiornamento contestuale della base dati di Polo, ma agendo direttamente sul catalogo collettivo con modalità di lavoro centralizzate” (Ibid.).

2.2 Il Gruppo di lavoro tecnico e l'aggiornamento delle normative per l'Authority File

Una porzione significativa delle utenze personali di livello 95 o 97 è stata assegnata dall'ICCU ai componenti del *Gruppo di lavoro tecnico per la gestione e manutenzione dell'Authority file di SBN. Nomi di persona*,¹⁵ finalizzato al controllo, alla bonifica e alla manutenzione dell'archivio delle voci di autorità relative ai nomi di persona in SBN. Il Gruppo di lavoro, coordinato dall'ICCU e trasversale a numerose istituzioni dislocate sul territorio italiano, è stato costituito il 9 luglio 2020 e vi partecipano bibliotecari¹⁶ dell'Istituto e di diverse istituzioni, cui potranno successivamente aggiungersi anche altri componenti interessati a partecipare. Per garantire uniformità di lavoro rispetto agli standard e ai protocolli sviluppati dall'Istituto nel corso degli anni, a tutti i componenti del *Gruppo di lavoro tecnico* è stata erogata dal personale dell'Ufficio *authority* un'apposita formazione collettiva e individuale, sia in presenza sia in modalità con collegamento remoto. Oltre al corso in presenza sul lavoro di autorità in SBN svoltosi presso l'ICCU nel 2019,¹⁷ i componenti del *Gruppo di lavoro tecnico* si sono potuti avvalere di documentazione tecnica predisposta dai referenti dell'Istituto per questo genere di attività.

Per il trattamento dei nomi personali in *authority*, oltre alle REICAT (*Regole italiane di catalogazione*), il perno catalografico attorno al quale ruota il lavoro di autorità è costituito dalle normative relative alla registrazione dei Nomi di persona pubblicate dall'ICCU nel 2020.¹⁸ Da un punto di vista strettamente operativo, è possibile inoltre fare riferimento alle *Linee guida per la compilazione delle registrazioni di autorità dei nomi di persona in SBN*¹⁹ che contengono indicazioni pratiche per la registrazione delle voci controllate in Interfaccia Diretta, un'applicazione che consente di operare sulla banca dati centrale del sistema Indice SBN, interrogando, modificando e fondendo record. Sono state successivamente elaborate, a partire da casi dubbi e problematici discussi durante le riunioni del *Gruppo di lavoro tecnico*, delle FAQ sul trattamento dei nomi in SBN mediante Interfaccia Diretta, in progressivo sviluppo.²⁰ Poiché un aspetto centrale del lavoro di autorità è rap-

¹⁵ <https://www.iccu.sbn.it/export/sites/iccu/documenti/2020/9196700Costituzione-gruppo-di-lavoro-AF....pdf>.

¹⁶ Con l'unica eccezione di Camillo Carlo Pellizzari di San Girolamo, dottorando di ricerca presso la Scuola Normale Superiore, cui è stata assegnata dall'ICCU un'utenza personale di livello 95 per lavorare sulle duplicazioni dell'Indice SBN e al quale è stato proposto di partecipare alle attività del Gruppo dopo una formazione personale erogata dal personale dell'Ufficio *Authority*.

¹⁷ https://www.iccu.sbn.it/export/sites/iccu/documenti/2019/AUT_2019_corso2.pdf.

¹⁸ <https://www.iccu.sbn.it/it/eventi-novita/novita/Normative-per-la-registrazione-di-Authority-dei-Nomi-di-persona/>.

¹⁹ <https://www.iccu.sbn.it/export/sites/iccu/documenti/2020/Linee-guida-AF-nomi-di-persona-062020.pdf>.

²⁰ https://www.iccu.sbn.it/export/sites/iccu/documenti/2021/FAQ-Authority-Nomi_SBN_DEF.pdf.

presentato dall'adeguamento²¹ alle REICAT per quanto concerne le forme dei Nomi di persona – con specifico riguardo al trattamento normalizzato dei numeri romani, della punteggiatura, delle qualificazioni e dei titoli che accompagnano un nome –,²² l'attività del *Gruppo di lavoro tecnico* si è soprattutto concentrata sulla bonifica di stringhe testuali afferenti a gruppi compatti di nomi (es. re, regine, santi, imperatori, papi...) in conformità con le nuove regole mutate dagli standard nazionali²³ e internazionali²⁴.

2.3 Linked Open Data e API per l'interrogazione e il riuso dei dati dei servizi bibliografici nazionali

Ora, a differenza di altre agenzie catalografiche nazionali preposte alla creazione e all'esposizione di *authority data* che si avvalgono centralmente di ingenti risorse umane e finanziarie convogliate con progetti *ad hoc* sui dati di autorità, l'esperienza italiana si caratterizza piuttosto per il suo essere, in primo luogo, una catalogazione “partecipata”, in cui la gestione dell'*authority file* condiviso dell'Indice SBN si inserisce – o meglio, si *dovrebbe* inserire – nel lavoro quotidiano del catalogatore. Il che presuppone non solo la presenza di “catalogatori appositamente formati, ai quali viene richiesto di conoscere le normative catalografiche e le modalità di inserimento dei dati all'interno della struttura del formato elettronico, ma anche di attenersi a un protocollo prestabilito nell'eseguire i controlli sulla correttezza e sulla completezza delle informazioni” (Weston 2015, 304). Se dunque in Francia,²⁵ in Germania²⁶ e in altre significative realtà internazionali gli scenari sono caratterizzati dalla progettazione e realizzazione, da parte di una forte *governance* centrale, di ambiziosi progetti catalografici volti al controllo, all'apertura e alla condivisione di dati qualitativamente verificati, il panorama italiano si caratterizza per un'impostazione di lavoro molto differente, perché diverse ne sono in primo luogo le premesse storiche. In SBN, infatti, l'*authority work* prende avvio nei Poli SBN²⁷ suddivisi per regioni, ai quali spetta il compito di coordinare il lavoro delle biblioteche afferenti al Polo, mentre

²¹ Superando dunque, a livello catalografico, le principali aporie esistenti tra le REICAT e l'Indice SBN prima della stesura delle nuove Norme ICCU, aporie esplicitate tra l'altro in Petrucciani, Turbanti, 2009, in particolare p. 9, ma ora in gran parte risolte.

²² Si vedano, a titolo puramente esemplificativo, VEAV017537 Gregorius VII <papa> (forma accettata corrente) e SBNV069011 Gregorius <papa ; 7.> (rinvio dalla forma usata nella precedente normativa); BVEV040314 Paulus IV <papa> (forma accettata corrente) e SBNV071349 Paulus <papa ; 4.> (rinvio dalla forma usata nella precedente normativa); RAVV059612 Vittorio Emanuele III <re d'Italia> (forma accettata corrente) e RAVV684564 Vittorio Emanuele <re d'Italia ; 3.> (rinvio dalla forma usata nella precedente normativa); CFIV078825 Karl V <imperatore> (forma accettata corrente) e RAVV652699 Karl <imperatore ; 5.> (rinvio dalla forma usata nella precedente normativa). Si ringrazia Carla Baccini per aver fornito una possibile lista di esemplificazioni.

²³ <https://www.iccu.sbn.it/it/normative-standard/>.

²⁴ Cfr. IFLA. «Names of persons» <https://www.ifla.org/g/cataloguing/names-of-persons/>; IFLA. «Guidelines for Authority Records and References», GARR <http://archive.ifla.org/VII/s13/garr/>; IFLA. «Dichiarazione di Principi Internazionali di Catalogazione» https://repository.ifla.org/bitstream/123456789/85/1/icp_2016-it.pdf 5. Punti di accesso.

²⁵ La Bibliothèque nationale de France rende i suoi dati “gratuiti e aperti” al fine di “accroître la visibilité des données de la BnF, par une meilleure exposition sur le Web, fédérer les données de la BnF, au sein et au-delà des catalogues”. In tal senso, la BnF intende: “contribuer à la coopération et l'échange de métadonnées par la création de liens entre des ressources structurées et de confiance; faciliter la réutilisation des métadonnées (sous Licence ouverte) par des tiers”.

²⁶ https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.

²⁷ <https://www.iccu.sbn.it/it/SBN/poli-e-biblioteche/index.html>.

all'ICCU compete sia l'aspetto di supervisione nazionale, sia la possibilità di progettare specifiche iniziative di carattere complessivo, con particolare attinenza all'aggiornamento delle Norme catalografiche per le voci di autorità e alla creazione di Gruppi di lavoro.²⁸

Dal punto di vista dell'*apertura* dei dati, merita di essere segnalato che, in ottemperanza alle politiche sui dati aperti che governano l'agire della pubblica amministrazione,²⁹ l'ICCU ha reso disponibili i Linked Open Data (LOD)³⁰ e le API³¹ per l'interrogazione e il riuso dei dati dei servizi bibliografici nazionali. I punti di accesso alle risorse sono stati pubblicati con i relativi manuali, al fine di consentire il più ampio accesso al patrimonio culturale delle biblioteche e degli istituti culturali che alimentano le basi dati gestite dall'Istituto. Del resto, come stabilito dal Codice dell'Amministrazione Digitale (CAD),³² i dati aperti sono disponibili e utilizzabili da parte di chiunque, anche per finalità commerciali, in formato disaggregato; sono accessibili attraverso le tecnologie digitali, comprese le reti telematiche pubbliche e private, in formati aperti e provvisti dei relativi metadati. Sul sito dell'ICCU è stato inoltre pubblicato il manuale descrittivo del triplestore dell'Ecosistema digitale³³.

2.4 Azioni compiute per il miglioramento delle voci di autorità dei nomi di persona in SBN

Nell'ambito della più ampia cornice istituzionale fin qui delineata, volta a garantire la condivisione di dati, sono stati predisposti da parte dell'ICCU una serie di interventi finalizzati a migliorare la qualità dei dati di *authority*. In particolare:

- **Visibilità di tutti i VID in Opac SBN.** Nell'ottica di promuovere una sempre maggiore apertura e interoperabilità dei dati bibliografici, l'ICCU ha deciso³⁴ di rendere visibili e accessibili in Opac SBN tutte³⁵ le voci relative agli autori presenti in Indice SBN, indipendentemente dal loro livello di autorità. Fino al mese di giugno 2023, le voci visibili in OPAC erano quelle con livello 90, 95 e 97. Grazie a questa nuova, importante, apertura, che rende disponibili anche le voci a livello 71 (*Medio*) e 51 (*Minimo*), gli utenti dell'Opac SBN avranno la possibilità di accedere a tutte le informazioni, anche minime, inserite nelle registrazioni relative ai nomi di persona e di ente. Le altre basi dati, inoltre, potranno creare dei collegamenti con tutti i record SBN relativi ai nomi di persona e di ente e con i dati ad essi correlati. L'auspicio è che l'interoperabilità e la collaborazione con altre piattaforme possa portare a un miglioramento reciproco in termini di integrazione, correzione e aggiornamento dei dati.
- **Aggiornamento delle Norme catalografiche con attenzione alle Datazioni.** Come noto, nel momento in cui scriviamo,³⁶ il dato relativo alle specificazioni cronologiche nelle voci di autorità di SBN popola i campi 200\$f e 300 U/A. In particolare, coerentemente

²⁸ <https://www.iccu.sbn.it/it/normative-standard/authority-control/>

²⁹ <https://www.dati.gov.it/contesto-normativo>.

³⁰ <https://triplestore.iccu.sbn.it/sparql>.

³¹ <https://api.iccu.sbn.it/devportal/apis>.

³² <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2005-03-07:82>.

³³ <https://www.iccu.sbn.it/export/sites/iccu/documenti/2022/Triplestore-SRI-descrizione.pdf>.

³⁴ E comunicato ai Poli con apposita Lettera siglata della Direzione dell'ICCU nel mese di giugno 2023.

³⁵ Con esclusione del livello 05.

³⁶ Maggio 2023.

con le indicazioni fornite nelle REICAT e nella normativa ICCU,³⁷ le qualificazioni cronologiche che affiancano la stringa testuale del nome (200\$f, AUTHORIZED ACCESS POINT – PERSONAL NAME, Dates) sono utilizzate per distinguere casi di omonimia, mentre le Datazioni³⁸ arricchiscono la voce di autorità con l'anno di nascita e/o di morte o con il periodo di attività della persona. Dal punto di vista dello standard catalografico, l'Ufficio Authority SBN, in accordo con tutto il personale dell'*Area di attività per l'elaborazione e diffusione degli standard e delle norme catalografiche*,³⁹ ha recentemente aggiornato la Tabella di trascrizione delle Datazioni al fine di realizzare un progressivo adeguamento allo standard internazionale U/A. Coerentemente con questo progressivo adeguamento con la sintassi prevista dal nuovo⁴⁰ campo 104 (CODED DATA FIELD: MAIN DATES OF ENTITY) dell'U/A, si è deciso di inserire unicamente informazioni veicolabili con caratteri numerici, demandando piuttosto alla Nota informativa⁴¹ eventuali specificazioni di carattere testuale (per es. l'indicazione dell'anno o del periodo di attività di un autore o di un tipografo, oppure l'indicazione del secolo in cui è vissuto un autore, verranno inseriti nella Nota informativa nel seguente modo: attivo nel 1860, attivo tra il 1900 e il 1920, vissuto nel XX secolo, etc.).⁴² Nell'ottica degli opportuni adeguamenti degli applicativi software alle Norme comuni⁴³ dell'ICCU e agli standard internazionali, il progetto Indice 3 dell'Istituto (e di cui diremo più approfonditamente in seguito) terrà conto di tali novità catalografiche.

- **Persistenza, tramite permalink, del VID soppresso tramite fusione.** In caso di fusione tra VID, il VID soppresso continua a essere tracciato come reindirizzamento al VID esito della fusione mediante permalink.⁴⁴
- **Individuazione di omonimi non identificati c.d. "occulti".** Il personale dell'ICCU svolge, con cadenza regolare, attività di bonifica, manuale e/o massiva, per l'individuazione (e successiva correzione) degli "omonimi non identificati" che sono erroneamente qualificati con nomenclature non standard (ad es. *omonimo non identificato, autore indifferenziato, omonimi indifferenziati, omonimo indifferenziato, omonimi indistinti*). Si tratta, evidentemente, di un lavoro in itinere e potenzialmente infinito rispetto al quale tutta la comunità dei catalogatori SBN è chiamata a collaborare.
- **Implementazione dei campi U/A.** Nell'ambito degli interventi di manutenzione evolutiva che saranno attuati per la realizzazione del progetto Indice 3, la presenza dei campi U/A sarà maggiormente strutturata al fine di granularizzare le informazioni presenti nella voce.

³⁷ [https://norme.iccu.sbn.it/index.php?title=Norme comuni/Authority file/Nomi/Registrazione di authority/Qualificazioni](https://norme.iccu.sbn.it/index.php?title=Norme%20comuni/Authority%20file/Nomi/Registrazione%20di%20authority/Qualificazioni)

³⁸ [https://norme.iccu.sbn.it/index.php?title=Norme comuni/Authority file/Nomi/Registrazione di authority/Datazioni](https://norme.iccu.sbn.it/index.php?title=Norme%20comuni/Authority%20file/Nomi/Registrazione%20di%20authority/Datazioni)

³⁹ <https://www.iccu.sbn.it/it/istituto/area-standard-norme-catalografiche-e-didattica/index.html>

⁴⁰ Introdotta dall'U/A nel 2021. Cfr. https://cdn.ifla.org/wp-content/uploads/U_A_104_update2021_ONLINE_FINAL.pdf

⁴¹ [https://norme.iccu.sbn.it/index.php?title=Norme comuni/Authority file/Nomi/Registrazione di authority/Nota informativa](https://norme.iccu.sbn.it/index.php?title=Norme%20comuni/Authority%20file/Nomi/Registrazione%20di%20authority/Nota%20informativa)

⁴² In caso di data incerta si potrebbe dover utilizzare il punto interrogativo (?), previsto dallo standard U/A come nono carattere aggiuntivo.

⁴³ [https://norme.iccu.sbn.it/index.php?title=Norme comuni](https://norme.iccu.sbn.it/index.php?title=Norme%20comuni)

⁴⁴ Come ad esempio: PUVV413378 Freyer, Johannes (accorpato) rinvia a TO0V064754 Freyer, Hans (accorpante).

Tra gli altri aspetti: sarà adottato il campo U/A 017 (OTHER IDENTIFIER);⁴⁵ saranno resi ripetibili i codici Lingua⁴⁶ (101 LANGUAGE OF THE ENTITY) e Paese (102 NATIONALITY OF THE ENTITY).

- **Visibilità, nell’Opac SBN, delle Fonti.** Per ciascun VID sono state sciolte le sigle dei repertori bibliografici presenti nel campo U/A 810 (SOURCE DATA FOUND) “Fonti”.
- **Arricchimento, tramite Wikidata, di VID.** Nel maggio 2023 ICCU ha fornito ai colleghi⁴⁷ di Wikidata uno scarico di 103148 VID (numero successivamente implementato a 106490) che avevano un link attivo da Wikidata. Tra questi VID sono stati selezionati quelli che erano privi contemporaneamente di Nota informativa (U/A 300), Fonti (U/A 810) e Nota del catalogatore (U/A 830), che sono risultati essere 20632. Dagli elementi di Wikidata abbinati a ciascuno di questi VID sono stati estratti nel giugno 2023 tramite query SPARQL⁴⁸ i valori delle seguenti proprietà, ove presenti: data di nascita (P569), data di morte (P570), luogo di nascita (P19), luogo di morte (P20), occupazione (P106), ordine religioso (P611), paese di cittadinanza (P27), ISNI (P213). L’ICCU sta ora procedendo a un’analisi a campione dei VID “arricchiti” restituiti con i valori desunti da Wikidata al fine di importarli nell’Indice SBN con l’indicazione della fonte da cui sono state tratte le informazioni aggiuntive.

3. Da Alfabetica all’Infrastruttura software del patrimonio culturale italiano (ISPC): nuove sfide per i dati di autorità

Con la pubblicazione del portale Alfabetica (<https://alfabetica.it/web/alfabetica>), l’ICCU ha scelto di valorizzare le proprie voci di autorità soprattutto mediante il canale dei Protagonisti (Castro 2022). Significativa, infatti, è l’opportunità di esplorare i dati di un personaggio non solo mediante la voce di autorità dell’Opac SBN, ma anche attraverso le informazioni presenti in Wikidata rispetto alla medesima entità, oltre che mediante altri riferimenti esterni.

Ad esempio, con riguardo al record corrispondente all’identificativo SBN IT\ICCU\CFIV\038491, che individua la poetessa e narratrice Ada Negri, il canale Protagonisti consente di visualizzare sia i dati specifici della voce di autorità in SBN (estremi cronologici, nota biografica), sia altre informazioni (come ad esempio l’indicazioni dei luoghi e delle date di nascita e morte) desumibili dai riferimenti esterni (collegamenti con Wikidata⁴⁹, con Wikipedia⁵⁰, con il VIAF⁵¹ e con l’ISNI⁵²).

⁴⁵ Cfr. https://cdn.ifla.org/wp-content/uploads/files/assets/uca/unimarc_updates/AUTHORITIES/a_017_new2017.pdf. “This field contains an identifier associated with the entity named in the 2XX field which cannot be accommodated in another field and a qualification which distinguishes between identifiers when more than one is contained in a record. The field also contains the source of the identifier”.

⁴⁶ Attualmente invece è un codice facoltativo, non ripetibile, che indica la lingua utilizzata dalla persona nel creare un’espressione destinata alla pubblicazione, alla diffusione ecc.

⁴⁷ L’ICCU ringrazia Camillo Carlo Pellizzari di San Girolamo e Stefano Bargioni per il prezioso lavoro svolto.

⁴⁸ Tutte le query usate sono raccolte in https://www.wikidata.org/w/index.php?title=Property_talk:P396&oldid=1923724268#-Query_per_estrazione_dati.

⁴⁹ <https://www.wikidata.org/wiki/Q346250>.

⁵⁰ https://it.wikipedia.org/wiki/Ada_Negri.

⁵¹ <https://viaf.org/viaf/73891042/>.

⁵² <https://isni.org/isni/0000000368580577>.

L'immagine della scrittrice è tratta da Wikimedia Commons.
Si vedano le Figure 1 e 2.

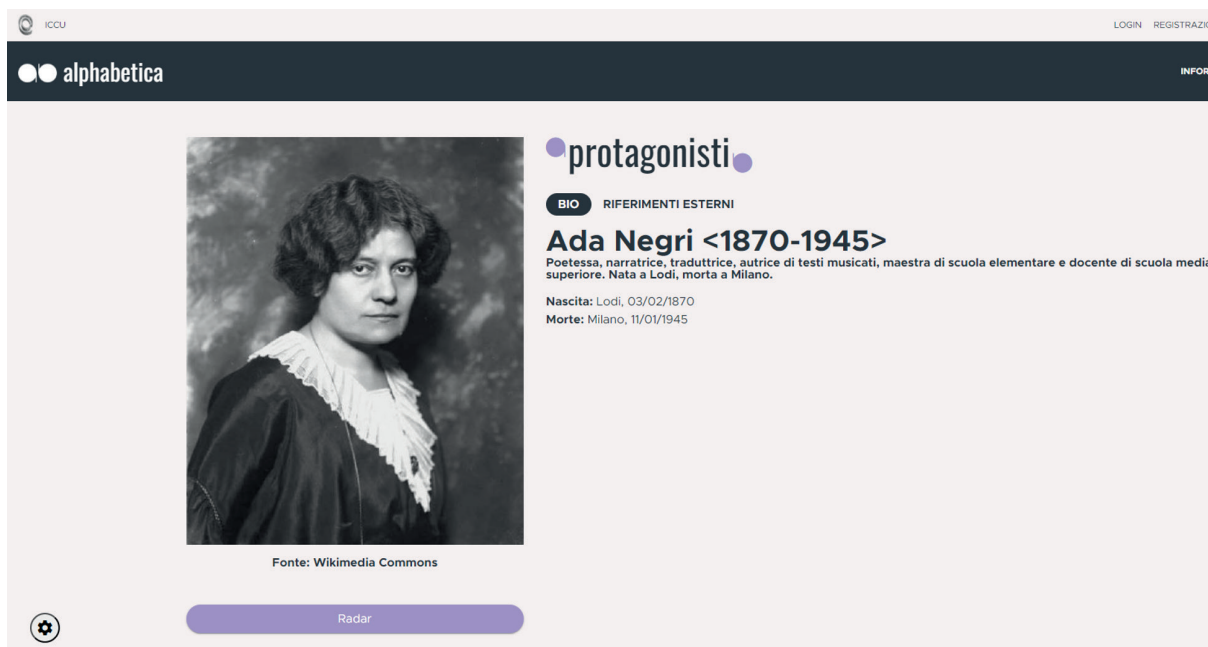


Figura 1. Alfabetica. Canale Protagonisti. Visualizzazione della registrazione di autorità dell'Indice SBN, arricchita dalle due voci "Nascita" e "Morte" che provengono da Wikidata.

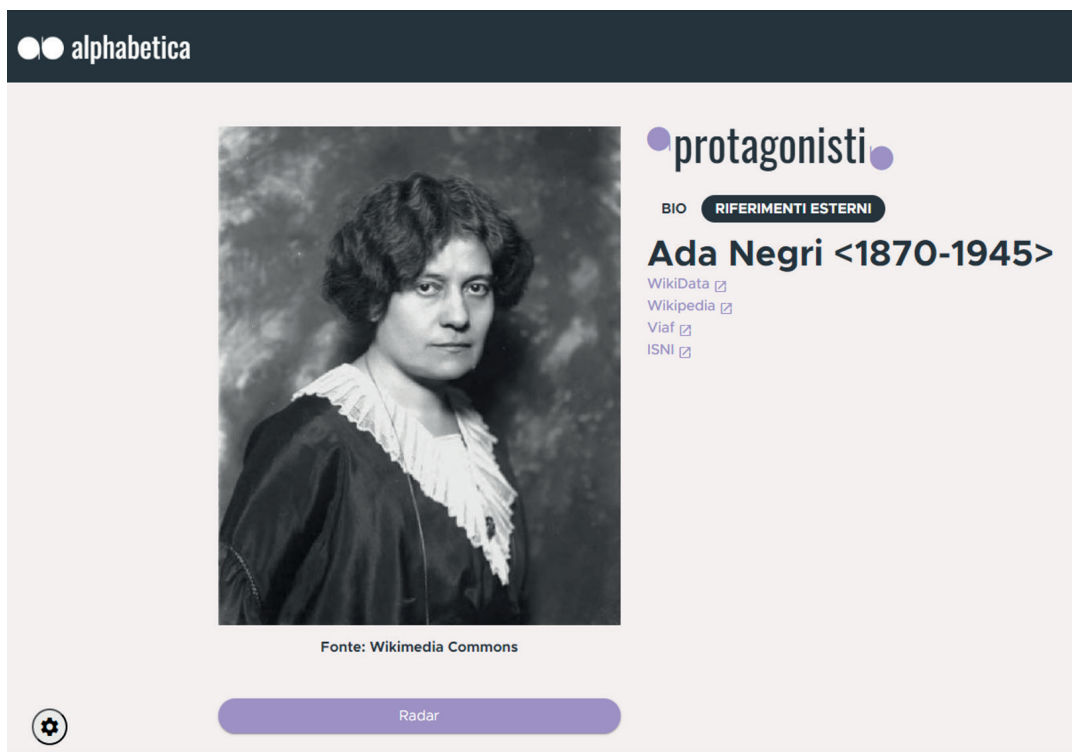


Figura 2. Alfabetica. Canale Protagonisti. Visualizzazione della registrazione di autorità dell'Indice SBN corredata da riferimenti esterni.

Di particolare interesse in Alfabetica lo strumento del cosiddetto “Radar semantico”, che consente di ampliare la restituzione dei risultati di ricerca, suggerendo all’utente “altre risorse” correlate con quella selezionata. Il Radar del canale Protagonisti consente all’utente una visualizzazione grafica delle principali relazioni dei protagonisti con le proprie opere e con quelle di altri personaggi illustri, con i luoghi e con le date delle loro manifestazioni più rilevanti. Compiendo, ad esempio, una ricerca sulla poetessa e narratrice Ada Negri, Alfabetica restituisce all’utente una sintesi della produzione editoriale dell’autrice in formato grafico, che può essere navigata espandendo il Titolo dell’opera per ottenere dettagli sulle singole manifestazioni. Il Radar mostra le principali relazioni del protagonista, a partire dalle opere di maggior rilievo (Figura 3).

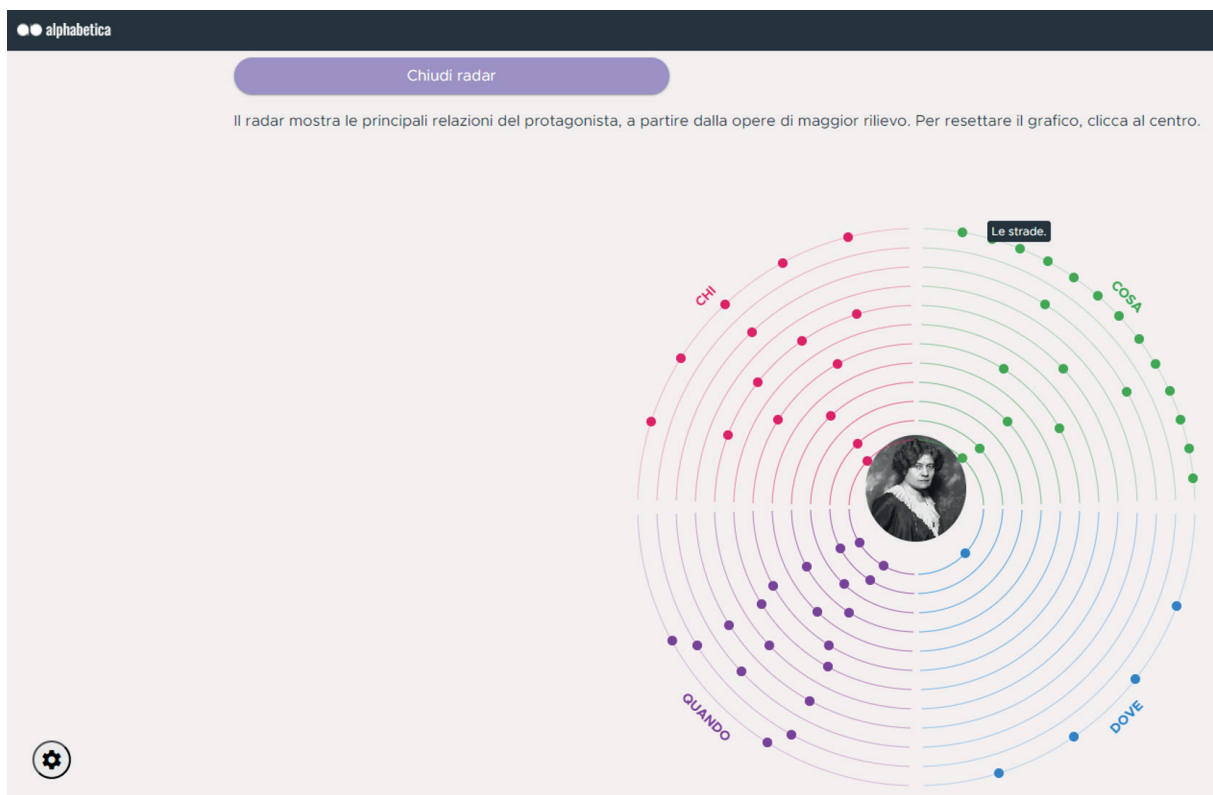


Figura 3. Alfabetica. Canale Protagonisti. Radar semantico.

Grazie all’integrazione con la banca dati Manus Online, l’utente potrà apprezzare anche il testo manoscritto dell’originale con firma autografa (Roma, Biblioteca Angelica, Manoscritti, ms. 2422/1-7. Firma autografa).⁵³

⁵³ <https://alfabetica.it/protagonisti-risultati/-/s/results?input=CFIV038491%21%3B%21Ada%21%3B%21Negri#navigation-1093>

3.1 L'evoluzione dei sistemi centrali del Catalogo collettivo del Servizio bibliotecario nazionale – il Progetto Indice 3

A partire dai primi mesi del 2023, l'ICCU sta lavorando alla realizzazione di un altro sfidante progetto finalizzato all'Evoluzione dei sistemi centrali del Catalogo collettivo del Servizio bibliotecario nazionale – il Progetto Indice 3. Tra gli obiettivi attesi figurano: la reingegnerizzazione tecnologica dell'Indice SBN utilizzando componenti open-source; il potenziamento della componente di indicizzazione documentale e quindi della capacità di ricerca sull'Indice; l'aumento della granularità delle attuali funzionalità esposte dalle API; la rivisitazione delle attuali funzioni con l'obiettivo di semplificare la profilazione dei Poli, eliminando le funzionalità desuete e implementandone di nuove. Il Progetto Indice 3 si innesta in un disegno progettuale più ampio che vede l'ICCU impegnato nella riorganizzazione delle basi dati di propria competenza. Il 4 ottobre 2018 infatti è stato pubblicato, come noto, il Bando di Gara Europea per la realizzazione del "Sistema di Ricerca Integrato e Catalogo generale delle risorse descritte nelle principali banche dati gestite dall'ICCU", da cui la realizzazione del portale Alfabetica, che è stato presentato al pubblico il 16 dicembre 2021 presso la Sala Giovanni Spadolini al Collegio Romano. Con la realizzazione di questo progetto, che costituisce uno strumento di visibilità nazionale del lavoro trentennale svolto dalle istituzioni bibliotecarie, si preannuncia un momento di grande innovazione per l'apertura dei servizi alle biblioteche per un'utenza sempre più ampia. Mediante la realizzazione di Indice 3, la componente applicativa dell'Indice SBN sarà reingegnerizzata secondo un rigoroso approccio a micro-servizi e contestuale ristrutturazione del livello di Storage (Database Management System DBMS SQL) che riguarderà database, *application server* e interfacce. Sono stati inoltre individuati altri due obiettivi che riguardano la completa revisione del pannello di amministrazione dell'Indice e la reingegnerizzazione tecnologica dell'Interfaccia Diretta dell'Indice.

All'interno di questo più ampio scenario, tra le nuove implementazioni previste dal progetto Indice 3 figurano anche (oltre alla gestione del legame di soggetto a livello di Titolo dell'opera⁵⁴ conformemente a quanto previsto dalle *Linee guida sull'indicizzazione per soggetto e la classificazione nel Servizio bibliotecario nazionale*)⁵⁵ interventi strategici e puntuali volti a migliorare le voci di autorità tra i quali, come già accennato, l'adozione del campo U/A 017, che permetterà l'inserimento di uno o più identificativi standard con specificazione del tipo di identificativo e del suo valore.

3.2 La gestione Authority file, vocaboli e thesauri nell'Infrastruttura software del patrimonio culturale (ISPC)

Nell'ambito della più generale cornice progettuale volta a definire l'Infrastruttura software del Patrimonio culturale (ISPC) a cura dell'Istituto centrale per la digitalizzazione del patrimonio

⁵⁴ Sarà implementata una funzione di spostamento del legame del soggetto dal BID del documento, ovvero dalla manifestazione (monografia, periodico, spoglio), al Titolo dell'opera, prevedendo inoltre che negli applicativi client sia resa visibile all'utente la presenza di un soggetto legato all'opera senza necessità di espandere il reticolo dell'opera stessa. Inoltre, quando si soggetterà la manifestazione il sistema controllerà che nell'opera collegata non sia già presente lo stesso legame alla stringa di soggetto. Questa funzione sarà resa disponibile in Interfaccia Diretta.

⁵⁵ https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Linee_guida_sull%27indicizzazione/FAQ_e_Appendici/Appendici/Progetti_ICCU_avviati_e_da_sviluppare.

culturale, Digital Library, è stato predisposto un modello concettuale per la componente di gestione Authority File, Vocabolari e Thesauri. Tale componente è parte integrante dei Processori di dominio e la sua funzione è quella di processare e mettere a disposizione *dataset* di *authority* e di strumenti terminologici che verranno inviati all'ISPC attraverso flussi MIC e non MIC (a prescindere dallo specifico dominio di riferimento). Gli strumenti terminologici di riferimento sono i Vocabolari e i Thesauri relativi a ogni singolo dominio (archivistico, bibliografico e relativo ai beni sonori e audiovisivi). La componente Authority File, Vocabolari e Thesauri è funzionale alla creazione di un punto di accesso unificato tramite la creazione del Catalogo generale di dominio, di grafi di conoscenza per singolo dominio e di grafi di conoscenza cross-domain.

L'obiettivo è una delle sfide principali del mondo della cultura e della sua catalogazione, ossia fornire un punto di accesso e di riferimento univoco, tale da consentire la fruizione e la navigazione trasversale di tutte le banche dati confluite nella piattaforma ISPC (e non solo, aggiungendo anche riferimenti esterni). La componente Authority File, Vocabolari e Thesauri sarà alimentata da: invio massivo asincrono (componente di ingestione); invio puntuale (interfacce di cooperazione applicativa) e allineamenti periodici per Vocabolari e Thesauri. I dati da processare saranno memorizzati in *DataLake* (repository dei dati archiviati) dai processi di acquisizione e notificati a ISPC.

4. Conclusioni

Come si desume dalle premesse storiche qui brevemente delineate e dagli scenari che contraddistinguono i più recenti sviluppi nell'ambito del lavoro di *authority* in SBN, certamente permangono dei ritardi rispetto alle potenzialità del Catalogo nazionale rappresentato dall'Opac SBN. Tuttavia, le azioni dispiagate recentemente contribuiranno senz'altro a velocizzare i processi di adeguamento rispetto agli standard internazionali. L'impegno profuso dall'ICCU, dai Poli e dalle più ampie comunità di bibliotecari e studiosi resta un perno imprescindibile per attivare delle sinergie su ampia scala che potranno innescare un circuito virtuoso volto al miglioramento delle voci di autorità.

Riferimenti bibliografici

- Bianchini, Carlo, Stefano Bargioni, e Camillo Carlo Pellizzari di San Girolamo. 2022. «Le voci di autorità dei nomi di persona in SBN e Alfabetica: problemi e prospettive.» *Bibliothecae.it* 11 (1): 247–314. <https://doi.org/10.6092/ISSN.2283-9364/15078>.
- Caffo, Rossella. 2013. «SBN tra presente e futuro.» *Biblioteche oggi* 31 (7): 8–13. <https://doi.org/10.3302/0392-8586-201307-008-1>.
- Castro, Elisabetta. 2022. «Alfabetica: uno strumento poliedrico di accesso ai servizi bibliografici nazionali.» *DigItalia* 17 (1): 39–47.
- ICCU, e Wikimedia Italia. 2015. «Convenzione tra ICCU e Associazione Wikimedia Italia - Associazione per la diffusione della conoscenza libera.» https://www.iccu.sbn.it/export/sites/iccu/documenti/2016/convenzione_Wikimedia.pdf.
- ICCU, e Wikimedia Italia. 2018. «Convenzione tra ICCU e Associazione Wikimedia Italia - Associazione per la diffusione della conoscenza libera.» <https://www.iccu.sbn.it/export/sites/iccu/documenti/2020/ICCU-Wikimedia.pdf>.
- ICCU, e Wikimedia Italia. 2022. «Convenzione tra ICCU e Associazione Wikimedia Italia - Associazione per la diffusione della conoscenza libera.» https://www.iccu.sbn.it/export/sites/iccu/documenti/2022/ICCU_Wikimedia_2022.pdf.
- Guerrini, Mauro. 2003. *Authority control: definizione ed esperienze internazionali: atti del convegno internazionale*, Firenze, 10-12 febbraio 2003, a cura di Mauro Guerrini e Barbara B. Tillett, con la collaborazione di Lucia Sardo. Firenze: Firenze University Press; Roma: AIB.
- Guerrini, Mauro. 2013. «Per un nuovo catalogo SBN e per una nuova Bibliografia nazionale italiana.» *DigItalia* 8 (2): 185–90.
- IFLA (International Federation of Library Associations and Institutions). 2017. «Names of persons» <https://www.ifla.org/g/cataloguing/names-of-persons/>.
- IFLA (International Federation of Library Associations and Institutions). 2001. «Guidelines for Authority Records and References», GARR <http://archive.ifla.org/VII/s13/garr/>.
- IFLA (International Federation of Library Associations and Institutions). 2016. «Dichiarazione di Principi Internazionali di Catalogazione» https://repository.ifla.org/bitstream/123456789/85/1/icp_2016-it.pdf.
- Mataloni, Maria Cristina, ed Egidio Incelli. 2022. «OPAC SBN: innovare i servizi nel solco di una solida tradizione.» *DigItalia* 17 (1): 58–65.
- Petruciani, Alberto, e Simona Turbanti. 2009. «Per l'applicazione delle REICAT: i cambiamenti nelle forme dei nomi degli autori.» *Biblioteche oggi* 27 (6): 7-22.
- Weston, Paul Gabriele. 2015. «Authority Data.» In *Biblioteche e biblioteconomia: principi e questioni*, a cura di Giovanni Solimine e Paul Gabriele Weston, 281-313. Roma: Carocci.

Subject indexing of normative acts

Ignazio Pirronitto^(a)

a) University of Florence, <https://orcid.org/0009-0001-7938-6617>

Contact: Ignazio Pirronitto, actius.13@hotmail.it

Received: 21 April 2023; **Accepted:** 25 October 2023; **First Published:** 15 January 2024

ABSTRACT

Subject indexing can be related to normative acts in the development of semantic metadata aimed at non-book resources. The survey demonstrates how the indexing process by subject can be applied to this special type of resources, so as to facilitate their management in a knowledge organization. The study attempts to apply the Italian system of subject indexing in each of its steps, and practical examples are suggested on how to provide semantic access points that enable the search for regulatory acts by subject. A specific in-depth study is devoted to bringing out all the semantic aspects that can be considered with regards to the subject of a norm and how this can be declined in order to bring out firstly, what the “aboutness” of the resource is, secondly, what the “isness” of a normative act is, and finally, what is required to create strings of subject built based on the “ofness” of the resource.

KEYWORDS

Legislative acts; Subject indexing; Nuovo Soggettario; Non-book resources.

L'indicizzazione per soggetto di atti normativi

ABSTRACT

L'indicizzazione per soggetto può essere applicata agli atti normativi e questo aspetto rappresenta un ulteriore sviluppo della metadattazione semantica rivolta a risorse non librerie. Il contributo dimostra come può essere applicato il processo di indicizzazione per soggetto a questo speciale tipo di risorse, così da favorirne la gestione in un sistema di organizzazione della conoscenza che ne permetta la descrizione e il recupero. Nello studio è tentata l'applicazione del sistema italiano di indicizzazione per soggetto in ogni suo passaggio e si suggeriscono pratici esempi su come fornire punti di accesso semantico che consentano la ricerca per argomento di atti normativi. È mostrato come effettuare l'analisi concettuale di un atto normativo soffermandosi su alcuni aspetti particolari e tipici del contenuto concettuale di questo tipo di risorse per procedere, in seguito, alla corretta individuazione dei ruoli di ogni termine al fine di creare stringhe di soggetto che siano in grado di rappresentare in modo completo il contenuto concettuale di una norma. Un approfondimento specifico è dedicato a far emergere quale sia l'*aboutness* dell'atto normativo, quale sia la *isness* e suggerire come creare stringhe di soggetto costruite basandosi sulla *ofness*.

PAROLE CHIAVE

Atti normativi; Indicizzazione per soggetto; Nuovo Soggettario; Risorse non librerie.

Introduzione

In una realtà in cui le possibilità di accesso alla conoscenza e all'informazione si moltiplicano in modo esponenziale diventa indispensabile trovare i giusti strumenti utili a gestire e organizzare le informazioni e i dati che quotidianamente ci vengono proposti. Il fatto che la Rete conceda a tutti di partecipare alla generale organizzazione della conoscenza per mezzo del social bookmarking e delle attività di tagging, operazioni possibili in un numero sempre crescente di siti internet e di social media, potrebbe condurre a un ripensamento riguardo alla necessità e all'efficacia dei sistemi di classificazione e d'indicizzazione adoperati da anni in biblioteche, archivi e musei. Lo sviluppo delle folksonomie¹, in quanto classificazioni di informazioni create dalla comunità degli utenti, se da un lato appaiono strumenti più accessibili e amichevoli, dall'altro lato non sono in grado di garantire un preciso e puntuale recupero delle informazioni, pertanto, proprio in questo aspetto è confermata l'utilità e la validità dei *Simple Knowledge Organization System* (SKOS). La sfida, in questo senso, non consiste solo nel classificare le informazioni ma nel riuscire a recuperarle con esattezza anche a distanza di tempo e questo scopo è raggiungibile soltanto tramite l'impiego di un sistema di organizzazione della conoscenza basato su principi oggettivi, condivisi e noti. Altro aspetto da considerare è il fatto che sistemi di classificazione e di indicizzazione semantica sono necessari per gestire non soltanto la 'risorsa libro' poiché il concetto stesso di risorsa ormai denota qualsiasi res che presenti un potenziale contenuto informativo per l'utente sia questa un libro, un brano musicale, un film, una fotografia o una scultura (ICCU 2023).

Accogliendo lo stimolo promosso dalla seconda edizione della Guida al Nuovo Soggettario (BNCF 2021) di ampliare le procedure di indicizzazione per soggetto anche a risorse non bibliografiche è proposta, a seguire, la questione relativa all'indicizzazione per soggetto di atti normativi, scelti appositamente per dimostrare la possibilità di descrivere il contenuto semantico anche di risorse dalla natura tanto particolare. Lo studio, condotto su un campione di dieci atti normativi, trae spunto dall'attività della Biblioteca Innocenti Library "A. C. Moro", sita presso l'Istituto degli Innocenti di Firenze, da anni impegnata nella catalogazione di atti normativi, e si pone l'obiettivo di mostrare come sia praticabile l'indicizzazione per soggetto di atti normativi soffermandosi sulle questioni e sulle riflessioni che scaturiscono dalla descrizione di questa particolare tipologia di risorse.

A ogni risorsa il suo Soggetto

Tra le questioni che hanno smosso negli anni gli animi di bibliotecari e di esperti del settore vi è la consapevolezza che ormai in biblioteca non siano fruibili solo risorse librarie ma queste cedono in modo costante un po' del loro spazio fisico e virtuale a risorse differenti per natura e supporto (Gorman 2018, 52). Le origini di questo sviluppo sono riconducibili a diversi fattori, quali la nascita di nuove tipologie di risorse che consentono la trasmissione di un contenuto informativo non più unicamente su supporto di natura cartacea ma anche in formato digitale per mezzo di

¹ Termine derivato dal vocabolo inglese *folksonomy* e nato dall'unione delle parole di 'folk' e 'tassonomia'. Rappresenta l'atto con cui gli utenti di Internet classificano liberamente le informazioni in essa reperite mediante l'impiego di parole chiave, definite tag. Vedi https://www.treccani.it/vocabolario/folksonomia_%28Neologismi%29/.

ebook, PDF e risorse online. Inoltre, lo sviluppo tecnologico consente la fruizione di un testo non più soltanto mediante lettura ma anche ascoltandolo direttamente dalla voce del suo autore o di un lettore, come è possibile fare grazie a un audiolibro. A ciò va aggiunto la tendenza delle biblioteche a praticare politiche di acquisizione delle risorse che rendano le proprie collezioni in grado di offrire una scelta più variegata, stimolante e adatta alle maggiori esigenze informative dei lettori, siano esse legate all'età o ad altri fattori. Da ciò derivano le buone pratiche di avere in biblioteca una versione per bambini di un'opera destinata agli adulti, magari una versione espressa in fumetti. Proporre una variegata gamma di risorse ha un ruolo positivo anche nel promuovere l'inclusione di utenti con bisogni specifici che possono fruire di testi stampati secondo i criteri di alta leggibilità, con grandi caratteri tipografici o espresse nel linguaggio della Comunicazione Aumentativa e Alternativa (CAA) (Costantino 2011). Infine, l'ultimo fattore da considerare sul tema è la metamorfosi in corso di molte biblioteche le quali assumono maggiormente le funzioni di centri di documentazione chiamati a custodire in modo centralizzato anche collezioni di altri enti e che spesso vanno ben oltre ai semplici fondi librari. In queste occasioni la mission della biblioteca viene estesa alla descrizione, valorizzazione e conservazione di risorse quali fotografie, film, oggetti di ogni tipologia materiale che si affiancano alle tradizionali raccolte librarie.

La presenza di risorse di natura diversa ha reso necessario, nel tempo, l'adeguamento delle norme e delle regole di catalogazione al fine di renderle efficaci anche nella descrizione di ognuna di esse. L'esigenza di tali integrazioni è giustificata dalla consapevolezza che ogni tipo di risorsa ha attributi propri che se non fossero opportunamente identificati e descritti potrebbero fornire all'utente informazioni non chiare o incomplete. Va notato che in questo aspetto il ramo della catalogazione descrittiva ha avuto molta più attenzione rispetto a quello della catalogazione semantica, ciò è testimoniato anche dalle diverse edizioni dello standard ISBD pubblicate per la descrizione di tipologie diverse di risorse. Queste, revisionate fino al primo decennio del Duemila, negli anni successivi hanno fatto emergere l'esigenza di dare vita a un nuovo standard che le unificasse tutti, armonizzandole e rendendo possibile l'impiego di un unico standard per tutte le risorse senza perdere la corretta descrizione della natura di ognuna di esse. In tal senso, l'obiettivo è raggiunto solo nel 2011 per mezzo della definitiva versione dello standard *ISBD Consolidated edition* (IFLA 2012). Un'ulteriore apertura nella descrizione di risorse diverse in biblioteca è rappresentata da RDA che dal 2010 si propone come standard internazionale per la descrizione e l'accesso alle risorse progettato per produrre dati di descrizione in modo compatibile con la tecnologia del web semantico. Essendo uno strumento rivolto non solo a biblioteche ma ad ogni istituzione che conserva e gestisce la memoria registrata, esso nasce già predisposto alla definizione dei dati necessari alla descrizione di risorse diverse per tipologia e natura. Sotto questo aspetto, un confronto tra i due standard permette di osservare che mentre ISBD impone di presentare i dati riguardo alla natura della risorsa seguendo un ordine rigido e non modificabile, il più recente standard RDA prevede soltanto l'obbligo di definire chiaramente la natura della risorsa, lasciando all'agenzia catalografica la libertà di presentare questa informazione con un testo, con un codice alfanumerico o con un'icona rappresentativa. (Guerrini e Bianchini 2016, 149).

Sul versante dell'indicizzazione per soggetto la situazione si presenta ben diversa, pertanto, limitandoci all'esperienza italiana dal secondo Dopoguerra, uno dei motivi che hanno portato a un trattamento non omogeneo delle risorse sotto questo aspetto è riscontrato nella scelta di assegnare il soggetto esclusivamente a opere che presentassero un potenziale contenuto informativo su un

dato argomento. Da ciò è derivata la prassi, sostenuta e alimentata principalmente da Carlo Revelli, secondo cui le opere di fiction e determinati generi letterari, quali la narrativa e la poesia, non andassero soggettate anche quando il tema in esse trattato fosse ben definito (Revelli 2004a, 315). La giustificazione di tale scelta risiedeva nel fatto che lo scopo principale di questo tipo di opere non fosse fornire informazioni su un dato argomento ma esprimere l'immaginazione e la creazione artistica dell'autore (Fedeli 2015). Questa politica catalografica, ancora ampiamente diffusa, era alimentata anche da una limitata consapevolezza delle potenzialità dell'indicizzazione per soggetto in quanto servizio all'utente, poiché era ponderata sul presupposto, dimostratosi poi inesatto, che l'utente avesse bisogno di accessi semantici per ottenere soltanto risorse di natura informativa legata a esigenze di studio o di ricerca. L'esperienza sul campo e le indagini sulle richieste degli utenti hanno dimostrato che insieme a opere su un dato argomento, molti utenti chiedono anche opere di un certo genere, per esempio un romanzo giallo o libri per bambini ma anche risorse di natura diversa, quale un film su un tema o film tratti da altre opere letterarie. In seguito, è lo stesso Revelli a rivedere le sue posizioni, riconoscendo il beneficio derivante dalla soggettazione di opere di fiction principalmente per l'utenza delle biblioteche di pubblica lettura (Revelli 2004b).

Il secondo fattore da prendere in considerazione è la consapevolezza che in Italia l'indicizzazione per soggetto è stata praticata molto raramente anche nei confronti di opere di antico regime tipografico o create prima del Ventesimo secolo, comportando la conseguente esclusione di una cospicua parte di patrimonio delle biblioteche italiane dai vantaggi derivanti dal recupero mediante soggetto. In questo caso, la scelta era stata prettamente di politica catalografica poiché lo standard internazionale relativo alle procedure di analisi concettuale delle opere (ISO 5963:1985) non prevede criteri di selezione cronologica. Successivamente anche nei confronti di questo tipo di risorse vi è stato un cambio di approccio, poiché, dal 2016 la Bibliografia Nazionale Italiana ha scelto di creare accessi per soggetto anche per le nuove edizioni di opere che sono state create prima del Ventesimo secolo e, attualmente, tale pratica è ampiamente diffusa anche tra le biblioteche aderenti al Servizio Bibliotecario Nazionale. A favore di questo punto nell'indicizzazione di questo tipo di risorse il sistema del Nuovo Soggettario prevede che sia impiegato il termine del Thesaurus 'Opere anteriori al 1900' da inserire nella stringa di soggetto con il ruolo di forma bibliografica.

L'ultimo fattore riguarda direttamente il soggetto in quanto elemento di descrizione, poiché, essendo individuato in base alla pertinenza di un'opera a un determinato argomento, per la maggior parte dei casi, è creato con il solo scopo di esplicitare l'*aboutness* della stessa. Non a caso il metodo di analisi concettuale praticato in Italia prevede che l'enunciato di soggetto punti a fornire una risposta precisa alla domanda "di che cosa parla/tratta quest'opera?". Solo negli ultimi anni sono stati meglio definiti e riconosciuti i vantaggi derivanti anche da altre forme di indicizzazione per soggetto i quali, oltre a indicare l'argomento dell'opera (*aboutness*) possano rappresentare ciò di cui l'opera è la rappresentazione (*ofness*) e ciò che l'opera è in base alla sua forma (*isness*).

Tornando sulla considerazione fatta in partenza riguardo all'esistenza in biblioteca di risorse diverse per natura e tipologia è possibile asserire, prendendo in prestito una delle regole di Ranganathan, che ad ogni risorsa deve essere dato il suo soggetto, espresso nella forma più utile e accessibile all'utente. Il superamento di alcune prassi del passato può consentire di individuare il giusto soggetto anche per un romanzo, un film, un quadro e persino un gioco da tavolo, rendendo l'indicizzazione per soggetto un tipo di descrizione efficace ed efficiente anche nei multiformi aspetti della metadattazione orientata verso le esigenze dell'ecosistema digitale e dell'universo bibliografico.

Catalogazione di atti normativi: una proposta per il futuro

Entrando nel cuore della questione, si intende puntare l'attenzione su una particolare tipologia di risorsa, ossia, l'atto normativo considerato in questa sede non sotto il suo più familiare aspetto giuridico bensì come risorsa informativa inserita tra le altre già note dell'universo bibliografico. Le riflessioni in seguito qui sviluppate prendono come spunto iniziale l'attività di catalogazione svolta dalla Biblioteca Innocenti Library di Firenze per conto del Centro nazionale di documentazione e analisi per l'infanzia e l'adolescenza, da anni impegnata nella catalogazione di atti normativi, attività indispensabile allo scopo di fornire a esperti e utenti un modo efficace e in costante aggiornamento che consenta loro di accedere al contenuto delle norme in vigore su materie relative all'infanzia e adolescenza, tematiche fondamentali per le attività dello stesso Centro.² La singolarità del servizio, che la biblioteca svolge tramite propri strumenti di indicizzazione semantica, emerge anche alla luce del fatto che le principali banche dati ad accesso libero di norme italiane quali la Gazzetta Ufficiale³ in versione telematica e Normattiva⁴ consentono soltanto la ricerca per materia e per parola o frase nel titolo o nel testo dell'atto normativo.

L'applicazione della descrizione bibliografica a norme e leggi potrebbe mutarsi in un servizio in grado di potenziare in modo notevole la risposta di biblioteche e di centri di documentazione nei confronti degli utenti. Uno dei suoi punti di forza risiede proprio nel fatto che esso permetterebbe all'utente di avere accesso diretto, mediante catalogo, alle principali fonti del diritto rendendo inoltre possibile il superamento del limite rappresentato dal fatto che attualmente in biblioteca è possibile trovare soltanto risorse 'sul diritto' o 'di giurisprudenza' e non le fonti propriamente dette, rappresentate da costituzioni, statuti, leggi, decreti e regolamenti. L'attività appare potenzialmente sostenibile anche nell'ottica del servizio di reference in quanto ricercare atti normativi per soggetto può soddisfare sia l'esigenza del giurista specializzato, interessato alle leggi su temi specifici, sia l'utente con bisogni informativi più generici. In altri termini, i vantaggi potenziali di questo servizio sarebbero i seguenti:

- Fornire l'accesso a norme tramite una 'ricerca per argomento' in grado di soddisfare le richieste dell'utente a prescindere dal suo livello di specializzazione.
- Consentire il recupero di norme di ambito internazionale, europeo, nazionale e regionale, dando la possibilità di conoscere come lo stesso diritto o principio giuridico sia declinato nei diversi ambiti territoriali.
- L'atto normativo è trattato a tutti gli effetti come risorsa bibliografica e la descrizione non prevede solo l'indicazione della sua esistenza in catalogo ma, a differenza delle banche dati di Enti legislatori, esso è descritto in un regime di *knowledge organization*. In questo aspetto risiede la più evidente unicità del servizio posto che sul territorio italiano, escludendo i fornitori di servizi a pagamento, l'unico progetto che ad esso si avvicina, sotto l'aspetto del recupero di risorse di natura legislativa tramite accesso semantico, è quello condotto dalla Biblioteca del Senato della Repubblica "Giovanni Spadolini" la quale mediante l'impiego di TESEO (TESauro SENato per l'Organizzazione dei documenti parlamentari) consente

² <https://www.gazzettaufficiale.it/eli/id/2007/07/23/007G0118/sg>

³ <https://www.gazzettaufficiale.it/>

⁴ <https://www.normattiva.it/>

la ricerca per argomento all'interno delle banche dati parlamentari, ma il servizio è limitato alla sola documentazione di iter di disegni di legge e a documenti di procedure non legislative, escludendo norme e leggi promulgate.⁵

Tipologie di norme catalogabili

Il trattamento catalografico di un atto normativo dovrebbe essere fondato su alcune scelte che l'agenzia catalografica è chiamata a definire sin da principio, prendendo in considerazione due importanti aspetti. Il primo è mirato a definire quali atti catalogare, chiarendo se saranno oggetto di indicizzazione semantica soltanto gli atti normativi primari, quali leggi o regolamenti dell'Unione europea, o se il trattamento è esteso anche a quelli secondari quali decreti o deliberazioni. Tale scelta è correlata anche all'ambito territoriale in cui l'atto produce i suoi effetti, pertanto, sotto questo aspetto, l'agenzia bibliografica è chiamata a decidere se catalogare in totalità gli atti di ambito comunale, regionale, nazionale o internazionale o solo atti selezionati e scelti su criteri prestabiliti. È bene precisare che i due aspetti presentano un legame inscindibile dovuto al fatto che alcune forme di atto normativo sono prodotte solo da determinati organi, come mostrato dal seguente elenco che presenta alcuni degli organi di maggior rilievo con alcune tipologie di atti da questi prodotti e che potrebbero essere oggetto di indicizzazione semantica.

Organo	Tipo di atto emanato
United Nations General Assembly (UNGA)	Decisioni, Raccomandazioni, Risoluzioni
United Nations Human Rights Council (UNHRC)	Risoluzioni
Human Rights Committee (HRC)	Commenti Generali, Pareri
Council of Europe	Decisioni, Raccomandazioni, Risoluzioni
Parlamento europeo	Decisioni, Direttive, Pareri, Posizioni, Regolamenti
Consiglio dell'Unione europea	Decisioni, Direttive, Pareri, Raccomandazioni, Regolamenti
Commissione europea	Decisioni, Direttive di esecuzione, Raccomandazioni, Regolamenti di esecuzione
Banca centrale europea	Decisione, Orientamenti, Pareri, Raccomandazione, Regolamenti
Parlamento italiano	Leggi, Leggi costituzionali
Presidenza della Repubblica	Decreti del Presidente della Repubblica
Governo italiano e Ministeri	Decreti legge, Decreti legislativi, Decreti del Presidente del Consiglio, Decreti Ministeriali, Regolamenti
Regioni e Province Autonome	Leggi regionali, Regolamenti regionali, Deliberazioni di Giunta/Consiglio

Figura 1. Esempi di Enti con relativi atti normativi emanati

⁵ <https://www.senato.it/tesauro/teseo.html>

Il secondo punto da definire riguarda la quantità, ovvero, la scelta di catalogare in toto gli atti normativi prodotti dagli organi selezionati o se optare per un secondo criterio di selezione che potrebbe essere di natura tematica o disciplinare, per fare un esempio, scegliendo di catalogare solo gli atti riguardanti le politiche sociali o solo atti che legiferano sul tema della sicurezza nei luoghi di lavoro o altri temi. In questo aspetto è possibile anche intravedere un'auspicabile collaborazione tra bibliotecario e giurista in un sistema virtuoso nel quale al secondo potrebbe essere affidato il compito di selezionare gli atti normativi da catalogare mentre al primo resta la parte tecnica della metadattazione.

Descrivere gli atti normativi

Le norme giuridiche sono una tipologia di risorsa del tutto particolare che impone al catalogatore di focalizzare la sua attenzione su alcuni aspetti specifici che le comuni risorse bibliografiche non presentano, tra queste, una prima particolarità è rappresentata proprio dal loro contenuto informativo. Se in biblioteca le risorse sono comunemente distinte in opere che trattano un tema e in opere che hanno contenuti di finzione o di fantasia, le norme presentano un contenuto particolare che spesso è più difficilmente descrivibile concettualmente con un solo tema. Ciò è legato a una delle caratteristiche tipiche della norma consistente nel fatto che il suo contenuto non è espresso soltanto in termini di “discorso su qualcosa” ma si manifesta come una regola o insieme di regole finalizzate a disciplinare in astratto un comportamento futuro di tutti i soggetti cui essa rivolge i propri effetti giuridici. Alla luce di questo è possibile dedurre che il contenuto informativo di una norma spesso non è un solo tema o argomento ma è una situazione, rilevante dal punto di vista giuridico, che deve ancora accadere e che concettualmente può coinvolgere una pluralità di temi.

Sebbene il punto centrale degli esempi proposti in seguito sia l'analisi concettuale e soggettazione di atti normativi si ritiene opportuno fare una riflessione anche su alcune questioni legate alla catalogazione descrittiva di questo tipo di risorse soffermandosi anche su come esse possano essere inquadrare seguendo il modello concettuale IFLA LRM (Mangione et al. 2021).

- È da considerare ‘Opera’ il testo della norma nella sua interezza. Poiché una norma è il prodotto finale dell'esercizio del potere legislativo, l'entità ‘Opera’ corrisponde al testo pubblicato tramite i mezzi ufficiali a conclusione dello stesso.
- È da ritenere ‘Espressione’ il testo della norma espresso nella lingua ufficiale dell'organo emanante. Per le norme di ambito nazionale e regionale non emergono dubbi in tal senso, non accade ugualmente nel caso della normativa internazionale, in genere promulgata in più lingue, tra le quali è consigliabile scegliere la versione in lingua inglese in assenza del testo espresso in lingua italiana in forma ufficiale.
- La ‘Manifestazione’ è rappresentata solitamente dal file in estensione PDF o dalla pagina html che contiene il testo della norma catalogata. L'accesso alla risorsa può essere fornito tramite URL o permalink.
- L'indicazione del ‘Titolo preferito’ può coincidere con la comune forma di citazione dell'atto stesso (REICAT 2009, 9.1.2.5 D) mentre la sua epigrafe può essere descritta come ‘complemento del titolo’.

- ‘Creatore’ della norma è inteso l’organo che emana l’atto. Al fine di una più chiara esposizione va precisato che sebbene le leggi italiane presentino la formula “il Presidente della Repubblica promulga” sono intestate all’organo che esso rappresenta (REICAT 2009), ovvero, il Parlamento italiano espresso nella forma ‘Italia’. Un’intestazione presentata in forma gerarchizzata è impiegata anche per enti che hanno organi subordinati che producono atti normativi differenti, pertanto, una deliberazione della Giunta della Regione Toscana è intestata nella forma ‘Toscana <Regione> : Giunta regionale’, di contro, una legge regionale promulgata dalla stessa regione avrà come creatore ‘Toscana <Regione> : Consiglio regionale’. Il modello di relazione in questo caso è LRM-R5 ‘Opera è stata creata da Agente’.



Figura 2. Esempio di formulazione di responsabilità e del titolo proprio

Il soggetto di una norma

Indicizzare per soggetto questo tipo di risorse richiede alcune considerazioni preliminari necessarie a una corretta descrizione semantica delle stesse. Seguendo la procedura prevista per ogni tipo di risorsa, le operazioni di soggettazione di una norma devono avere come punto di partenza l’analisi concettuale della stessa che, in questi casi specifici, è bene che non sia limitata solo all’individuazione del tema di base (*aboutness*) ma è necessario che essa conduca al riconoscimento degli aspetti relativi alla sua natura (*isness*) e, se il caso lo richiede, è bene analizzare anche come il suo contenuto può essere interpretato a favore delle esigenze dell’utente (*ofness*). L’esplicitazione dei tre aspetti è ritenuta necessaria per rappresentare in modo coesteso ed esaustivo tutti i concetti emersi dall’analisi concettuale della norma. Entrando nel merito delle suddette tre forme di indicizzazione, per ognuno di esse è bene prendere in considerazione i seguenti aspetti:

- *Aboutness*. Può coincidere in modo più o meno esatto con il precetto⁶ della norma stessa. Nell’individuazione del tema di base va riconosciuta la poca rilevanza del titolo, essendo composto soltanto dalla dichiarazione del tipo di atto, dalla data di emanazione e dal numero progressivo assegnatogli. In questo caso, infatti, il ruolo di maggior rilievo è svolto dal complemento del titolo, il quale, è spesso sufficiente per comprendere di cosa tratta la norma, tuttavia, è bene precisare anche che non basta soffermarsi alla lettura dell’epigrafe ma è bene esaminare il testo dell’atto nella sua interezza. Questo aspetto assume una

⁶ Il precetto è definito come il comando o il divieto che la norma impone di osservare. (Bin e Pitruzzella 2023).

maggior importanza anche per il fatto che spesso gli atti normativi possono contenere aspetti tematici o addirittura altri documenti, quali regolamenti o linee guida, che per motivi di sintesi non sono indicati nell'epigrafe ma che vanno comunque rappresentati concettualmente con il proprio soggetto. Il grado di specificità del tema di base spesso è anche correlato all'ambito dell'atto esaminato, per fare un esempio, una legge statale sulle politiche sociali per le famiglie può avere, per motivi di competenza amministrativa, un approfondimento tematico alquanto generale e astratto che richiede un soggetto di contenuto concettuale generico, al contrario, una deliberazione provinciale che si occupa di uno specifico strumento sociale o di uno specifico progetto realizzato nell'ambito delle politiche sociali per le famiglie è descritto con un soggetto contenente termini dal contenuto concettuale più specifico.

- *Isness*. In sede di catalogazione delle norme specificare nella stringa di soggetto 'ciò che l'opera è' assume una rilevanza non trascurabile giustificata principalmente dal fatto che la presenza della *isness* potrebbe essere l'unico elemento della stringa che consenta di 'identificare' quella risorsa e di selezionarla qualora soddisfi il bisogno informativo dell'utente. Nella situazione di seguito illustrata a scopo esemplificativo è possibile notare come la dichiarazione di cosa sia la risorsa possa fare la differenza a beneficio dell'utente:

Enunciato 1	<i>Quest'opera tratta dell'accoglienza dei minori stranieri non accompagnati in Italia.</i>
Soggetto 1	Minori stranieri non accompagnati - Accoglienza - Italia

Figura 3. Esempio di enunciato di soggetto e stringa privi di indicazione della *isness*

Enunciato 2	<i>Questo è un decreto ministeriale che tratta dell'accoglienza dei minori stranieri non accompagnati in Italia.</i>
Soggetto 2	Minori stranieri non accompagnati - Accoglienza - Italia - Decreti ministeriali

Figura 4. Esempio di enunciato di soggetto e stringa nei quali è rappresentata la *isness* della risorsa

Le due stringhe di soggetto appaiono identiche sotto il profilo dell'*aboutness* in quanto rappresentano lo stesso argomento. Spostando, invece, i termini del confronto sul piano della *isness* è possibile notare che esse comunicano due informazioni differenti poiché la prima stringa di soggetto rappresenta il contenuto concettuale di una risorsa che tratta del tema suddetto mentre la seconda stringa, arricchita dal termine 'Decreti ministeriali', consente di recuperare un decreto ministeriale che tratta lo stesso tema. Nella prassi catalografica la rappresentazione della *isness* nella stringa di soggetto deve essere affidata ai termini del Thesaurus che possono assumere il ruolo di forma bibliografica, per esempio, 'Decreti ministeriali', 'Risoluzioni', 'Deliberazioni'. Nel caso in cui l'agenzia catalografica decida di descrivere atti analoghi ma prodotti da creatori diversi è possibile potenziare la precisione di questa forma d'indicizzazione completando la stringa con l'indicazione dell'ente emanante espresso in forma non gerarchizzata come mostrato nella successiva figura 6.

- *Ofness*. L'ultima forma d'indicizzazione possibile nei confronti di una norma è quella mirata a far emergere 'qual è il significato del suo contenuto'. Prima di chiarire in che modo può essere espressa la *ofness* è bene soffermarsi un po' su questa forma di descri-

zione semantica. Essa è percepita come un'estensione dell'*aboutness*⁷ e la sua esistenza ha lo scopo di consentire la rappresentazione di alcuni aspetti della risorsa che altrimenti non sarebbe possibile esprimere con la 'circularità' e che possono avere ugualmente un potenziale interesse per gli utenti. Il suo impiego è previsto prevalentemente per indicizzare risorse iconografiche e, storicamente, trova la sua origine nel momento in cui la *Library and information science* (LIS) si è confrontata con il problema di come rappresentare nel modo corretto il contenuto semantico di 'ciò di cui tratta l'opera' e di 'ciò che l'opera rappresenta' (Shatford 1986, 39-62). Mutuando le teorie di iconologia dello storico dell'arte Erwin Panofsky (Panofsky 1999), la biblioteconomia è giunta a teorizzare che l'analisi concettuale di una risorsa iconografica può essere condotta nei tre seguenti livelli (BNCF 2021, 32):

1. Descrizione generica delle entità raffigurate: animali, persone, luoghi.
2. Identificazione specifica delle entità e azioni raffigurate: un noto personaggio, un luogo identificato, un evento storico conosciuto.
3. Interpretazione del significato dell'opera: l'opera rappresenta un tema, è allegoria di un concetto, è importante in questo contesto.

Rivolgendo di nuovo l'attenzione alle norme è possibile analizzare concettualmente il loro testo sotto la lente del terzo livello di *ofness* al fine di indicare all'utente cosa essa rappresenti o, in altri termini, come può essere interpretato il suo testo. È bene precisare sin da subito che questo tipo di analisi concettuale delle norme rappresenta un terreno su cui è necessario muoversi in modo molto cauto, principalmente per il fatto che individuare questo tipo di contenuto semantico può condurre il catalogatore in un'operazione che vada oltre le proprie competenze, trascinandolo nel ciclone dell'interpretazione della legge.⁸ Sul tema la giurisprudenza italiana ha dedicato molta della sua attenzione, a dimostrazione della delicatezza e importanza dell'argomento, pertanto, in questa sede, l'unico riferimento proposto è quello contenuto nel Decreto con cui è stato approvato il testo del Codice civile italiano che così si esprime nell'articolo 12 "nell'applicare la legge non si può ad essa attribuire altro senso che quello fatto palese dal significato proprio delle parole secondo la connessione di esse, e dalla intenzione del legislatore".⁹ Alla luce di quanto detto appare evidente che il compito del catalogatore nei confronti delle norme deve restare asettico e neutrale, tuttavia, solo in alcuni casi particolari e soltanto se ciò comporta un beneficio agli utenti è possibile creare un ulteriore accesso semantico di questo tipo. L'utilità di questo tipo di orientamento descrittivo risiede nel fatto che spesso il legislatore emana una norma su un dato tema che può avere effetti o riferimenti anche per temi che pur essendo diversi dal tema di base potrebbero essere potenziali chiavi di ricerca per l'utente. Un esempio è rappresentato dal seguente caso:

⁷ <https://www.isko.org/cyclo/subject#refS>

⁸ Questo tema presenta una notevole produzione scientifica, in questa sede si fa riferimento a due recenti contributi (Pino 2021) (Astone 2020) e si rimanda alla bibliografia pregressa.

⁹ <https://www.gazzettaufficiale.it/eli/id/1942/04/04/042U0262/sg>

Enunciato	La norma tratta dell'erogazione di borse di studio agli studenti da parte della Regione Siciliana.
Soggetto 1	Borse di studio - Erogazione [agli] studenti [da parte della] Sicilia <Regione>
Soggetto 2	Studenti - Diritto allo studio - Tutela - Sicilia

Figura 5. Esempio di stringa di soggetto (Soggetto 2) costruita sulla ofness della risorsa

La norma esaminata ha come tema di base i concetti espressi dall'enunciato indicato in alto e dal quale è stata costruita la stringa di soggetto 1 quale principale accesso semantico, tuttavia, estendendo l'analisi concettuale all'ambito in cui si colloca la norma è stato possibile riconoscere, 'interpretandone' il significato, che essa potrebbe avere rilevanza semantica anche nei confronti del tema della tutela al diritto allo studio, pertanto, al solo scopo di ampliare i punti di accesso e recupero alla norma, è stata costruita anche la seconda stringa di soggetto. La scelta trova fondamento scientifico nel principio di predittività (BNCF 2021, 29) grazie al quale, in questo specifico caso, è molto alta la possibilità che l'utente provi a cercare la norma partendo dal secondo soggetto e non dal primo.

Ruoli logici nel soggetto dell'atto normativo

L'analisi concettuale di una norma comporta un'attenzione maggiore anche nell'individuazione dei ruoli logici. Spesso l'enunciato di soggetto risulta molto complesso perché, come già ricordato, più che trattare un tema, la norma contiene regole necessarie a gestire situazioni anche complesse e che possono coinvolgere una molteplicità di entità. Seguendo lo schema dei ruoli proposto nella Guida al Nuovo Soggettario (BNCF 2021), di seguito sono illustrati i ruoli logici identificabili con maggior frequenza nell'analisi concettuale di un campione di atti normativi.

- **Concetto chiave.** È in genere il precetto della norma, ovvero, la regola o il divieto che impone di osservare e che può assumere il ruolo di azione senza entità verso cui è diretta, per esempio, 'Circolazione stradale'. A seconda dei casi questo ruolo è svolto anche da entità diverse. Per esempio, norme con contenuto relativo a strumenti di previdenza sociale possono presentare stringhe il cui concetto chiave è rappresentato dal termine che identifica lo stesso strumento sociale come nel caso di '**Assegni sociali** - Erogazione [alle] Famiglie' oppure '**Sussidi** - Erogazione [alle] Lavoratrici madri'. Norme che tutelano diritti o concetti ad essi connessi possono avere stringhe in cui il concetto chiave è rappresentato dalla persona che detiene il diritto come nel caso del soggetto '**Lavoratori** - Diritto di sciopero - Tutela'.
- **Elemento transitivo.** In linea di massima è rappresentato dal termine che rappresenta il concetto di azione transitiva compiuta o fatta compiere dal legislatore per mezzo dell'atto. È frequente il caso di stringhe che vanno costruite secondo lo schema 'azione su azione', ciò si verifica quando l'azione o l'intervento proposto dalla norma è finalizzato a correggere o ridurre un comportamento che nella stringa è a sua volta rappresentato da un altro concetto di azione, per esempio, nel caso di un soggetto di questo tipo '**Bambini** - Violenza sessuale - **Prevenzione**'.

- **Beneficiario.** Ruolo svolto dal termine rappresentante i destinatari dell'intervento normato dall'atto e che rappresenta generalmente persone o enti. Per esempio, 'Finanziamenti - Erogazione [ai] **Comuni**' oppure 'Borse di studio - Erogazione [agli] **Studenti**'.
- **Agente.** Ruolo assegnato al termine che rappresentano il concetto di entità che compie l'azione normata e, nei casi analizzati, indica un'entità di diritto pubblico. Riproponendo l'esempio precedente, la stringa potrebbe essere costruita in questo modo 'Borse di studio - Erogazione [agli] **Studenti** [da parte della] **Liguria <Regione>** - Deliberazioni'.
- I **Ruoli complementari 'Luogo'** e **'Tempo'** possono essere impiegati solo quando strettamente necessario. Nel primo caso, l'indicazione del luogo dovrebbe essere impiegata soltanto quando non è usato il nome dell'ente che emana il provvedimento in quanto, nel caso contrario, il concetto di luogo coinciderebbe con la porzione di territorio politico su cui essa produce i suoi effetti giuridici. Per esempio, nella stringa 'Immigrati - Accoglienza - **Sicilia**' è necessario l'impiego del termine per rappresentare il concetto di luogo, ma nel caso della stringa 'Funzionari pubblici - Formazione - Finanziamenti [da parte del] Ministero dell'interno [di] Italia' la presenza del nome proprio dell'ente che finanzia la formazione dei funzionari pubblici rende superflua l'indicazione del luogo, posto che la norma produce effetti solo in Italia. Considerazioni analoghe possono essere fatte anche nei confronti dei termini relativi al ruolo di tempo e poiché la norma ha effetto giuridico fino alla sua abrogazione, l'indicazione del tempo dovrebbe essere impiegata soltanto per quei casi i cui effetti giuridici sono espressamente limitati nel testo dell'atto. Per esempio la stringa 'Famiglie - Tutela - Ruolo [dell'] Osservatorio nazionale sulla famiglia - Italia - **2020-2023** - Decreti ministeriali' rappresenta il contenuto concettuale di un decreto ministeriale che tratta del ruolo dell'Osservatorio nazionale sulla famiglia in Italia, per gli anni rappresentati nella stringa, in caso contrario il ruolo di tempo non andrebbe indicato.
- Il ruolo di **Forma bibliografica** può essere impiegato per rappresentare nella stringa la *isness* della norma catalogata. Nei casi di medesimi atti prodotti da organi differenti, la stringa può essere completata dal nome dell'ente emanante. Questo ulteriore elemento, del tutto opzionale, consentirebbe nello scorrimento della lista dei soggetti non soltanto la rapida e puntuale distinzione della tipologia di atto indicizzato ma anche l'identificazione e la selezione dell'organo che lo ha prodotto, come mostrato nella figura seguente.

Diritto allo studio - Raccomandazioni	Diritto allo studio - Raccomandazioni [delle] Nazioni Unite
Diritto allo studio - Raccomandazioni	Diritto allo studio - Raccomandazioni [del] Consiglio d'Europa
Diritto allo studio - Raccomandazioni	Diritto allo studio - Raccomandazioni [dell'] Unione europea
Mantenimento della pace - Risoluzioni	Mantenimento della pace - Risoluzioni [del] Consiglio d'Europa
Mantenimento della pace - Risoluzioni	Mantenimento della pace - Risoluzioni [dell'] Unione europea

Figura 6. Esempio di espansione del ruolo di Forma bibliografica

Esempi esplicativi di indicizzazione per soggetto di atti normativi

Nelle pagine seguenti è proposta una selezione di dieci atti normativi analizzati sotto la lente della catalogazione semantica a supporto e dimostrazione di quanto finora esposto. Al fine di agevolare la lettura sono di seguito fornite indicazioni sulle modalità di analisi e sulle scelte grafiche fatte nel proporre l'analisi di ogni singolo atto normativo.

- Gli atti sono presentati secondo il criterio di ambito crescente, per cui, sono proposte nel seguente ordine: norme regionali e delle Province autonome di Bolzano e Trento, norme nazionali, atti dell'Unione europea, atti emanati da organi di ambito internazionale.
- L'analisi degli atti mostra l'enunciato di soggetto seguito dalla illustrazione dettagliata di tutti gli elementi della stringa di soggetto. I termini sono tutti tratti dal Thesaurus del Nuovo Soggettario, i cui ruoli sono indicati tra parentesi tonde seguendo lo schema dei ruoli presente nella seconda edizione della Guida. In fondo è sempre presente il paragrafo 'Osservazioni' in cui sono illustrate le scelte fatte in sede di analisi concettuale e alcune proposte di integrazioni al Thesaurus.
- Tutti i link presenti nelle pagine hanno come ultima data di consultazione: 25/10/2023.

Esempio 1

Sicilia <Regione> : Assemblea Regionale.

L.R. 19 novembre 2021, n. 27 : Interventi per la prevenzione ed il contrasto ai fenomeni del bullismo e del cyberbullismo sul territorio della Regione.

Link alla norma

<https://www.inapp.gov.it/strumenti/normativa/norme-regionali/sicilia-legge-regionale-19-novembre-2021-n-27>

Soggetto

Bullismo [e] Cyberbullismo [e] Pedopornografia - Prevenzione [e] Riduzione - Progetti - Finanziamenti [della] Sicilia <Regione> - Legislazione regionale

Analisi concettuale

Enunciato di soggetto

Questa è una legge regionale (*isness*) che tratta dei finanziamenti della Regione Siciliana di progetti per la prevenzione e la riduzione del bullismo e del cyberbullismo e della pedopornografia (*aboutness*).

Stringa di soggetto

(concetto chiave) **Bullismo** [e] **Cyberbullismo** [e] **Pedopornografia** - (elemento transitivo) **Prevenzione** [e] **Riduzione** - (strumento) **Progetti** - (elemento transitivo) **Finanziamenti** [della] (agente) **Sicilia <Regione>** - (forma bibliografica) **Legislazione regionale**

Osservazioni

La stringa presenta tre termini con il ruolo di concetto chiave, citati in forma coordinata, in quanto fra di essi non intercorre alcuna relazione transitiva o di appartenenza. Rispetto ai primi due, la presenza del termine ‘Pedopornografia’ è giustificata dal fatto che esso è il termine che più si avvicina al valore semantico del concetto di ‘cyberpedofilia’, tema che seppur non presente nell’epigrafe della legge è attestato nell’articolo 1 che illustra in modo più completo il contenuto della stessa. Anche il ruolo di elemento transitivo è rappresentato da due termini ‘Prevenzione’ e ‘Riduzione’, necessari a rappresentare i concetti delle due azioni che il precetto della legge impone, ovvero, “la prevenzione ed il contrasto”. La struttura della stringa è completata dal termine ‘Progetti’ con ruolo di strumento collegato a sua volta al termine ‘Finanziamenti’ che detiene il ruolo di secondo elemento transitivo e che rappresenta anche l’azione della quale il termine ‘Sicilia <Regione>’ ne è agente. La rappresentazione della *isness*, data dal ruolo di forma bibliografica, potrebbe essere affidata al termine ‘Legislazione regionale’, per il quale si propone un’integrazione della nota sintattica nel Thesaurus che ne indichi la possibilità di impiego anche per rappresentare questo ruolo complementare. Nel complesso, è possibile notare come tutta la stringa sia costruita seguendo il principio della dipendenza dall’oggetto/meta secondo cui i termini che rappresentano la meta vanno citati prima dell’azione, mentre i termini che rappresentano lo strumento e l’agente vanno citati dopo (BNCF 2021, 189).

Esempio 2

Trento <Provincia autonoma> : Consiglio provinciale.

L.P. 12 febbraio 2020, n. 1 : Interventi a sostegno dei coniugi separati o divorziati in difficoltà.

Link alla norma

https://www.consiglio.provincia.tn.it/layouts/15/dispatcher/doc_dispatcher.aspx?app=clex&at_id=34525

Soggetto

Sussidi - Erogazione [ai] *Genitori divorziati* [e ai] *Genitori separati* [da parte di] Trento <Provincia autonoma> - Legislazione provinciale

Analisi concettuale

Enunciato di soggetto

Questa è una legge provinciale (*isness*) che tratta dell’assegnazione di sussidi economici ai genitori divorziati e ai genitori separati da parte della Provincia Autonoma di Trento (*aboutness*).

Stringa di soggetto

(concetto chiave) **Sussidi** - (elemento transitivo) **Erogazione** [ai] (beneficiario) **Genitori divorziati** [e ai] **Genitori separati** [da parte di] (agente) **Trento <Provincia autonoma>** - (forma bibliografica) **Legislazione provinciale**

Osservazioni

La stringa è costruita attorno all'elemento transitivo 'Erogazione' il quale presenta un tipo di relazione a tre argomenti (AIB e GRIS 1997, par. 2.2.3) che lo vede contemporaneamente legato al concetto chiave 'Sussidi', al ruolo di agente rappresentato dal termine 'Trento <Provincia autonoma>' e ai termini 'Genitori divorziati' e 'Genitori separati' che assumono il ruolo di beneficiari dell'azione. Per questo tipo di casi l'ordine di citazione dei termini prevede che al concetto chiave faccia subito seguito l'elemento transitivo e poi il beneficiario seguito, a sua volta, dall'agente, introdotto dall'elemento connettivo [da parte di]. Il termine 'Sussidi' è impiegato per rappresentare concettualmente il contenuto del sintagma 'interventi a sostegno' presente nell'epigrafe ma dal significato troppo ampio e soggetto ad ambiguità, pertanto, al fine di individuare con esattezza il concetto da esprimere nella stringa, è necessario leggere anche il testo della norma in cui è indicato, nell'articolo 2, che il tipo di benefici erogati sono dei sussidi economici. Dunque, la stringa propone il termine 'Sussidi' per il ruolo di concetto chiave mentre il ruolo di elemento transitivo è rappresentato dal termine 'Erogazione'. Scegliendo tra i termini del NS il ruolo del primo beneficiario sarebbe al momento rappresentabile dal termine 'Divorziati' mentre il concetto di genitori separati non è rappresentato con un termine specifico nel NS. La necessità di rappresentare in modo preciso tali concetti rende gli elementi 'Genitori divorziati'¹⁰ e 'Genitori separati'¹¹ termini che potrebbero essere proposti per l'introduzione nel Thesaurus. Infine, si potrebbe rappresentare la (*isness*) di 'leggi provinciali' con un nuovo termine 'Legislazione provinciale'¹² che potrebbe essere introdotto nel Thesaurus con una necessaria nota d'ambito che ne indichi l'attribuzione esclusiva alle province autonome di Bolzano e di Trento.

Esempio 3

Italia

L. 20 gennaio 2016, n. 12 : Disposizioni per favorire l'integrazione sociale dei minori stranieri residenti in Italia mediante l'ammissione nelle società sportive appartenenti alle federazioni nazionali, alle discipline associate o agli enti di promozione sportiva.

Link alla norma

<https://www.gazzettaufficiale.it/eli/id/2016/02/01/16G00016/sg>

Soggetti

1. Minori stranieri - Accoglienza [da parte delle] Società sportive - Italia - Legislazione
2. Minori stranieri - Integrazione sociale - Impiego [dello] Sport - Italia - Legislazione

¹⁰ L'ipotesi di collocazione nel Thesaurus del termine proposto è la seguente: Genitori divorziati. Macrocategoria: Agenti; Categoria: Persone e gruppi; TT Persone; BT Genitori.

¹¹ L'ipotesi di collocazione nel Thesaurus del termine proposto è la seguente: Genitori separati. Macrocategoria: Agenti; Categoria: Persone e gruppi; TT Persone; BT Genitori.

¹² L'ipotesi di collocazione nel Thesaurus del termine proposto è la seguente: Legislazione provinciale. Macrocategoria: Cose; Categoria: Strumenti; TT Strumenti; BT Legislazione.

Analisi concettuale

Enunciato di soggetto 1

Questa è una legge (*isness*) che tratta dell'ammissione alle società sportive dei minori stranieri residenti in Italia (*aboutness*).

Stringa di soggetto 1

(concetto chiave) **Minori stranieri** - (elemento transitivo) **Accoglienza** [da parte delle] (agente) **Società sportive** - (luogo) **Italia** - (forma bibliografica) **Legislazione**

Osservazioni

L'analisi concettuale dell'epigrafe, in questo caso, non consente una chiara identificazione di quale sia l'*aboutness* e quale l'*ofness*, pertanto, è necessario leggere per intero il testo dell'atto dal quale è possibile estrapolare il primo enunciato di soggetto. Analizzando i singoli elementi si nota che al termine 'Minori stranieri' è assegnato il ruolo di concetto chiave; l'azione di "ammissione" è rappresentato dal termine 'Accoglienza' mentre il concetto "società sportive" è rappresentato dal termine 'Società sportive' poiché il termine 'Associazioni sportive' è termine non preferito nel NS.¹³ Tra i ruoli complementari vi sono il luogo espresso dal termine 'Italia' mentre il compito di rappresentare la *isness* potrebbe essere affidato al termine 'Legislazione' per il quale potrebbe essere avanzata una proposta d'integrazione della nota sintattica del Thesaurus che dichiarerebbe il suo potenziale impiego anche nel ruolo di forma bibliografica.¹⁴

Enunciato di soggetto 2

Questa è una legge (*isness*) che può essere interpretata come una norma che tratta il tema dell'integrazione sociale dei minori stranieri residenti in Italia mediante le attività sportive (*ofness*).

Stringa di soggetto 2

(concetto chiave) **Minori stranieri** - (elemento transitivo) **Integrazione sociale** - (elemento transitivo) **Impiego** [dello] (strumento) **Sport** - (luogo) **Italia** - (forma bibliografica) **Legislazione**

Osservazioni

Questa stringa è costruita con lo scopo di descrivere cosa contiene concettualmente la legge interpretandone il suo contenuto nell'ottica di fornire un accesso semantico anche alla *ofness* dell'atto. Il risultato dell'analisi concettuale si presenta, in questo caso, come un'espansione dell'*aboutness* che permette di costruire una stringa la cui esistenza è giustificata dal fatto che essa rispetta il principio della predittività in quanto, seppur meno pertinente, ha maggiori possibilità di essere impiegata come punto di accesso dall'utente medio.

¹³ <https://thes.bncf.firenze.sbn.it/termine.php?id=17032>

¹⁴ L'indicazione proposta in Nota sintattica potrebbe essere fornita in analogia a quella del termine 'Fotografie', pertanto, dopo le indicazioni riguardo al suo impiego nel ruolo di ELEM. TRANS. e di PARTE/PROPR., potrebbero seguire l'indicazione "FORMA BIBL. Segue, se presenti, i termini che rappresentano il luogo, il tempo, la forma intellettuale o il caso, es. Educazione civica - Insegnamento - Scuole - Italia - Legislazione".

Esempio 4

Italia

L. 29 luglio 2020, n. 77 : Istituzione di una Commissione Parlamentare di inchiesta sulle attività connesse alle comunità di tipo familiare che accolgono minori. Disposizioni in materia di diritto del minore ad una famiglia.

Link alla norma

<https://www.gazzettaufficiale.it/eli/id/2020/08/28/20G00126/sg>

Soggetti

1. Commissione parlamentare di inchiesta sulle attività connesse alle comunità di tipo familiare che accolgono minori - Costituzione - Italia - Legislazione
2. Case alloggio - Controllo [da parte della] Commissione parlamentare di inchiesta sulle attività connesse alle comunità di tipo familiare che accolgono minori - Italia - Legislazione
3. Diritti dell'infanzia e dell'adolescenza - Italia - Legislazione

Analisi concettuale

Enunciato di soggetto 1

Questa è una legge (*isness*) tratta dell'istituzione in Italia della Commissione parlamentare di inchiesta sulle attività connesse alle comunità di tipo familiare che accolgono minori (*aboutness*).

Stringa di soggetto 1

(concetto chiave) **Commissione parlamentare di inchiesta sulle attività connesse alle comunità di tipo familiare che accolgono minori** - (elemento transitivo) **Costituzione** - (luogo) **Italia** - (forma bibliografica) **Legislazione**

Osservazioni

Il ruolo di concetto chiave è assegnato al sintagma 'Commissione parlamentare di inchiesta sulle attività connesse alle comunità di tipo familiare che accolgono minori' il quale è anche il nome proprio con cui è identificata la commissione in oggetto. Nonostante la presenza nel Thesaurus del termine 'Commissioni parlamentari', in questo caso specifico, è stato scelto il nome proprio della Commissione al fine di evitare una perdita di precisione e rilevanza nel richiamo delle risorse che trattano dello stesso organo. Il ruolo dell'elemento transitivo non è dato al termine 'Istituzione' poiché identificato come non preferito a favore del termine preferito 'Costituzione'.¹⁵

Enunciato di soggetto 2

Questa è una legge (*isness*) interpretabile come (*ofness*) una risorsa che tratta del controllo in Italia delle comunità familiari da parte della Commissione parlamentare di inchiesta sulle attività connesse alle comunità di tipo familiare che accolgono minori.

¹⁵ <https://thes.bncf.firenze.sbn.it/termine.php?id=58147>

Stringa di soggetto 2

(concetto chiave) **Case alloggio** - (elemento transitivo) **Controllo** [da parte della] (agente) **Commissione parlamentare di inchiesta sulle attività connesse alle comunità di tipo familiare che accolgono minori** - (luogo) **Italia** - (forma bibliografica) **Legislazione**

Osservazioni

Il secondo soggetto è costruito come estensione del valore semantico dell'*aboutness* e la scelta si è resa necessaria poiché la descrizione fornita con il primo soggetto risulta per l'utenza alquanto tautologica perché non dice nulla di più riguardo al concetto chiave. Leggendo il testo della norma appare evidente che l'istituzione della Commissione non sia fine a se stessa, pertanto, è possibile identificare questo ulteriore soggetto che ha come tema di base l'espansione dei concetti presenti in modo celato nell'*aboutness*, ovvero, ciò di cui si occupa la commissione. La scelta è giustificata dal fatto che vi sono maggiori probabilità che l'utente effettui una ricerca di risorse sul 'controllo delle case alloggio in Italia' piuttosto che usare come punto di accesso il nome della Commissione che potrebbe anche ignorare. Analizzando i singoli termini, la stringa rappresenta il concetto di 'comunità di tipo familiare' mediante il termine 'Comunità alloggio' poiché l'alternativo termine 'Comunità familiari' è identificato come non preferito. Eventuali dubbi sulla scelta di tale termine sono chiariti dalla seguente nota d'ambito "Strutture di tipo assistenziale che ospitano, in piccole comunità, persone che necessitano di accoglienza, assistenza, ecc. (come bambini in attesa di affidamento o adozione, ragazze madri, malati di mente, disabili), e gestiscono le proprie attività attraverso assistenti sociali, volontari, ecc., spesso con l'appoggio di strutture pubbliche".¹⁶ L'elemento di azione nella stringa è rappresentata dal termine 'Controllo' cui segue l'agente dato dal nome della Commissione. Chiude la stringa il ruolo di luogo e quello di forma bibliografica assegnato al termine 'Legislazione' secondo la proposta di integrazione avanzata negli esempi precedenti.

Enunciato di soggetto 3

Questa è una legge che (*isness*) tratta del diritto alla famiglia dei minori in Italia (*aboutness*).

Stringa di soggetto 3

(concetto chiave) **Diritti dell'infanzia e dell'adolescenza** - (luogo) **Italia** - (forma bibliografica) **Legislazione**

Osservazioni

Il contenuto della legge presenta due argomenti distinti dei quali il primo tratta della Commissione suddetta e rappresentato dal primo soggetto mentre il secondo è dedotto dalla dicitura in epigrafe "Disposizioni in materia di diritto del minore ad una famiglia" rappresentato, in questo caso, dalla terza stringa. Nel NS il termine che si avvicina maggiormente al concetto di diritto alla famiglia è 'Diritti dell'infanzia e dell'adolescenza' il quale è stato scelto anche se presenta un contenuto concettuale più generico e ampio, riferendosi ai diritti dei minori nella sua complessità. Chiudono la stringa l'elemento che rappresenta il luogo e quello relativo alla forma bibliografica indicata dal termine 'Legislazione'.

¹⁶ <https://thes.bncf.firenze.sbn.it/termine.php?id=49969>

Esempio 5

Italia : Ministero dell'Istruzione.

D.M. 5 gennaio 2021, n. 6 : Individuazione degli uffici di livello dirigenziale non generale dell'amministrazione centrale del Ministero dell'istruzione.

Link alla norma

<https://www.gazzettaufficiale.it/eli/id/2021/02/26/21A01183/sg>

Soggetto

Italia : Ministero dell'Istruzione - Uffici - Organizzazione - Decreti ministeriali

Analisi concettuale

Enunciato di soggetto

Questo è un Decreto ministeriale (*isness*) che tratta dell'organizzazione degli uffici del Ministero dell'Istruzione italiano (*aboutness*).

Stringa di soggetto

(concetto chiave) **Italia : Ministero dell'Istruzione** - (parte/proprietà) **Uffici** - (elemento transitivo) **Organizzazione** - (forma bibliografica) **Decreti ministeriali**

Osservazioni

L'analisi concettuale della risorsa richiede un approfondimento nella lettura del suo testo al fine di individuarne l'*aboutness*, pertanto, il concetto dell'azione 'Organizzazione' è identificabile leggendo il testo degli articoli 1 e 2. Inoltre il termine proposto è ritenuto più pertinente rispetto al termine più ambiguo "Individuazione degli uffici" presente in epigrafe. Il ruolo di concetto chiave è attribuito al nome del ministero in oggetto, citato in stringa in forma gerarchizzata come previsto dal Manuale applicativo al NS (BNCF 2023, par. A.6.2). Al concetto chiave fa seguito il termine 'Uffici' il quale pur essendo il reale oggetto dell'azione nella stringa è citato a seguire con il ruolo di parte/proprietà, secondo il principio della dipendenza dal possessore (BNCF 2021, 192). L'azione compiuta è rappresentata dal termine 'Organizzazione' tratto dalla categoria 'Attività'. Tra i ruoli complementari, in questo caso, non è stato necessario porre nella stringa il termine rappresentante il concetto di luogo in quanto è implicitamente rappresentato nel concetto chiave mentre è rappresentata la *isness* della risorsa per mezzo del termine 'Decreti ministeriali' con ruolo di forma bibliografica.

Esempio 6

Consiglio dell'Unione europea

Conclusione 8 marzo 2021 n. 6795/21 : Conclusione del Consiglio sul rafforzamento dell'applicazione della Carta dei diritti fondamentali.

Link alla norma

<https://data.consilium.europa.eu/doc/document/ST-6795-2021-INIT/it/pdf>

Soggetti

1. Convenzione europea dei diritti dell'uomo <1950> - Applicazione - Paesi dell'Unione europea - Atti atipici [del] Consiglio dell'Unione europea
2. Diritti umani - Tutela - Paesi dell'Unione europea - Atti atipici [del] Consiglio dell'Unione europea

Analisi concettuale

Enunciato di soggetto 1

Questa è una conclusione del Consiglio dell'Unione europea (*isness*) che tratta dell'applicazione della Convenzione europea dei diritti dell'uomo nei Paesi dell'Unione europea (*aboutness*).

Stringa di soggetto 1

(concetto chiave) **Convenzione europea dei diritti dell'uomo <1950>** - (elemento transitivo) **Applicazione** - (luogo) **Paesi dell'Unione europea** - (forma bibliografica) **Atti atipici [del] Consiglio dell'Unione europea**

Osservazioni

Il ruolo di concetto chiave è assegnato al titolo del documento citato nella norma e presentato secondo la forma prevista dalle regole italiane (REICAT 2009) mentre il concetto di azione è rappresentato dal termine 'Applicazione', tratto dalla gerarchia introdotta dall'etichetta di nodo [Attività di applicabilità generale nell'ambito giuridico e amministrativo].¹⁷ Tra i ruoli complementari sono presenti quello di luogo rappresentato dal termine 'Paesi dell'Unione europea' e quello di forma bibliografica che esprime la *isness*, in questo caso, per via del termine 'Atti atipici'. La scelta del tale termine per rappresentare il concetto di 'Conclusioni' è consentita dalla nota d'ambito "Atti, molto vari fra loro, emessi o adottati dalle istituzioni dell'Unione europea, non sempre previsti dai Trattati e generalmente nati dalla prassi"¹⁸. Infine è bene evidenziare come in questo esempio, così come nei successivi proposti, al fine aumentare la precisione della stringa si è scelto di completare l'elemento di forma bibliografia aggiungendo il nome dell'organo che ha prodotto l'atto nella forma 'Atti atipici [del] Consiglio dell'Unione europea'. Questo espediente, del tutto opzionale, consente durante la ricerca per scorrimento della lista delle stringhe di soggetto di individuare, oltre alla forma bibliografica, anche l'ente emanante, posto a seguito del termine 'Atti atipici' secondo il principio della dipendenza dall'oggetto/meta (BNCF 2021, 189).

Enunciato di soggetto 2

Questa è una conclusione del Consiglio dell'Unione europea (*isness*) che tratta della tutela dei diritti umani nei Paesi dell'Unione europea (*aboutness*).

¹⁷ <https://thes.bncf.firenze.sbn.it/termine.php?id=15137>

¹⁸ <https://thes.bncf.firenze.sbn.it/termine.php?id=64775>

Stringa di soggetto 2

(concetto chiave) **Diritti umani** - (elemento transitivo) **Tutela** - (luogo) **Paesi dell'Unione europea**
- (forma bibliografica) **Atti atipici** [del] **Consiglio dell'Unione europea**

Osservazioni

Questo soggetto è creato allo scopo di fornire un ulteriore punto di accesso semantico alla risorsa, favorendone il recupero anche tramite i termini 'Diritti umani' e 'Tutela' i quali pur avendo un valore semantico più generico hanno una maggiore probabilità d'impiego nelle ricerche da parte degli utenti. In questo caso la forma d'indicizzazione impiegata è basata sulla *ofness* della risorsa, pertanto, interpretandone il suo significato si può dedurre che l'applicazione della Convenzione europea dei diritti dell'uomo è semanticamente correlato al concetto più generico di 'tutela dei diritti umani' di cui la Convenzione può essere inteso come strumento di tutela. Alla luce di quanto esposto, il ruolo di concetto chiave è assegnato al termine 'Diritti umani' seguito dall'elemento transitivo 'Tutela'. Conclude la stringa il termine 'Paesi dell'Unione europea' con il ruolo di luogo e il termine 'Atti atipici' impiegato per rappresentare la *isness* della risorsa.

Esempio 7

Parlamento europeo.

Risoluzione P9_TA(2020)0031 del 12 febbraio 2020 : Risoluzione del Parlamento europeo del 12 febbraio 2020 su una strategia dell'UE per porre fine alle mutilazioni genitali femminili nel mondo.

Link alla norma

https://www.europarl.europa.eu/doceo/document/TA-9-2020-0031_IT.html

Soggetto

Mutilazioni genitali femminili - Prevenzione [e] Riduzione - Risoluzioni [del] Parlamento europeo

Analisi concettuale

Enunciato di soggetto

Questa è una risoluzione del Parlamento europeo (*isness*) che tratta della prevenzione e riduzione delle mutilazioni genitali femminili nel mondo (*aboutness*).

Stringa di soggetto

(concetto chiave) **Mutilazioni genitali femminili** - (elemento transitivo) **Prevenzione** [e] **Riduzione** - (forma bibliografica) **Risoluzioni** [del] **Parlamento europeo**

Osservazioni

Il ruolo di concetto chiave è assegnato al termine 'Mutilazioni genitali femminili'¹⁹ cui fanno segui-

¹⁹ <https://thes.bncf.firenze.sbn.it/termine.php?id=7433>

to i due termini ‘Prevenzione’ e ‘Riduzione’, entrambi con ruolo composto di elemento transitivo e impiegati per rappresentare il contenuto concettuale dell’espressione presente in epigrafe “porre fine alle mutilazioni genitali femminili nel mondo”. Il ruolo di luogo deducibile dall’espressione “nel mondo” non è rappresentabile poiché non identifica un’area territoriale definita. Conclude la stringa il termine ‘Risoluzioni’ definito nel NS “Decisioni finali prese da un congresso, un’assemblea, un organo collegiale”²⁰ e scelto per rappresentare la *isness* dell’atto analizzato.

Esempio 8

Council of Europe : Committee of Ministers

Recommendation CM/Rec(2018)5 : Recommendation on 14 March 2018 of the Committee of Ministers to member States concerning children with imprisoned parents.

Link alla norma

https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016807b3175

Soggetto

Figli - Rapporti [con i] *Detenuti genitori* - Tutela - Europa - Raccomandazioni internazionali [del] Comitato dei ministri [del] Consiglio d’Europa

Analisi concettuale

Enunciato di soggetto

Questa è una raccomandazione del Comitato dei Ministri del Consiglio d’Europa (*isness*) che tratta della tutela, in Europa, dei rapporti dei figli minorenni con i genitori detenuti (*aboutness*).

Stringa di soggetto

(concetto chiave) **Figli** - (elemento transitivo) **Rapporti** [con i] (agente) *Detenuti genitori* - (elemento transitivo) **Tutela** - (luogo) **Europa** - (forma bibliografica) **Raccomandazioni internazionali** [del] **Comitato dei ministri** [del] **Consiglio d’Europa**

Osservazioni

Oltre a quanto espresso in epigrafe, il tema di base è identificabile in modo completo nell’articolo 1 nel seguente preambolo “it is necessary to protect the child’s right to, and need for, an emotional and continuing relationship with their imprisoned parent” e che consente di ricavare la formulazione dell’enunciato di soggetto. Il ruolo di concetto chiave è assegnato al termine ‘Figli’ impiegato per rappresentare il concetto ‘child’ e scelto al posto del termine ‘Bambini’ poiché da un punto di vista semantico presenta una relazione logica più forte con il termine ‘Detenuti genitori’. Il termine ‘Rapporti’ è il primo elemento transitivo che lega il concetto chiave all’agente. Per rappresentare nella stringa il concetto di ‘imprisoned parents’ si propone il termine ‘Detenuti genitori’ per il

²⁰ <https://thes.bncf.firenze.sbn.it/termine.php?id=20587>

quale si suggerisce l’inserimento nel Thesaurus²¹. Il concetto di azione ‘protect’ è rappresentato dal termine ‘Tutela’ anch’esso con il ruolo di elemento transitivo. Il ruolo complementare di luogo è tratto dal contenuto della risorsa e assegnato al termine ‘Europa’, infine, completa la stringa il termine ‘Raccomandazioni internazionali’ che rappresenta la *isness* della risorsa seguito dal nome proprio dell’ente produttore espresso nella forma indicata dal Manuale applicativo (BNCF 2023).

Esempio 9

United Nations : General Assembly.

Resolution A/RES/76/166 : Resolution adopted by the General Assembly on 16 December 2021 : The right to food

Link alla norma

<https://digitallibrary.un.org/record/3954949>

Soggetto

Sicurezza alimentare - Risoluzioni [dell’] Assemblea Generale [delle] Nazioni Unite

Analisi concettuale

Enunciato di soggetto

Questa è una risoluzione dell’Assemblea Generale delle Nazioni Unite (*isness*) che tratta del diritto al cibo (*aboutness*).

Stringa di soggetto

(concetto chiave) **Sicurezza alimentare** - (forma bibliografica) **Risoluzioni [dell’] Assemblea Generale [delle] Nazioni Unite**

Osservazioni

Il ruolo di concetto chiave è assegnato al termine ‘Sicurezza alimentare’ tratto dalla gerarchia ‘Diritti umani’ e dalla categoria ‘Strumenti’, scelto per rappresentare il concetto di “right to food”²². Il ruolo complementare di forma bibliografica è rappresentato dal termine ‘Risoluzioni’ mentre il nome dell’ente emanante è espresso nella forma discorsiva indicata dal Manuale applicativo.

Esempio 10

United Nations : General Assembly.

Resolution A/RES/73/327 : Resolution adopted by the General Assembly on 25 July 1919 : International Year of the elimination of child labour, 2021.

²¹ L’ipotesi di collocazione nel Thesaurus del termine proposto è la seguente: Detenuti genitori. Macrocategoria: Agenti; Categoria: Persone e gruppi; TT Persone; BT Detenuti.

²² <https://thes.bncf.firenze.sbn.it/termine.php?id=63523>

Link alla norma

<https://undocs.org/en/A/RES/73/327>

Soggetti

1. Anno internazionale per l'eliminazione del lavoro minorile, 2021 - Costituzione - Risoluzioni [dell'] Assemblea Generale [delle] Nazioni Unite
2. Minori - Lavoro - Abolizione - Risoluzioni [dell'] Assemblea Generale [delle] Nazioni Unite

Analisi concettuale

Enunciato di soggetto 1

Questa è una risoluzione dell'Assemblea Generale delle Nazioni Unite (*isness*) che tratta dell'istituzione dell'Anno internazionale per l'eliminazione del lavoro minorile nel 2021 (*aboutness*).

Stringa di soggetto 1

(concetto chiave) **Anno internazionale per l'eliminazione del lavoro minorile, 2021** - (elemento transitivo) **Costituzione** - (forma bibliografica) **Risoluzioni [dell'] Assemblea Generale [delle] Nazioni Unite**

Osservazioni

Il ruolo di concetto chiave è assegnato al nome dell'evento istituito, espresso nella lingua italiana 'Anno internazionale per l'eliminazione del lavoro minorile, 2021' la cui corretta denominazione è stata verificata dalla pagina in lingua italiana dell'Organizzazione Internazionale del lavoro.²³ Pur essendo un evento a carattere occasionale la forma del nome dell'evento non prevede, in questo caso, l'impiego delle parentesi unciniate per qualificare l'anno in quanto è parte integrante del suo nome (REICAT 2009). Il ruolo di elemento transitivo è rappresentato dal termine 'Costituzione' tratto dalla categoria 'Attività' poiché, il termine 'Istituzione' è non preferito mentre l'alternativo termine 'Creazione' è non utilizzabile in quanto il suo ambito d'uso è filosofico/religioso. Conclude la stringa l'elemento con ruolo di forma bibliografica 'Risoluzioni' con cui è espressa la *isness* della risorsa seguito dal nome dell'organo emanante espresso in forma discorsiva.

Enunciato di soggetto 2

Questa è una Risoluzione dell'Assemblea Generale delle Nazioni Unite (*isness*) che può essere interpretata come una risorsa che tratta dell'abolizione del lavoro minorile (*ofness*).

Stringa di soggetto 2

(concetto chiave) **Minori** - (parte/proprietà) **Lavoro** - (elemento transitivo) **Abolizione** - (forma bibliografica) **Risoluzioni [dell'] Assemblea Generale [delle] Nazioni Unite**

²³ https://www.ilo.org/rome/risorse-informative/comunicati-stampa/WCMS_768733/lang--it/index.htm

Osservazioni

Questo secondo soggetto è stato fornito con lo scopo di aumentare i punti di accesso semantico alla risorsa anche attraverso i termini che rappresentano il concetto generale di ‘abolizione del lavoro minorile’. La stringa, in questo caso, rappresenta non il tema di base ma la sua *ofness*, ovvero, un tema ritenuto rilevante per gli utenti del catalogo e individuato ‘interpretando’ il contenuto della Risoluzione. Il concetto chiave è rappresentato dalla combinazione dei termini ‘Minori’ e ‘Lavoro’ poiché ‘Lavoro minorile’ è indicato come termine non preferito.²⁴ Il concetto di azione è rappresentato dall’elemento transitivo ‘Abolizione’ tratto dalla gerarchia [Attività relative al cambiamento]²⁵. Anche in questo caso la rappresentazione della *isness* è affidata al termine ‘Risoluzioni’ corredato dal nome dell’ente.

Conclusioni

L’indicizzazione per soggetto può realmente essere applicata a ogni tipo di risorsa informativa e negli esempi sopra indicati è mostrato come sia applicabile agli atti di natura normativa o amministrativa e che fornire un accesso semantico al contenuto della norma possa garantire un potenziamento notevole nella ricerca e recupero di questo tipo di risorse. Pertanto, certi dei benefici derivanti dall’indicizzazione per soggetto, la sfida per chi volesse applicare questa proposta di descrizione di norme e leggi consiste soltanto nel fare le giuste riflessioni su come ottenere i migliori risultati di ricerca derivati da tale attività. Negli esempi precedenti è mostrata la forma di indicizzazione postcoordinata, preferita dalle maggiori agenzie bibliografiche italiane, ma nulla vieta la sua applicazione anche mediante l’impiego di singoli termini di Thesaurus proposti all’utenza nella forma precoordinata. La flessibilità dello studio si mostra anche nei confronti del tipo di atto da descrivere e, sebbene in questo le analisi proposte riguardino solamente atti di ambito regionale, nazionale e internazionale, le riflessioni proposte sono valide anche per atti di ambito locale quali Deliberazioni di Giunta comunale o un Ordinanze del sindaco. In conclusione l’atto normativo può realmente avere “diritto di cittadinanza” tra le risorse il cui contenuto concettuale può essere rappresentato mediante una stringa di soggetto.

²⁴ <https://thes.bncf.firenze.sbn.it/termine.php?id=34445>

²⁵ <https://thes.bncf.firenze.sbn.it/termine.php?id=28441>

Riferimenti bibliografici

AIB (Associazione italiana biblioteche) e GRIS (Gruppo di ricerca sull'indicizzazione per soggetto). 1997. *Guida all'indicizzazione per soggetto*. Roma: Associazione italiana biblioteche. <https://www.sba.unifi.it/mod-Areafiles-display-lid-308-cid-79.html>.

Astone, Francesco. 2020. *Appunti sull'interpretazione della legge*. Pisa: Pacini giuridica.

Bin, Roberto, e Giovanni Pitruzzella. 2023. *Le fonti del diritto*. Torino: Giappichelli.

BNCF (Biblioteca nazionale centrale di Firenze). 2021. *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto*. 2a edizione. Roma: Associazione italiana biblioteche; Firenze: Biblioteca nazionale centrale di Firenze. https://www.bncf.firenze.sbn.it/wp-content/uploads/2020/01/Nuovo-soggettario_Guida.pdf.

BNCF (Biblioteca nazionale centrale di Firenze). 2023. *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto. Manuale applicativo*. https://thes.bncf.firenze.sbn.it/Manuale_applicativo.pdf.

Commissione permanente per la revisione delle regole italiane di catalogazione, a c. di. 2009. *Regole italiane di catalogazione: REICAT*. Roma: ICCU, 2009. <https://www.iccu.sbn.it/export/sites/iccu/documenti/2015/REICAT-giugno2009.pdf>.

Costantino, Maria Antonella. 2011. *Costruire libri e storie con la CAA. Gli IN-book per l'intervento precoce e l'inclusione*. Trento: Erickson.

Fedeli, Sara. 2015. "La soggettazione della narrativa per bambini e ragazzi." *JLIS.it* 6 (3), 101-20. DOI: <https://doi.org/10.4403/jlis.it-11151>.

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

Guerrini, Mauro, e Carlo Bianchini. 2016. *Manuale RDA*. Milano: Editrice Bibliografica.

ICCU (Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche). 2023. *Linee guida sull'indicizzazione per soggetto e la classificazione nel Servizio bibliotecario nazionale (SBN)*. A cura del Gruppo di lavoro per la catalogazione semantica in SBN. Roma: ICCU. https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Linee_guida_sull'indicizzazione/Indicizzazione_per_soggetto.

IFLA (International Federation of Library Associations and Institutions). 2012. *International Standard Bibliographic Description. Edizione consolidata*. Raccomandata dall'ISBD Review Group, approvata dallo Standing Committee dell'IFLA Cataloguing section, edizione italiana a cura dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche. Roma: ICCU. https://www.iccu.sbn.it/export/sites/iccu/documenti/2012/ISBD_NOV2012_online.pdf.

Mangione, Dario, Cristiana Paola, Ignazio Pirronitto, Martina Ponzalli, Ada Schettini, e Francesca Sorrentino. 2021. *IFLA LRM. Library Reference Model*. Roma: Associazione italiana biblioteche.

Panofsky, Erwin. 1999. *Studi di iconologia: i temi umanistici nell'arte del Rinascimento*. Introduzione di Giovanni Previtali, traduzione di Renato Pedio. Torino: Einaudi.

Pino, Giorgio. 2021. *L'interpretazione del diritto*. Torino: Giappichelli.

Revelli, Carlo. 2004a. *Il catalogo*. Milano: Editrice Bibliografica.

Revelli, Carlo. 2004b. "La mattanza dei catalogatori. Una funzione che rischia la dequalificazione." *Biblioteche oggi* 22 (5), 7-15. <http://www.bibliotecheoggi.it/2004/20040500701.pdf>.

Shatford, Sara. 1986. "Analyzing the subject of a picture. A theoretical approach." *Cataloging & Classification Quarterly* 6 (3), 39-62.

Hachette v. Internet Archive

Fabio Mercanti^(a)

a) Sapienza università di Roma

Contact: Fabio Mercanti, fabio.mercanti@uniroma1.it

Received: 16 July 2023; Accepted: 04 October 2023; First Published: 15 January 2024

ABSTRACT

In June 2020, some major publishers – Hachette Book Group, HarperCollins, John Wiley & Sons, Penguin Random House – sued the Internet Archive for copyright infringement. The case is about the activity of the Open Library and of the National Emergency Library managed by Internet Archive. These offered digitisation and lend of printed books owned by the libraries, some of which were protected by copyright. This activity was carried out on the basis of Controlled Digital Lending, a theory not provided by US copyright law. In its March 2023 ruling, the Court agreed with the plaintiffs.

In the first part of the article, the activities of the Open Library and the National Emergency Library are briefly reconstructed and the development of the litigation is put into context.

The second part is dedicated to Controlled Digital Lending: the implementation process is defined, the position of its theorists is explored and the legal framework on which it is based is reconstructed.

The third part presents the main reasons for the prosecution and the Court's decision.

Finally, some reflections are proposed.

KEYWORDS

Hachette v. Internet Archive; Controlled digital lending; Open library; Digital lending

Il caso Hachette v. Internet Archive

ABSTRACT

Nel giugno 2020 alcuni grandi editori – Hachette Book Group, HarperCollins, John Wiley & Sons, Penguin Random House – hanno accusato Internet Archive di violazione del copyright. Alla base dell'accusa c'è l'attività della Open Library e, durante il primo periodo della pandemia da Covid-19, della National Emergency Library, entrambe gestite da Internet Archive. Queste hanno offerto in prestito digitalizzazioni di libri a stampa posseduti dalle biblioteche, alcuni dei quali protetti da copyright. Tale attività è stata svolta sulla base del Controlled Digital Lending il quale, secondo i suoi teorizzatori, è fondato sulla first sale doctrine e sul fair use, ma non è previsto dalla legge sul copyright degli Stati Uniti. Con la sentenza del marzo 2023 la Corte ha dato ragione ai querelanti.

Nella prima parte dell'articolo viene sinteticamente ricostruita l'attività della Open Library e della National Emergency Library e si contestualizza lo sviluppo del contenzioso.

La seconda parte è dedicata al Controlled Digital Lending: si definisce il processo di implementazione, si approfondisce la posizione dei suoi teorizzatori e si ricostruisce il quadro giuridico sul quale si basa.

Nella terza parte si presentano invece le principali ragioni dell'accusa e la decisione della Corte.

Infine si propongono alcuni stimoli per ampliare la riflessione a partire dal caso giudiziario trattato.

PAROLE CHIAVE

Hachette v. Internet Archive; Controlled digital lending; Open library; prestito digitale;

La Open Library e i presupposti del caso Hachette v. Internet Archive

«The idea is to build the library of Alexandria version 2»

Internet Archive (d'ora in poi in forma abbreviata IA) è un'organizzazione no-profit nata nel 1996. Il fondatore di IA è Brewster Kahle, laureato presso il Massachusetts Institute of Technology (MIT). Insieme ad altri è stato anche il fondatore del progetto WAIS – un sistema di ricerca di testi venduto ad AOL a metà anni Novanta – e poi di Alexa internet, acquistata da Amazon a fine anni Novanta (si tratta del servizio di analisi del traffico web, non dell'assistente virtuale proposto da Amazon).

IA offre diversi servizi importanti tra cui la Open Library, la quale mette a disposizione informazioni bibliografiche di un vastissimo catalogo di libri nonché la possibilità di prenderne in prestito una copia digitale; e la Wayback Machine, ovvero un archivio digitale del web in modo che le pagine web siano anche ricercabili e accessibili agli utenti. Offre inoltre servizi di digitalizzazione, archiviazione e fruizione di contenuti. In generale, il *core business* di IA è quindi l'archiviazione e l'accesso ai contenuti da parte degli utenti.

In questo articolo ci si occupa in particolare della Open Library (d'ora in poi in forma abbreviata OL), la quale è al centro del caso Hachette v. Internet Archive. OL è nata nel 2007 fondata, tra gli altri, da Aaron Swartz e dallo stesso Kahle. L'idea è quella di creare «one web page for every book ever published»¹ ricca di informazioni bibliografiche. Ben presto viene fornita anche la possibilità di accesso a libri posseduti dalle biblioteche in versione digitalizzata.

La seconda metà del primo decennio del Duemila è stato un periodo molto importante per il libro digitale, la lettura digitale, l'editoria e le biblioteche. Infatti, sono stati lanciati sul mercato il Kindle e l'iPhone e nel giro di pochi anni altri modelli di e-reader, smartphone e tablet. Sempre in quegli anni è stata rilasciata la seconda versione di ePub, il formato aperto per la creazione di e-book (Roncaglia 2010, 154-163)². Inoltre è rapidamente cresciuto un nuovo mercato di e-book commerciali³. Negli stessi anni si è sviluppata negli Stati Uniti l'attività di Overdrive, un aggregatore di contenuti per il servizio di prestito digitale bibliotecario⁴. A cavallo tra i due decenni anche in Europa sono nati servizi simili⁵. In generale, gli e-book pongono le biblioteche davanti a novità sostanziali che non riguardano solo il formato del libro, ma anche l'acquisizione e il prestito, possibili grazie ad accordi con i titolari dei diritti (Roncaglia 2015, 5-7; Mercanti 2022).

Si tratta quindi di un periodo particolarmente intenso e in cui matura una visione ancora più ambiziosa della OL:

The idea is to build the library of Alexandria version 2. The idea of all books, all music, all video, all lectures, well: kind of everything, available to anybody, anywhere that would want to have access. Of

¹ <https://openlibrary.org/about>.

² Per ulteriori dettagli sulle versioni di OEB e ePub si può far riferimento al sito dell'International Digital Publishing Forum: <https://idpf.org/epub-older-versions> e del W3C: <https://www.w3.org/TR/epub-overview-33/>. Si precisa che in questo articolo con il termine 'e-book' si intendono solo libri digitali nativi.

³ Con 'commerciale' si fa riferimento all'editoria *trade* non *scholar*, professionale o scolastica.

⁴ <https://www.overdrive.com/>.

⁵ Ad esempio, eReolen in Danimarca, Onleihe in Germania, MLOL in Italia e altri. Per una panoramica sul prestito digitale in diversi paesi si vedano (Mount 2014; 2016; EBLIDA-EGIL 2022; 2023).

course, it's not gonna be done by one organisation, but we hope to play a role by helping move forward libraries, ourselves and making as much technology as required to be able to fulfil this goal (Pohl 2012).

Secondo Kahle il progetto OL è assolutamente coerente con quanto le biblioteche fanno abitualmente, ovvero acquistare libri e prestarli alle persone (Albanese 2011a). Ha però evidenziato due tipologie di problemi. Il primo è la costruzione della collezione, la quale sarebbe realizzabile grazie a un servizio di digitalizzazione proposto da IA a 10 cent per pagina. L'altro problema è relativo alla messa a disposizione delle opere, in particolare di quelle protette da copyright. Infatti la biblioteca dovrebbe offrire non solo «old stuff» ma anche «access right up to the current day» (Albanese 2011b). IA ha quindi avviato un programma di digitalizzazione con la collaborazione delle biblioteche stesse, le quali possiedono i libri a stampa.

Negli anni OL è diventata una realtà sempre più affermata e strutturata grazie alla partnership con biblioteche e altre istituzioni. I libri arrivano dalle biblioteche, sono acquistati o donati e spesso le digitalizzazioni sono possibili anche grazie al finanziamento della Kahle/Austin Foundation⁶. Il prestito può durare un'ora ed è rinnovabile, oppure 14 giorni, terminati i quali non è più possibile accedere al libro. Se il libro è già in prestito per 14 giorni la richiesta dell'utente entra in lista d'attesa. Il libro in prestito per un'ora può essere letto tramite BookReader, l'interfaccia proposta da IA, mentre quello in prestito per 14 giorni può essere anche scaricato, ovviamente in una versione protetta con un DRM (Digital Rights Management)⁷. I libri possono avere una versione digitale basata sul testo estratto grazie alla tecnologia OCR (Optical Character Recognition) ma si tratta di un testo che, se non supervisionato, può contenere errori.

Il prestito tramite la OL avviene in base alla teoria del Controlled Digital Lending (d'ora in poi in forma abbreviata CDL), la quale non è prevista dalla legge sul copyright statunitense ma che, secondo i teorizzatori, si basa sui principi della *first sale doctrine* e del *fair use*. Il CDL prevede il prestito 'controllato' – quindi in base a determinati criteri e pratiche – di copie digitalizzate di libri a stampa⁸. L'attività di digitalizzazione e prestito di OL ha attirato sia sostegno che critiche da parte di associazioni di autori negli Stati Uniti e in altri paesi. Authors Alliance ha sostenuto il CDL come uno strumento vantaggioso sia per i lettori che per gli autori, soprattutto per quanto riguarda le opere non disponibili in commercio (Authors Alliance 2018). Diversamente, The Authors Guild ha criticato la distribuzione da parte di IA di copie integrali di opere protette da copyright senza l'autorizzazione dei titolari dei diritti (Authors Guild 2018; 2019). Anche la National Writers Union si è dimostrata contraria al CDL (National Writers Union 2019). Nel Regno Unito la Society of Authors ha criticato la pratica di OL e giudicato illegale il CDL in UK (Society of Authors 2019). L'Australian Society of Authors si è definita «deeply concerned» per quanto realizzato dalla OL (Australian Society of Authors 2019). A inizio 2019 diverse associazioni di editori, autori, traduttori ecc. internazionali hanno firmato un appello rivolto a bibliotecari e lettori sostenendo che il CDL è un modello che danneggia gli autori (*Appeal on Controlled Digital Lending*, 2019). L'As-

⁶ Come si può leggere in filigrana nelle prime pagine di alcuni libri presi in prestito. La Kahle/Austin Foundation è la fondazione no profit di Brewster Kahle e sua moglie Mary Austin.

⁷ <https://help.archive.org/help/borrowing-from-the-lending-library/>.

⁸ Per un approfondimento si rimanda al capitolo *Il controlled digital lending*.

sociation of American Publishers (AAP), dal canto suo, ha definito il CDL una «flawed theory» (AAP 2019) e una «manufactured theory» (AAP 2023a) e lo ha criticato fortemente.

Nel frattempo l'attività di OL è cresciuta ulteriormente anche grazie all'acquisizione di Better World Books da parte di Better World Libraries, partner di Internet Archive. A seguito dell'acquisizione, Better World Books ha iniziato a offrire un flusso costante di libri a IA per la digitalizzazione⁹.

Dalla Open Library alla National Emergency Library

A inizio 2020, a causa della pandemia da Covid-19, anche le biblioteche negli Stati Uniti hanno smesso di offrire servizi in presenza (ALA 2020). Il 24 marzo IA ha annunciato la National Emergency Library (d'ora in poi in forma abbreviata NEL), una particolare 'versione' della OL che non prevede liste di attesa. In pratica viene ammesso il prestito da parte di più utenti in contemporanea, oltrepassando quindi i limiti previsti dalla OL e dal CDL. Nel complesso, IA offriva 1,4 milioni di libri moderni, più titoli di pubblico dominio. IA ha affermato che la digitalizzazione è stata effettuata con particolare attenzione alle opere del XX secolo che non hanno una versione digitale in commercio. I titoli in commercio possono invece essere presi in prestito tramite servizi come Overdrive. Nella visione di IA la NEL si rivolgeva in particolare agli studenti («to serve the nation's displaced learners» (Freeland 2020a)) e al pubblico generalista che non poteva accedere fisicamente alle biblioteche locali a causa della chiusura o per via della quarantena. Tale opportunità era rivolta a persone in tutto il mondo, non solo agli utenti statunitensi. In occasione del lancio della NEL, IA aveva annunciato che la sospensione delle liste d'attesa sarebbe durata fino al 30 giugno o oltre, se la pandemia fosse continuata (Freeland 2020a).

Dai commenti al post in cui veniva presentata la NEL è emerso il forte entusiasmo degli utenti, ma qualche autore aveva evidenziato che la disponibilità dei suoi libri online gratuitamente avrebbe potuto avere un impatto negativo sui suoi incassi. Agli autori veniva comunque data la possibilità di richiedere l'eliminazione di un titolo dalla NEL (Freeland 2020a)¹⁰. Il *sentiment* riguardo la NEL è quindi tutt'altro che omogeneo. Chiaramente sono molti gli entusiasti. Jill Lepore – storica e giornalista del New Yorker – ha addirittura proposto un paragone con il Council on Books in Wartime, il programma che prevedeva la rivendita di libri a basso prezzo ai soldati statunitensi durante la Seconda guerra mondiale (Lepore 2020)¹¹. Diversamente, a livello politico, il senatore Thom Tillis (presidente della Senate Judiciary Committee Subcommittee on Intellectual Property) si è detto preoccupato dal fatto che la NEL operi al di fuori della legge sul copyright (Albanese 2020a). Lo stesso Kahle ha risposto che la NEL è basata sul *fair use* ed è di grande utilità nel

⁹ Better World Books è stata fondata da tre studenti universitari dell'Indiana all'inizio del Duemila, i quali raccoglievano libri usati per rivenderli online. Questo avrebbe permesso di finanziarsi e di finanziare anche attività di alfabetizzazione e distribuzione di libri a persone in difficoltà. Questa idea di business è cresciuta fortemente nel corso degli anni, soprattutto grazie al recupero dei libri scartati dalle biblioteche per far spazio alle novità. Per maggiori dettagli si veda il sito web di Better World Books: https://www.betterworldbooks.com/go/Internet_Archive e (Freeland 2019; Enis 2019).

¹⁰ Ad esempio si veda il commento di Meghan McCarthy – autrice e illustratrice per bambini – del 26 marzo e la risposta di Chris Freeland di IA (Freeland 2020a).

¹¹ È interessante notare che la NEL e il contesto in cui nasce faccia tornare in mente a Lepore il contesto bellico della Seconda guerra mondiale.

momento difficile che si stava attraversando e in cui il patrimonio delle biblioteche, sul quale si è investito molto denaro pubblico, non era accessibile (Albanese 2020b).

Nello stesso tempo l'azione di IA di 'liberalizzare' in maniera assoluta i prestiti ha attirato forti critiche da parte di autori ed editori (Alter 2020). L'Association of American Publishers ha parlato di «aggressive, unlawful, and opportunistic attack on the rights of authors and publishers in the midst of the novel coronavirus pandemic» (AAP 2020a), ciò mentre gli editori stavano facendo molto per facilitare l'accesso e fornire materiali integrativi anche per la didattica. Quello che IA proponeva viene definito come «a cynical play to undermine copyright, and all the scientific, creative, and economic opportunity that it supports» (AAP 2020a).

Anche gli autori di Authors Guild si sono definiti sconvolti da quello che stava facendo IA in un contesto così difficile in cui non potevano promuovere i libri in presenza e non avevano altre fonti di reddito se non i diritti sulle copie vendute. Gli autori hanno sostenuto che IA metteva a disposizione anche titoli moderni protetti da copyright e che stava usando la scusa della pandemia «to push copyright law further out to the edges» (Authors Guild 2020a). Inoltre, si rifiutavano di definire IA come una biblioteca poiché non acquista né acquisisce in licenza i libri che fornisce in prestito (Authors Guild 2020a).

IA ha difeso la NEL sostenendo di essere una biblioteca, che le copie a stampa dei libri sono state acquistate o donate, che il prestito è avvenuto sulla base del CDL e che il CDL non viola la legge. Inoltre, ha ribadito che agli autori veniva data la possibilità di richiedere la rimozione dei propri titoli dalla NEL e che, dopo il 30 giugno, salvo una durata prolungata della pandemia, si sarebbe tornati al regime delle liste d'attesa previsto dalla OL (Freeland 2020b).

Hachette v. Internet Archive

A inizio giugno 2020 alcuni editori – Hachette Book Group, HarperCollins, John Wiley & Sons, Penguin Random House – hanno intentato una causa contro IA (Harris 2020; Milliot 2020). Il 10 giugno Kahle ha comunicato la chiusura anticipata della NEL, che è avvenuta il 16 giugno e non il 30 (o successivamente) come previsto (Kahle 2020b; Albanese 2020c). OL avrebbe quindi continuato a prestare i libri digitalizzati nella stessa modalità attiva prima del lancio della NEL, ripristinando quindi le liste d'attesa.

Dall'11 giugno IA ha iniziato a pubblicare post sull'impatto positivo della NEL riportando le testimonianze di chi ha beneficiato del programma come studenti, insegnanti, ricercatori, lettori ecc. (Freeland 2020c; 2020d; 2020e).

Nonostante la chiusura della NEL gli editori non hanno ritirato la causa contro IA. Nel marzo 2023 il giudice John Koeltl del distretto di New York ha accettato le istanze dei querelanti (Del Vecchio 2023)¹².

IA ha definito la sentenza come «a blow to all libraries and the communities we serve» (Freeland

¹² Hachette Book Group, Inc. v. Internet Archive, 542 F.Supp. 1156 (2023). Per il documento integrale della sentenza si veda <https://storage.courtlistener.com/recap/gov.uscourts.nysd.537900/gov.uscourts.nysd.537900.188.0.pdf>. In questa pagina viene ricostruito tutto il percorso che ha portato al processo: <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/91795-hachette-v-internet-archive-all-our-coverage.html>. Per la documentazione relativa al caso si veda: <https://www.courtlistener.com/docket/17211300/hachette-book-group-inc-v-internet-archive/>.

2023). Secondo IA si tratta di una sentenza che ha un impatto sulle biblioteche che usano il CDL. Comunque IA continuerà a lottare «for the traditional right of libraries to own, lend, and preserve books» (Freeland 2023a). Gli editori, dal canto loro, hanno considerato la sentenza come «an affirmation of publishing» e del copyright (AAP 2023b).

L'11 agosto 2023 – circa quattro mesi dopo quella che IA ha definito una «disappointing decision» (Freeland 2023b) – le parti in causa hanno presentato una proposta congiunta, come più volte richiesto dal giudice Koeltl. Il procedimento comprende una ingiunzione permanente riguardante i “covered books”, un accordo monetario confidenziale e prevede il diritto di IA di ricorrere in appello¹³. I querelanti e IA hanno però una idea diversa su cosa siano i “covered books”: per i primi l’ingiunzione riguarda tutti i libri digitalizzati senza autorizzazione, per IA invece solo quelli oggetto della causa (i 127 Works in Suit¹⁴, come si specificherà meglio in seguito nell’analisi del caso). Il giudice ha limitato l’ingiunzione a quei titoli resi disponibili dagli editori in base a licenze per il formato elettronico¹⁵. IA si è attenuta a quanto previsto dall’ingiunzione, e ha deciso di presentare ricorso, in quanto convinti che «libraries should be able to own, preserve, and lend digital books outside of the confines of temporary licensed access» e che «the judge made errors of law and fact in the decision» (Freeland 2023b; 2023d).

Nel frattempo anche alcune etichette discografiche hanno fatto causa a IA (Belanger 2023).

Prima di analizzare le ragioni e i principali elementi dell’accusa e della difesa è necessario approfondire il concetto di CDL.

Il Controlled Digital Lending

Cosa è il Controlled Digital Lending

Il CDL consente a una biblioteca di far circolare un titolo digitalizzato al posto di uno fisico in modo controllato. Una biblioteca può quindi prestare contemporaneamente solo il numero di copie che ha acquistato legalmente o che le sono state donate. Pertanto, anche a seguito della digitalizzazione di un libro, non potranno circolare in prestito più copie di quelle che la biblioteca possedeva prima della digitalizzazione. Il CDL si basa quindi sul rapporto tra posseduto e prestato («owned to loaned’ ratio») e la copia digitale non può aggiungersi alle copie fisiche in circolazione. Per fare un esempio, se la biblioteca possiede due copie a stampa di un libro e ne realizza una versione digitalizzata non potrà prestare contemporaneamente tre copie ma due, o quelle a stampa o una a stampa e una digitale. Il prestito della copia digitale deve avvenire per un periodo di tempo limitato e per utente singolo e, inoltre, è previsto l’impiego di misure tecniche appropriate per im-

¹³ [Proposed] Consent judgment and permanent injunction subject to reservation of right of appeal. <https://storage.courtlistener.com/recap/gov.uscourts.nysd.537900/gov.uscourts.nysd.537900.214.1.pdf>.

¹⁴ Con ‘Works in Suit’ si intendono i 127 titoli protetti da copyright pubblicati dagli editori querelanti e resi disponibili tramite la OL che sono oggetto della causa, cfr. *Hachette Book Group, Inc. v. Internet Archive, United States District Court for the Southern District of New York, opinion and order: John G. Koeltl, District Judge*, p. 1. <https://storage.courtlistener.com/recap/gov.uscourts.nysd.537900/gov.uscourts.nysd.537900.188.0.pdf>. Per la lista completa si veda l’Exhibit A (cfr. nota 36).

¹⁵ «An injunction covering all in-print books, including those the Publishers have not made available for electronic licensing, risks going “beyond the scope of the issues tried in the case”». <https://storage.courtlistener.com/recap/gov.uscourts.nysd.537900/gov.uscourts.nysd.537900.216.0.pdf>.

pedire agli utenti di conservare una copia o distribuire copie aggiuntive. I fattori di ‘controllo’ non riguardano quindi solo il periodo di tempo limitato durante il quale il libro digitale è accessibile, ma la stessa circolazione delle copie nei vari formati.

Questi sono i punti fondamentali per una corretta implementazione del CDL da parte delle biblioteche: 1) assicurarsi che le opere originali siano acquisite legalmente; 2) applicare il CDL solo alle opere che sono possedute e non concesse in licenza; 3) limitare il numero totale di copie in qualsiasi formato che sono in circolazione in qualsiasi momento al numero di copie fisiche che la biblioteca possiede legalmente (è necessario mantenere un rapporto tra “posseduto e prestato”); 4) prestare ogni versione digitale a un singolo utente alla volta, proprio come verrebbe prestata una copia fisica; 5) limitare il periodo di tempo per ciascun prestito a uno analogo a quello previsto per il prestito fisico; 6) usare DRM per impedire la copia e la redistribuzione delle copie (Courtney and Hansen 2018; Courtney *et al.* 2018).

Il CDL quindi ammette che la biblioteca – o chi per lei – digitalizzi un libro a stampa e lo dia in prestito. Più che la creazione di una copia digitale, avviene quindi una digitalizzazione, ovvero la creazione di un nuovo documento per l’accesso a un testo, il quale permette un utilizzo nuovo e diverso rispetto all’edizione a stampa.

Su cosa si basa il Controlled Digital Lending

Il CDL viene puntualmente definito nel *Position statement on controlled digital lending by libraries* del 2018 e nel *White paper on controlled digital lending of library books* dello stesso anno. I due autori del *White paper* sono David R. Hansen e Kyle K. Courtney, mentre il *Position statement* vede come autori gli stessi Hansen e Courtney e altri. Tra questi anche Michelle Wu, che nel 2011 aveva proposto qualcosa di simile (Wu 2011). Successivamente, nel 2017, Wu ha sostenuto progetti *digitize-and-lend* per ampliare e facilitare le possibilità di accesso ai testi. La pratica è stata proposta come il logico sviluppo del servizio di prestito bibliotecario nell’era digitale (Wu 2017).

Il *Position statement* e il *White paper* sono il punto di riferimento essenziale per tutte le biblioteche che vogliono avviare un progetto basato sul CDL. Non sono documenti tecnici, ma gli autori illustrano i principi del CDL cercando di inserirlo in una cornice legale. Infatti, il CDL non è previsto dalla legge sul copyright statunitense né da norme che riguardano le biblioteche.

Secondo i suoi teorizzatori, il CDL si basa sulla *first sale doctrine* e sul *fair use*, i quali sono le fondamenta dell’attività di prestito bibliotecario¹⁶. In questa sede non è possibile proporre un approfondimento dettagliato su questi due principi ma si sintetizzano gli aspetti centrali utili a comprendere meglio il CDL.

Secondo la Copyright law il titolare del copyright ha il diritto esclusivo di distribuire – o autorizzare la distribuzione – al pubblico dell’opera protetta da copyright mediante vendita o altro trasferimento di proprietà, oppure tramite noleggio o prestito¹⁷. In base alla *first sale doctrine*, una volta che un soggetto ha acquistato legalmente una copia di un’opera, il titolare del copyright perde la possibilità di controllare la successiva circolazione di quella copia. Pertanto, chi acquista una copia

¹⁶ Per dettagli si vedano le pagine *Copyright for libraries* sul sito web della American Library Association. <https://libguides.ala.org/copyright/general>.

¹⁷ 17 U.S.C. § 106(3). <https://www.copyright.gov/title17/92chap1.html#106>.

ne diviene il possessore e può farne quello che vuole, anche rivederla per averne un utile: il libro fisico è suo¹⁸. In base a questo principio le biblioteche negli Stati Uniti possono prestare le copie dei titoli che decidono di acquistare; inoltre possono fare delle copie di queste secondo particolari condizioni¹⁹. La realtà statunitense è diversa da quella europea dove le biblioteche prestano i libri non solo sulla base dell'esaurimento del diritto di distribuzione (principio di esaurimento comunitario)²⁰, ma anche della deroga prevista al diritto esclusivo di prestito da parte di istituzioni pubbliche, così come definito dalla direttiva 2006/115/CE (già dalla 92/100/CEE)²¹.

Negli Stati Uniti il *fair use* prevede un uso limitato di materiali protetti da copyright senza il permesso dei titolari del copyright. Ciò può avvenire per scopi didattici, di ricerca e critica. Il legislatore prevede quattro parametri per poter valutare se un uso non autorizzato rientra nel *fair use* oppure no. Questi sono: 1) la finalità e il carattere dell'utilizzo, incluso se l'utilizzo è di natura commerciale o per scopi educativi e senza fini di lucro; 2) la natura dell'opera protetta da copyright; 3) la quantità e la sostanzialità dell'opera utilizzata in rapporto all'insieme dell'opera protetta da copyright; 4) l'effetto dell'uso sul mercato potenziale o sul valore dell'opera protetta da copyright²².

Alcuni casi su *first sale doctrine* e *fair use*: ReDigi, HathiTrust e Google Books

I casi che qui sinteticamente si illustrano sono importanti perché riguardano la *first sale doctrine* e il *fair use* e sono ampiamente citati nel *Position statement* e il *White paper*, nonché nei documenti che riguardano il caso Hachette v. Internet Archive.

Un caso al quale è necessario fare riferimento è certamente Capitol Records, LLC v. ReDigi Inc. ReDigi è stato un servizio che permetteva la rivendita di file musicali acquistati online, creando così una sorta di 'mercato dell'usato' di copie. La Corte ha definito che la *first sale doctrine* non si applica alle copie digitali di un'opera e un'attività come quella di ReDigi non poteva rientrare in tale principio. Inoltre, l'attività di ReDigi non era ammessa neanche in base ai fattori del *fair use* poiché le riproduzioni erano integrali e le loro rivendite in competizione con la vendita autorizzata dai titolari del copyright. Inoltre ReDigi era un attore commerciale, e le persone usavano il servizio con finalità di lucro²³.

È interessante ricordare che nei primi anni Dieci ReDigi si proponeva di avviare un servizio di rivendita di e-book (Rosen 2012; 2013), e lo stesso intendeva fare Amazon (Dewey 2013; Wohlsen 2013).

In Europa qualcosa di simile è accaduto con il caso Nederlands Uitgeversverbond and Groep Algemene Uitgevers v. Tom Kabinet Internet BV and others (C-263/18) che ha visto contrapposte due associazioni degli editori olandesi e la Tom Kabinet, la quale consentiva la rivendita online di

¹⁸ 17 U.S.C. § 109(a). <https://www.copyright.gov/title17/92chap1.html#109>.

¹⁹ 17 U.S.C. § 108. <https://www.copyright.gov/title17/92chap1.html#108>.

²⁰ Art. 4 della direttiva 2001/29/CE. <https://eur-lex.europa.eu/legal-content/IT/ALL/?uri=CELEX%3A32001L0029>.

²¹ Art. 6 della dir. 2006/115/CE. <https://eur-lex.europa.eu/legal-content/IT/ALL/?uri=CELEX%3A32006L0115>.

²² 17 U.S.C. § 107. <https://www.copyright.gov/title17/92chap1.html#107>.

²³ Capitol Records, LLC v. ReDigi Inc., 934 F. Supp. 2d 640 (S.D.N.Y. 2013). Il caso è andato avanti dal 2012 al 2018 con la sentenza del Secondo circuito. Per la documentazione relativa al caso si veda: <https://www.courtlistener.com/docket/4350067/capitol-records-llc-v-redigi-inc/>. Sull'impatto del caso ReDigi sul CDL si veda (Wu 2019).

e-book. In questo caso la Corte ha definito che l'esaurimento non si applica alle copie digitali di e-book e quindi ha deciso a favore delle associazioni degli editori²⁴.

Tornando agli Stati Uniti, un caso importante per affrontare il tema del *fair use* è certamente Authors Guild v. HathiTrust. HathiTrust Digital Library proponeva scansioni di libri cartacei, molti dei quali erano opere protette da copyright. HathiTrust usava queste digitalizzazioni per finalità di conservazione, per offrire un servizio di indicizzazione per la ricerca *full text* e per servizi rivolti a persone con *print disabilities* certificate, come non vedenti e ipovedenti. La Authors Guild e altre associazioni degli autori hanno accusato HathiTrust di offrire scansioni non autorizzate di opere protette da copyright. La Corte ha però definito che l'attività di HathiTrust rientrava nel *fair use* e quindi non si trattava di un caso di violazione del copyright. È inoltre necessario considerare che i servizi proposti da HathiTrust rientravano nel concetto di 'uso trasformativo'. Questo non fa parte dei quattro fattori del *fair use*, ma per certi versi è connesso al primo («the purpose and character of the use»). Esso riguarda i casi in cui la copia di un'opera protetta viene usata per offrire qualcosa di nuovo e con un altro scopo rispetto a quello dell'originale, senza sostituire l'uso originale dell'opera, quindi senza competizione con le copie dell'opera in commercio²⁵. Nel caso in questione la sentenza ha definito che i servizi di HathiTrust potevano rientrare nel concetto di uso trasformativo e nell'ambito del *fair use*. Ciò però non giustifica certamente HathiTrust per quanto riguarda azioni come la creazione di altre copie dei libri per altre finalità (De Robbio 2012; Albanese 2014)²⁶.

Un altro caso al quale è necessario far riferimento è certamente Authors Guild, Inc. v. Google, Inc. L'attività di Google con il servizio Google Books – «il primo tentativo di digitalizzazione libraria su scala globale» (Roncaglia 2010, 184) – è ben nota e la letteratura è decisamente ampia²⁷. In questo caso Authors Guild ha intentato una causa per violazione del copyright per la digitalizzazione di opere senza il consenso dei titolari dei diritti e senza pagare una licenza. Inoltre, le digitalizzazioni sono state archiviate, indicizzate e usate per ricerche testuali tramite un motore di ricerca. Sono molti i punti che avvicinano questa offerta a quella di HathiTrust, nonostante Google sia un'impresa a fini di lucro. Google Books non richiede però un pagamento diretto: oltre a ottenere informazioni bibliografiche, un utente può fare una ricerca testuale gratuita, quindi visualizzare brevi frammenti di testo (*snippets*) che comprendono la parola cercata. La Corte non ha ritenuto che quanto offerto da Google sia in contrasto con i quattro fattori del *fair use*. L'uso che una persona può fare di Google Books non è considerato come un sostituto competitivo del mercato. Inoltre, ha riconosciuto il ruolo trasformativo del servizio di ricerca testuale e presentazione tramite lo *snippet*. Il fatto che Google è attore economico non nega il *fair use*: molti casi che rientrano nel *fair use* – come le parodie o le citazioni – sono fatti commercialmente e a scopo di lucro²⁸.

²⁴ Court of Justice of the European Union, *Nederlands Uitgeversverbond e Groep Algemene Uitgevers contro Tom Kabinet Internet BV e a.*, C-263/18. Per la documentazione relativa al caso si veda: <https://curia.europa.eu/juris/liste.jsf?num=C-263/18>.

²⁵ U.S. Copyright Office Fair Use Index. <https://www.copyright.gov/fair-use/>.

²⁶ The Authors Guild, Inc. v. HathiTrust, 1:11-cv-06351, (S.D.N.Y.) and Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014). Per una sintesi del caso si veda: <https://www.copyright.gov/fair-use/summaries/authorsguild-hathitrust-2dcir2014.pdf>.

²⁷ Per una bibliografia fino al 2011 si veda Charles W. Bailey, Jr., Google Books Bibliography. <https://digital-scholarship.org/gbsb/gbsb.htm>.

²⁸ Authors Guild v. Google 721 F.3d 132 (2nd Cir. 2015). Per una sintesi del caso: <https://www.copyright.gov/fair-use/summaries/authorsguild-google-2dcir2015.pdf>.

Il *Position statement on controlled digital lending by libraries* e il *White paper on controlled digital lending of library books*

Dopo questo brevissimo ma necessario excursus, torniamo ora ad approfondire il *Position statement* e il *White paper*. Come anticipato all’inizio del capitolo, questi documenti sono il riferimento fondamentale per tutte quelle biblioteche che vogliono implementare il CDL (quindi non riguarda solo IA, per quanto con la OL ne abbia proposta la realizzazione più nota e diffusa)²⁹.

Il *Position statement* è il documento generale in cui viene presentato il CDL. In supporto al *Position statement* Hansen e Courtney hanno diffuso anche il *White paper*, un documento più articolato in cui, oltre ad approfondire i principi giuridici, vengono ripresi casi giuridici conclusi e in corso (alcuni dei quali già visti nel paragrafo precedente) e viene illustrato il cosiddetto Problema del libro del XX secolo (quest’ultimo tema è molto interessante e riguarda quelle opere tutelate dal diritto d’autore che sono *out of print* nella versione a stampa e non hanno una versione digitale, quindi sono fruibili tramite la biblioteca o, eventualmente, acquistabili al mercato dell’usato).

La *first sale doctrine* è alla base del prestito di libri. Come scrivono Hansen e Courtney nel *White paper* «Libraries were built on it» (Courtney e Hansen 2018). Restano però molte domande aperte per quanto riguarda il digitale. Infatti rimandano al caso ReDigi per evidenziare la difficoltà nel definire cosa accade quando si trasferisce un file digitale, e come questo trasferimento si configura giuridicamente. Si chiedono: «Is the transfer of a digital copy from one device to another the transfer of a particular copy or the creation of a new copy?» (Courtney e Hansen 2018). Nel caso ReDigi, seppur in base alle sue peculiarità, è stato comunque definito che la *first sale* non si applica alle copie digitali di un contenuto. Nel *White Paper* si sostiene però che le biblioteche «that seek to utilize CDL should still be able to apply the first sale doctrine’s rationale in the fair use context» (Courtney e Hansen 2018). Il prestito delle biblioteche è considerato speciale («we view library lending uses as special» (Courtney e Hansen 2018)) e anche il CDL si ‘regge’ sul *fair use*³⁰. Questo – particolarmente articolato e non semplice da valutare – richiede un’ampia trattazione da parte dei teorizzatori del CDL (che in questa sede può solo essere abbozzata). Secondo loro non si è mai discusso in maniera esaustiva sul rapporto tra la prima vendita e lo scopo e il carattere del servizio. Anche per quanto riguarda il caso ReDigi, ricordano che si trattava di un attore commerciale e l’uso proposto non era trasformativo, mentre le biblioteche che implementano il CDL lo fanno per scopi di ricerca e apprendimento e non commerciali. Inoltre, l’utente che fruisce del libro digitalizzato tramite la biblioteca non ottiene un beneficio economico, come invece il rivenditore del file digitale tramite ReDigi (o altri servizi simili). Dal canto suo, invece, il CDL permetterebbe alle biblioteche di svolgere la propria missione con un maggiore impatto sulla società, facilitando l’accesso anche a persone con disabilità o che non possono facilmente accedere ai libri a stampa. Per quanto riguarda l’uso trasformativo del CDL non sembra ci sia accordo tra gli autori del *White paper* e quelli del *Position statement* (Courtney and Hansen 2018, nota 87). Secondo gli autori del *White paper* un caso interessante per il CDL è piuttosto quello di HathíTrust, dove non è presente

²⁹ Si riporta ad esempio l’esperienza di biblioteche di diverse realtà (De Castell 2022, 1-35; Pang 2022, 154-166; Board e Stutzman 2020, 157-165).

³⁰ In altro contesto Courtney parla di ‘superpoteri’ delle biblioteche in relazione al *fair use* (Courtney 2020a).

l'uso commerciale e il servizio è proposto per un maggiore accesso per finalità di ricerca e apprendimento.

I teorizzatori del CDL sostengono che il secondo fattore del *fair use* abbia svolto raramente un ruolo significativo nella valutazione del *fair use*. Per quanto riguarda il CDL è difficile fare una valutazione perché le collezioni delle biblioteche comprendono opere diverse, quindi alcune rientreranno meglio di altre nel *fair use*.

Per quanto riguarda il terzo fattore i teorizzatori del CDL notano che in fase di giudizio viene messa in relazione la quantità dell'opera utilizzata agli scopi dell'utilizzo. Pertanto anche l'uso dell'opera intera può potenzialmente rientrare nel *fair use*, se l'uso è coerente con il primo fattore. Quindi secondo loro il CDL potrebbe rientrare nel terzo fattore del *fair use*, considerando anche le limitazioni previste (periodo di prestito limitato, impossibilità di copiare e diffondere ulteriormente l'opera ecc.). Il quarto fattore riguarda il rapporto tra l'utilizzo dell'opera e il suo mercato, ovvero il pericolo che un determinato utilizzo minacci il mercato di un'opera. Il CDL è considerato coerente a questo fattore poiché «the market effect of CDL is nearly identical to the market effect already favored under the first sale doctrine» (Courtney e Hansen 2018). Il CDL infatti è strutturato in modo che l'utilizzo che viene fatto della copia digitale è lo stesso previsto per la copia cartacea, utilizzo per quale i titolari del copyright hanno ottenuto un guadagno con la prima vendita (Courtney 2020b). Quindi il CDL non danneggerebbe il mercato in maniera diversa rispetto agli usi consentiti dalla legge poiché «there is no situation in which the library is getting use of two copies for the price of one» (Courtney e Hansen 2018). Inoltre bisogna considerare che non esiste un effettivo mercato per i libri del XX secolo protetti da copyright che non vengono più stampati e non hanno una versione digitale.

I teorizzatori evidenziano che il prestito di copie digitalizzate non può essere completamente equiparato a quello fisico. Infatti è privo di quegli attriti («frictions») tipici dell'analogico (ad esempio il doversi recare in biblioteca per prendere il libro in prestito, il fatto che il libro a stampa è 'fisicamente' in prestito a un'altra persona ecc.); le copie digitali non si degradano come quelle a stampa; c'è però il pericolo della pirateria e per questo è fondamentale la sicurezza degli archivi digitali. Questi aspetti vengono approfonditi nel *White paper*, sostenendo che nessuno di essi può essere considerato un ostacolo per un CDL ben progettato.

Gli autori dedicano inoltre un'ampia sezione del *White paper* alla progettazione del sistema e alla mitigazione del rischio. I principali rischi sono di natura legale: il rischio che una biblioteca venga citata in giudizio, che perda la causa e le relative conseguenze. Concretamente questo si traduce in una perdita in termini economici e di tempo, oltre che in un danno alla reputazione.

«What IA does is copyright infringement, plain and simple, and it must be stopped»³¹

Dopo avere delineato il contesto nel quale è iniziata l'azione giudiziaria e aver approfondito il CDL, in questo capitolo si cercherà di focalizzare in maniera sintetica le principali ragioni dell'accusa e la decisione della Corte.

³¹ Hachette Book Group, Inc. v. Internet Archive, 1:20-cv-04160, (S.D.N.Y. Jun 01, 2020) ECF No. 1, para 52. <https://www.courtlistener.com/docket/17211300/1/hachette-book-group-inc-v-internet-archive/>.

Defendant IA is engaged in willful mass copyright infringement. Without any license or any payment to authors or publishers, IA scans print books, uploads these illegally scanned books to its servers, and distributes verbatim digital copies of the books *in whole* via public-facing websites. With just a few clicks, any Internet-connected user can download *complete* digital copies of in-copyright books from Defendant³².

Come si può leggere da questo breve estratto, i querelanti contestano a IA la violazione del copyright con le aggravanti di essere massiva e volontaria. IA digitalizza libri a stampa protetti da copyright senza autorizzazione e al di fuori di quanto previsto dalla legge, quindi archivia e distribuisce al pubblico tali copie. Questa attività vede la riproduzione fedele e integrale delle opere a stampa. Gli editori la definiscono un atto di pirateria intenzionale, sistematico e su larga scala, poiché IA digitalizza una grande quantità di libri e li presta in tutto il mondo³³. Precisano quindi che l'azione legale non riguarda un'attività occasionale o la condivisione di opere di pubblico dominio o in licenza, ma – questo è un punto centrale – un'azione massiva che vede IA raccogliere «collection of truckloads of in-copyright books to scan, reproduce, and then distribute digital bootleg versions online»³⁴. Gli editori mettono a disposizione della Corte anche una lista di titoli protetti da copyright che sono risultati disponibili tramite la OL, molti dei quali sono opere di intrattenimento come romanzi³⁵. Seppur la causa sia stata intentata meno di tre mesi dopo il lancio della NEL, le accuse dei querelanti non riguardano esclusivamente questa esperienza 'di emergenza' ma, da un lato la crescita progressiva dell'attività di IA con la OL (digitalizzazione massiva, prestito in tutto il mondo ecc.) e dall'altro il CDL. La denuncia è quindi particolarmente ampia ed è necessario evidenziare alcuni ambiti entro i quali si muovono le accuse a IA. Uno di questi riguarda l'identità stessa di IA. Infatti, secondo i querelanti, questa non può essere definita una biblioteca, e così la OL. Inoltre, seppur IA si definisca una organizzazione no-profit, in realtà ha ingenti entrate che vengono da donatori tramite la Kahle/Austin Foundation e dalla vendita di servizi di digitalizzazione alle biblioteche³⁶. Inoltre, non può essere definita una biblioteca perché la digitalizzazione e il prestito non sono autorizzati e non avvengono secondo modalità previste dalla legge sul copyright. L'infondatezza del CDL è uno dei temi centrali dell'accusa. Si critica l'idea che il CDL – e la OL – si basi sulla *first sale doctrine* e rispetti i quattro fattori del *fair use*; inoltre non si riconosce l'uso trasformativo³⁷. Gli stessi querelanti affermano che il CDL sembra essere stato proposto in risposta alle obiezioni dei titolari del copyright alle attività di IA³⁸. E a seguito della pubblicazione del *Position statement* e dal *White paper* le associazioni degli editori e degli autori hanno iniziato ad attaccare con maggiore forza la OL e il CDL.

³² *Ivi*, para 2. Il corsivo è nell'originale.

³³ *Ivi*, para 3. IA mette a disposizione la possibilità di richiedere l'eliminazione di un titolo dalla OL, ma ciò avviene solo dopo la sua digitalizzazione e pubblicazione.

³⁴ *Ivi*, para 6.

³⁵ Si riportano i titoli di alcuni romanzi come *The lord of the flies* di William Golding, *Song of Solomon* di Toni Morrison, *Gone girl* di Gillian Flynn, *The road* di Cormac McCarthy; cfr. *Ivi*, para 21. Per la lista completa si veda l'Exhibit A allegato dall'accusa: <https://storage.courtlistener.com/recap/gov.uscourts.nysd.537900/gov.uscourts.nysd.537900.1.1.pdf>.

³⁶ Per il servizio di digitalizzazione: <https://archive.org/scanning>; Kahle sostiene che «Libraries paying for our scanning services is a major source of earned income for the Internet Archive» (Kahle 2020a).

³⁷ Hachette Book Group, Inc. v. Internet Archive, cit., para 11.

³⁸ *Ivi*, para 107.

Quanto proposto da IA con la OL è considerato dall'accusa completamente diverso da realtà già valutate in sede giudiziale come Google Books e Hathitrust. Gli editori mettono in dubbio anche l'effettiva applicazione e la verificabilità del "rapporto tra posseduto e prestato" alla base del CDL. A questo si aggiunge che con la NEL il concetto di CDL è stato spinto all'estremo eliminando le liste d'attesa.

Un altro tema centrale dell'accusa è l'impatto che la OL – e la NEL – ha sul mercato del libro e il suo ecosistema. Questo riguarda le mancate vendite per editori e librai, necessarie perché esista una produzione e un mercato editoriale di qualità; le mancate vendite per gli autori, le quali sono fondamentali non solo per vivere ma anche per continuare a scrivere opere; l'impatto negativo sul lavoro delle persone impiegate nel processo editoriale; la concorrenza della OL con sistemi di prestito digitale basati su *licensing*, i quali invece sono autorizzati dai titolari del copyright. Secondo i querelanti, IA opera con un «profound disregard of the boundaries and balance of core copyright principles» e, nonostante l'idea della '*free knowledge*', sta cercando di distruggere l'ecosistema del libro («the carefully calibrated ecosystem that makes books possible»)³⁹.

Le affermazioni e le accuse dei querelanti sono decisamente molto dure. Non lo sono di meno quelle della Association of American Publishers, secondo la quale IA colpisce duramente il quadro legale del copyright appropriandosi indebitamente degli investimenti degli editori e degli autori senza tenere conto della legge sul copyright (AAP 2020b). La Authors Guild, dal canto suo, parla di furto di IA ai danni degli autori e di pirateria: quello offerto da IA non è assolutamente un servizio pubblico, nonostante continui a definirsi una biblioteca (Authors Guild 2020b)⁴⁰.

A fine luglio 2020 Kahle ha chiesto agli editori di ritirare il contenzioso poiché non possono attaccare il lavoro delle biblioteche e ciò che le biblioteche hanno sempre fatto (Albanese 2020d). Ma da parte degli editori non c'è stato un passo indietro (anche perché, probabilmente, avrebbe per certi versi significato accettare il CDL e ulteriori sviluppi della OL).

IA si è difesa ufficialmente dalle accuse degli editori sostenendo il valore del proprio operato per offrire l'accesso alla cultura per finalità di studio e ricerca, affermando ancora una volta che tale attività è ammessa sulla base della *first sale doctrine* e del *fair use*. Inoltre ha messo in dubbio il danno economico subito dagli editori a causa della OL⁴¹.

La Corte – dopo aver ricostruito la realtà del *digital lending* bibliotecario negli Stati Uniti, i servizi offerti da IA e il suo rapporto con Better World Books, la tipologia di titoli offerti dalla OL ecc. – ha definito che l'operato della OL non è ammesso dalla *first sale doctrine* e ha messo in dubbio la capacità della stessa IA di controllare che la copia cartacea e quella digitale non siano in prestito contemporaneo⁴².

Riguardo al *fair use*, la Corte ha argomentato che nell'attività della OL non è evidente l'uso trasfor-

³⁹ *Ivi*, para 13.

⁴⁰ In particolare Douglas Preston, il presidente della Authors Guild, sostiene che quello che IA sta facendo è come gettare un mattone contro la vetrina di un negozio di alimentari e distribuire cibo per poi vantarsi di offrire un servizio pubblico.

⁴¹ Per questo hanno richiesto dati di vendite mensili dal 2011, dati che gli editori hanno rifiutato di produrre con questa accuratezza perché sarebbe un'operazione troppo costosa e perché i fattori che possono determinare l'andamento volubile del mercato di un libro mese per mese sono molteplici e difficilmente ponderabili nel report richiesto. Una attività di questo tipo è considerata legalmente irrilevante (Albanese 2021).

⁴² Hachette Book Group, Inc. v. Internet Archive, United States District Court for the Southern District of New York, Opinion and Order: John G. Koeltl, District Judge, p. 31-32, <https://storage.courtlistener.com/recap/gov.uscourts.nysd.537900/gov.uscourts.nysd.537900.188.0.pdf>.

mativo, come invece per HathiTrust⁴³. Piuttosto, nel sostenere l'uso trasformativo, IA «distorts the way courts have treated utility-expanding transformative uses»⁴⁴.

Nel definire la natura e le finalità dell'uso, la Corte ha posto l'attenzione non tanto sulla natura dell'attività di IA (aspetto fortemente contestato dagli editori), ma sulla possibilità di trarre profitto dai diritti d'autore senza pagare quanto dovuto. IA non paga nulla ai titolari del copyright, ma usa la OL «to attract new members, solicit donations, and bolster its standing in the library community»; e, nello stesso tempo, grazie all'affiliazione con Better World Books, vende libri tramite la OL e guadagna da queste transazioni⁴⁵.

Anche il secondo fattore del fair use è a favore dei querelanti, poiché le opere oggetto del procedimento messe a disposizione dalla OL sono edite, protette da copyright e di diverso genere, sia fiction che non fiction («factual or informational»)⁴⁶.

Per quanto riguarda il terzo fattore, è evidente che la OL offra digitalizzazioni integrali delle opere. E, seppur a volte copiare l'intera opera sia «ragionevolmente necessario» («reasonably necessary») anche per un uso trasformativo (come nei casi Google Books e HathiTrust), IA digitalizza l'opera per offrirla integralmente al pubblico. Anche in base a questo fattore l'attività di IA non sembra rientrare nel *fair use*⁴⁷.

Il quarto fattore riguarda il rapporto con il mercato e anche in questo caso l'attività di IA non è coerente con quanto previsto dal *fair use*. Facendo riferimento al caso ReDigi, la Corte sostiene che «when a secondary use competes in the rightsholder's market as an effective substitute for the original, it impedes the purpose of copyright». Infatti esiste già un sistema di prestito basato su licenze (ad esempio quello offerto dalle biblioteche tramite Overdrive). Inoltre, IA potrebbe espandere ulteriormente la propria attività insieme ad altri partner, e altri soggetti potrebbero offrire un servizio simile a quello di IA. Secondo la Corte «this plainly risks expanded future displacement of the Publishers' potential revenues»⁴⁸. Dal canto suo IA sostiene che il prestito tramite OL è diverso da quello basato sulle licenze e non compete con esso. Inoltre, la vendita dei libri a stampa proposti su OL non è diminuita e, nonostante la fine della NEL e la rimozione di certi titoli, i prestiti tramite Overdrive non sono diminuiti. Quindi per IA non c'è un impatto negativo sul mercato. Secondo la Corte però, per avere un «significantly competing substitute» non è necessario replicare il sistema di *licensing* di Overdrive o altri⁴⁹. Inoltre, i dati delle vendite e dei prestiti digitali non dimostrerebbero che gli editori hanno avuto perfino vantaggi dall'attività di IA ma «they show at best that the presence of the Works in Suit in IA's online library correlated, however weakly, with positive financial indicators for the Publishers in other areas. They do not show that IA's conduct caused these benefits to the Publishers»⁵⁰. Inoltre bisogna considerare che gli editori non definiscono il prezzo dei libri in commercio tenendo conto anche della creazione di una copia digitale⁵¹.

⁴³ *Ivi*, 19.

⁴⁴ *Ivi*, 22.

⁴⁵ *Ivi*, 26-27.

⁴⁶ *Ivi*, 34-36.

⁴⁷ *Ivi*, 37.

⁴⁸ *Ivi*, 39-40.

⁴⁹ *Ivi*, 41.

⁵⁰ *Ivi*, 43.

⁵¹ *Ivi*, 39-44.

In conclusione, né la *first sale doctrine* né i fattori del *fair use* ammettono l'attività di IA con la OL. Ciò è ancora più evidente con la NEL, nella quale i prestiti sono avvenuti in maniera simultanea e senza liste d'attesa⁵². IA può digitalizzare e distribuire opere in pubblico dominio, ma ciò che il *fair use* non ammette è la digitalizzazione massiva sulla base dell'acquisizione e la distribuzione «of complete copyrighted works in a way that does not transform those works and that creates directly competing substitutes for the originals. Because that is what IA has done with respect to the Works in Suit, its defense of fair use fails as a matter of law»⁵³. Come evidenziato dal giudice «each enumerated fair use factor favors the Publishers, and although these factors are not exclusive, IA has identified no additional relevant considerations. At bottom, IA's fair use defense rests on the notion that lawfully acquiring a copyrighted print book entitles the recipient to make an unauthorized copy and distribute it in place of the print book, so long as it does not simultaneously lend the print book. But no case or legal principle supports that notion. Every authority points the other direction»⁵⁴.

Conclusioni

Alla luce degli sviluppi del caso Hachette v. Internet Archive dobbiamo chiederci quale sarà il futuro di IA e dei servizi offerti, se e come potrà essere penalizzata in generale l'attività di IA. Il procedimento giudiziario riguarda la OL e le attività connesse, ma IA offre anche importanti strumenti e servizi come la Wayback Machine. Una risposta ce la fornisce direttamente IA: la sentenza comporta «a significant loss of access to valuable knowledge for the public», ma non ha un impatto negativo sulle altre attività di IA. Infatti, IA può continuare a digitalizzare libri per scopi di conservazione, può fornire accesso ai libri (anche a persone «with qualified print disabilities»), può continuare a fornire un accesso parziale ai testi in base a quanto previsto dal *fair use*, può mettere a disposizione testi di pubblico dominio. L'ingiunzione, infatti, non riguarda i libri *out of print* (Freeland 2023c).

Possiamo inoltre chiederci quale sarà il futuro del CDL, anche al di là dell'attività di IA. Questo concetto giuridico-biblioteconomico si spognerà progressivamente oppure si cercherà di implementarlo in modalità diverse dalla OL (ad esempio in specifici contesti bibliotecari, per un numero ristretto di utenti e per un numero limitato di opere)?

La vicenda giudiziaria e le dichiarazioni delle parti rendono evidenti tensioni tra i soggetti interessati in una 'battaglia' che vede coinvolte anche le biblioteche, ma nella quale queste rischiano di essere nell'ombra (o addirittura di essere il 'campo di battaglia'). Da una parte IA a seguito della sentenza ha parlato di perdita per sé, gli utenti e le biblioteche, e di una perdita di accesso alla conoscenza (Freeland 2023c). Ma di fatto l'ingiunzione riguarda opere di cui è disponibile il formato digitale in licenza, quindi potenzialmente accessibili tramite le biblioteche. E inoltre IA può continuare a offrire servizi alle biblioteche e accesso ai documenti secondo quanto previsto dalla legge. IA considera ingiusta la sentenza, nella quale però si riconosce che una attività di di-

⁵² *Ivi*, 46.

⁵³ *Ivi*, 45-46.

⁵⁴ *Ivi*, 45. Per ulteriori riflessioni e approfondimenti precedenti la sentenza si veda (Schard 2021, 53-58; Schwabach 2021, 187-216; Tondelli 2021, 420-440).

gitalizzazione massiva e prestito a seguito dell'acquisto di un libro da parte della biblioteca non è ammessa dalla Copyright law.

Dall'altra parte, gli editori accusano IA per la sua attività basata sulla teoria del CDL, la quale è in concorrenza con la disponibilità di libri digitali accessibili tramite le biblioteche sulla base di licenze conformi alla legge. Ma le biblioteche hanno più volte 'lottato' contro pratiche di *over-pricing* ed embargo applicate da parte degli stessi editori statunitensi, i quali si sono rifiutati di fornire e-book in licenza o li hanno forniti a prezzi notevolmente alti (IFLA 2012; Sisto 2022). Da parte loro, i teorizzatori del CDL sono stati critici verso il sistema delle licenze (Courtney 2020b) e, nello stesso tempo, IA ha sostenuto che le biblioteche dovrebbero prestare i libri digitali «outside of the confines of temporary licensed access» (Freeland 2023b). Il conflitto si genera proprio nella definizione di questi 'confini'. Nel mondo delle biblioteche IFLA vede il CDL come «an alternative to a licensing approach» (IFLA 2021) mentre EBLIDA sostiene che non lo è (EBLIDA-EGIL 2023)⁵⁵.

Il conflitto è inoltre legato alla presunta concorrenzialità tra vendita e prestito. Questa è stata ripresa anche nel contesto del *digital lending*, in cui le licenze presentano degli attriti (*frictions*) tali da differenziare un e-book in prestito da uno in commercio⁵⁶. Nel caso in questione si aggiunge una ulteriore potenziale concorrenzialità, quella tra accesso digitale tramite licenza e non. Ma in che misura sono realmente concorrenziali? Per rispondere a questa domanda non ci possiamo fermare alle sole questioni giuridiche, ma dobbiamo assumere anche uno 'sguardo biblioteconomico', poiché ciò di cui si sta discutendo è la creazione di forme e modelli di accesso da remoto a contenuti. Quindi dobbiamo analizzare il rapporto tra il CDL e il *digital lending*. Infatti, CDL e prestito digitale sono molto diversi per più di una ragione. Se ne evidenzia una tra le principali: il CDL riguarda digitalizzazioni di libri a stampa, invece il prestito digitale vari contenuti digitali nativi, come e-book e audiolibri. Nonostante questo gli editori, la Corte, i teorizzatori del CDL e la stessa IA mettono spesso a confronto queste due forme di prestito, ora per evidenziare una concorrenza tra di loro, ora per escluderla (Mercanti 2023).

Nell'analizzare le varie tipologie di libro digitale e le forme e modelli di accesso con uno sguardo biblioteconomico dobbiamo considerare i bisogni delle persone. Infatti è importante riflettere sul fatto che il libro a stampa, la sua digitalizzazione, la versione in e-book sono di fatto tre documenti che, seppur hanno per oggetto lo stesso testo, offrono tre opportunità di accesso e lettura diverse, in base alle possibilità, agli obiettivi delle persone e ai dispositivi digitali tramite i quali si fruisce del libro. La sfida, oggi e nel futuro, non è bloccare e vincolare le possibilità di accesso e fruizione ma capire come integrarle nel rispetto della legge e degli interessi dei vari attori coinvolti, anche cercando di sviluppare nuove forme di accesso e potenziando prodotti e servizi, e soprattutto rispondendo ai bisogni delle persone. Servono studi per capire come le persone usano le diverse forme di libro tra cartaceo e digitale, anche in maniera integrata tra acquisto (anche in modalità abbonamento) e accesso tramite la biblioteca, e come questi accessi e modalità di fruizione sono in relazione, si potenziano e si arricchiscono vicendevolmente.

⁵⁵ Resta fermo il fatto che la Copyright law statunitense e la legge sul diritto d'autore negli stati membri presentano delle differenze sostanziali, anche per quanto riguarda il prestito bibliotecario.

⁵⁶ Si pensi ad esempio ai modelli di licenza per utente singolo i quali impediscono a più utenti di prendere in prestito uno stesso e-book contemporaneamente.

Mantenendo lo ‘sguardo biblioteconomico’ ci possiamo chiedere non solo dove sono le biblioteche nel caso giuridico in questione, ma anche quale è il loro posizionamento riguardo le varie forme di accesso digitale a contenuti protetti dal diritto d’autore. Infatti, le biblioteche devono certamente innovare la propria offerta per offrire servizi qualitativamente di alto livello, anche insieme ad altri soggetti (pubblici o privati), ma devono farlo nel rispetto della legge sul diritto d’autore e in maniera sostenibile all’interno del sistema del libro e della lettura.

Nello stesso tempo, il caso Hachette v. Internet Archive porta a riflettere sul rapporto tra biblioteche, editoria e piattaforme di diversa tipologia, sulle relazioni tra le varie missioni, configurazioni e proposte. Si tratta di un rapporto non semplice, neanche quando sono coinvolti attori come IA che si propongono come una biblioteca e totalmente dalla parte delle biblioteche. Proprio per questo dovrebbe essere un tema centrale nella riflessione biblioteconomica contemporanea.

Riferimenti bibliografici

Albanese, Andrew. 2011a. "At ALA Midwinter, Brewster Kahle, Librarians Ponder The E-book Future." *Publishers Weekly*, 10 January. <https://www.publishersweekly.com/pw/by-topic/industry-news/trade-shows-events/article/45730-at-ala-midwinter-brewster-kahle-librarians-ponder-the-e-book-future.html>.

Albanese, Andrew. 2011b. "Brewster's Millions: ALA Preview 2011." *Publishers weekly*, 30 May. <https://www.publishersweekly.com/pw/by-topic/industry-news/trade-shows-events/article/47448-brewster-s-millions-ala-preview-2011.html>.

Albanese, Andrew. 2014. "Second Circuit Upholds HathiTrust Verdict." *Publishers weekly*, 10 June. <https://www.publishersweekly.com/pw/by-topic/digital/copyright/article/62820-second-circuit-upholds-hathitrust-verdict.html>.

Albanese, Andrew. 2020a. "U.S. Senator Thom Tillis Questions the Internet Archive's 'National Emergency Library'." *Publishers weekly*, 8 April. <https://www.publishersweekly.com/pw/by-topic/digital/copyright/article/82988-u-s-senator-thom-tillis-questions-the-internet-archive-s-national-emergency-library.html>.

Albanese, Andrew. 2020b. "Internet Archive Responds to Senator's Concern Over National Emergency Library." *Publishers weekly*, 14 April. <https://www.publishersweekly.com/pw/by-topic/digital/copyright/article/83027-internet-archive-responds-to-senator-s-concern-over-national-emergency-library.html>.

Albanese, Andrew. 2020c. "Internet Archive to End 'National Emergency Library' Initiative." *Publishers weekly*, 12 June. <https://www.publishersweekly.com/pw/by-topic/digital/copyright/article/83584-internet-archive-to-end-national-emergency-library-initiative.html>.

Albanese, Andrew. 2020d. "Internet Archive to Publishers: Drop 'Needless' Copyright Lawsuit and Work with Us." *Publishers weekly*, 23 July. <https://www.publishersweekly.com/pw/by-topic/industry-news/libraries/article/83929-internet-archive-to-publishers-drop-needless-copyright-lawsuit-and-work-with-us.html>.

Albanese, Andrew. 2021. "Internet Archive Seeking 10 Years of Publisher Sales Data for Its Fair Use Defense." *Publishers weekly*, 9 August. <https://www.publishersweekly.com/pw/by-topic/industry-news/libraries/article/87104-internet-archive-seeking-10-years-of-publisher-sales-data-for-its-fair-use-defense.html>.

Alter, Alexandra. 2020. "'Emergency' Online Library Draws Ire of Some Authors." *The New York Times*, 30 March. <https://www.nytimes.com/2020/03/30/books/internet-archive-emergency-library.html>.

American Library Association. 2020. *Executive Board recommends closing libraries to public*, 17 March. <https://www.ala.org/news/press-releases/2020/03/ala-executive-board-recommends-closing-libraries-public>.

Appeal on Controlled Digital Lending (CDL). 2019. <https://nwu.org/wp-content/uploads/2019/08/CDL-Appeal-15AUG2019-v104.pdf>.

AAP (Association of American Publishers). 2019. *Statement on Flawed Theory of “Controlled Digital Lending”*, 4 February. <https://publishers.org/news/statement-on-flawed-theory-of-controlled-digital-lending/>.

AAP (Association of American Publishers). 2020a. *Comment from AAP President and CEO Maria Pallante on the Internet Archive’s National Emergency Library*, 27 March. <https://publishers.org/news/comment-from-aap-president-and-ceo-maria-pallante-on-the-internet-archives-national-emergency-library/>.

AAP (Association of American Publishers). 2020b. *Publishers File Suit Against Internet Archive for Systematic Mass Scanning and Distribution of Literary Works: Ask Court to Enjoin and Deter Willful Infringement*, 1 June, <https://publishers.org/news/publishers-file-suit-against-internet-archive-for-systematic-mass-scanning-and-distribution-of-literary-works/>.

AAP (Association of American Publishers). 2023a. *Publishers and Internet Archive Submit Negotiated Judgment with Permanent Injunction to District Court in Hachette Book Group, et al, v. Internet Archive. Proposal Follows Decisive Ruling on IA’s Infringing Activities*, 11 August. <https://publishers.org/news/publishers-and-internet-archive-submit-negotiated-judgment-with-permanent-injunction-to-district-court-in-hachette-book-group-et-al-v-internet-archive/>.

AAP (Association of American Publishers). 2023b. *Reflections from the Association of American Publishers on Hachette Book Group v. Internet Archive: An Affirmation of Publishing*, 31 March. <https://publishers.org/reflections-from-the-association-of-american-publishers-on-hachette-book-group-v-internet-archive-an-affirmation-of-publishing>.

Australian Society of Authors. 2019. *Open Library: copyright infringement*, 21 January. <https://www.asauthors.org/news/open-library-copyright-infringement>.

Authors Alliance. 2018. *Authors Alliance supports controlled digital lending by libraries*, 28 September. <https://www.authorsalliance.org/2018/09/28/authors-alliance-supports-controlled-digital-lending-by-libraries>.

Authors Guild. 2018. *An Update on Open Library*, 18 January. <https://authorsguild.org/news/update-open-library/>.

Authors Guild. 2019. *Controlled Digital Lending Is Neither Controlled nor Legal*, 9 January. <https://authorsguild.org/news/controlled-digital-lending-is-neither-controlled-nor-legal/>.

Authors Guild. 2020a. *Internet Archive’s National Emergency Library Harms Authors*, 27 March. <https://authorsguild.org/news/internet-archives-uncontrolled-digital-lending/>.

Authors Guild. 2020b. *Authors Guild Affirms Support for Copyright Infringement Lawsuit Against Internet Archive Brought by Four Leading Book Publishers*, 1 June. <https://authorsguild.org/news/authors-guild-affirms-support-for-copyright-infringement-lawsuit-against-internet-archive-brought-by-four-leading-book-publishers/>.

Belanger Ashley. 2023. “Internet Archive’s legal woes mount as record labels sue for \$400M.” *Ars technical*, 15 August. <https://arstechnica.com/tech-policy/2023/08/record-labels-sue-internet-archive-for-digitizing-obsolete-vintage-records/>.

Board, Brandon, e Karl Stutzman. 2020. "Controlled Digital Lending: An E-book Solution When There Is No E-book?." *Atla Summary of Proceedings*, 74. <https://doi.org/10.31046/proceedings.2020.1857>.

Courtney, Kyle K., e David Hansen. 2018. *White Paper On Controlled Digital Lending Of Library Books*. <https://controldigitallending.org/whitepaper/>.

Courtney, Kyle K Lila Bailey, David Hansen, Mary Minow, Jason Schultz, e Michelle Wu. 2018. *Position Statement On Controlled Digital Lending By Libraries*. <https://controldigitallending.org/statement/>.

Courtney, Kyle K. 2020a. "COVID-19, Copyright, & Library Superpowers." 11 March. <https://kylecourtney.com/2020/03/11/covid-19-copyright-library-superpowers-part-i/> (parte I), <https://kylecourtney.com/2020/03/16/covid-19-copyright-and-library-superpowers-part-ii/> (parte II).

Courtney, Kyle K. 2020b. "Libraries Do Not Need Permission To Lend Books: Fair Use, First Sale, and the Fallacy of Licensing Culture." 18 March. <https://kylecourtney.com/2020/05/18/libraries-do-not-need-permission-fair-use-first-sale-and-the-fallacy-of-permission-culture/>.

De Castell, Christina, Joshua Dickison, Trish Mau, Mark Swartz, Robert Tiessen, Amanda Wakaruk, e Christina Winter. 2022. "Controlled Digital Lending of Library Books in Canada." *The Canadian Journal of Library and Information Practice and Research*, 17 (2). <https://doi.org/10.21083/partnership.v17i2.7100>.

Del Vecchio, Francesco. 2023. "È iniziato il processo contro la più grande biblioteca di internet." *Wired*, 24 marzo. <https://www.wired.it/article/internet-archive-processo-diritto-autore-libri>.

De Robbio, Antonella. 2012, "Il copyright non è per sempre. Il Caso HathiTrust." *Il Bo live*, 17 ottobre. <https://ilbolive.unipd.it/it/content/il-copyright-non-e-sempr-il-caso-hathitrust>.

Dewey, Caitlin. 2013. "Amazon patents resale marketplace for used e-books, songs and other digital goods." *The Washington Post*, 6 February. https://www.washingtonpost.com/business/technology/amazon-patents-resale-marketplace-for-used-e-books-songs-and-other-digital-goods/2013/02/06/22f7d12e-70a0-11e2-a050-b83a7b35c4b5_story.html.

Enis, Matt. 2019. "Better World Libraries, Internet Archive Partner, Acquires Better World Books." *Library Journal*. 2 December. <https://www.libraryjournal.com/story/better-world-libraries-internet-archive-partner-acquires-better-world-books>.

European Bureau of Library, Information and Documentation Associations; Expert Group on Information Law. 2022. *First European Overview on E-lending in Public Libraries. An interim report prepared by EBLIDA EGIL: Country profiles and Summary Tables*, The Hague. <http://www.eblida.org/News/2022/first-european-overview-elending-public-libraries.pdf> (trad. ita. "Prima panoramica europea sul prestito digitale nelle biblioteche pubbliche." 2022. *AIB Studi*, 62 (3): 667-677. <https://doi.org/10.2426/aibstudi-13789>, trad. non integrale).

European Bureau of Library, Information and Documentation Associations; Expert Group on Information Law. 2023. *Handbook of Comparative E-Lending Policies in European Public Libraries*, The Hague. <http://www.eblida.org/news/press-release-handbook-on-comparative-e-lending-policies-in-europe.htm>.

Freeland, Chris. 2019. "For the Love of Literacy—Better World Books and the Internet Archive Unite to Preserve Millions of Books." *Internet Archive Blogs*. 6 November. <https://blog.archive.org/2019/11/06/for-the-love-of-literacy-better-world-books-and-the-internet-archive-unite-to-serve-millions-of-books/>.

Freeland, Chris. 2020a. "Announcing a National Emergency Library to Provide Digitized Books to Students and the Public." *Internet Archive Blogs*. 24 March. <https://blog.archive.org/2020/03/24/announcing-a-national-emergency-library-to-provide-digitized-books-to-students-and-the-public/>.

Freeland, Chris. 2020b. "Internet Archive responds: Why we released the National Emergency Library." *Internet Archive Blogs*. 30 March. <https://blog.archive.org/2020/03/30/internet-archive-responds-why-we-released-the-national-emergency-library/>.

Freeland, Chris. 2020c. "Impacts of the temporary National Emergency Library and controlled digital lending." *Internet Archive Blogs*. 11 June. <https://blog.archive.org/2020/06/11/impacts-of-the-temporary-national-emergency-library>.

Freeland, Chris. 2020d. "More Impacts of the National Emergency Library." *Internet Archive Blogs*. 22 June. <https://blog.archive.org/2020/06/22/more-impacts-of-the-national-emergency-library>.

Freeland, Chris. 2020e. "Even More Impacts of the National Emergency Library and Controlled Digital Lending." *Internet Archive Blogs*. 10 August. <https://blog.archive.org/2020/08/10/even-more-impacts-of-the-national-emergency-library-and-controlled-digital-lending>.

Freeland, Chris. 2023a. "The Fight Continues." *Internet Archive Blogs*. 25 March. <https://blog.archive.org/2023/03/25/the-fight-continues>.

Freeland, Chris. 2023b. "Our Fight is Far From Over." *Internet Archive Blogs*. 11 August. <https://blog.archive.org/2023/08/11/our-fight-is-far-from-over/>.

Freeland, Chris. 2023c. "What the Hachette v. Internet Archive Decision Means for Our Library." *Internet Archive Blogs*. 17 August. <https://blog.archive.org/2023/08/17/what-the-hachette-v-internet-archive-decision-means-for-our-library/>.

Freeland, Chris. 2023d. "Internet Archive Files Appeal in Publishers' Lawsuit Against Libraries." 11 September. <https://blog.archive.org/2023/09/11/internet-archive-files-appeal-in-publishers-lawsuit-against-libraries/>.

Harris, Elizabeth A. 2020. "Publishers Sue Internet Archive Over Free E-Books." *The New York Times*. 1 June. <https://www.nytimes.com/2020/06/01/books/internet-archive-emergency-library-coronavirus.html>.

IFLA (International Federation of Library Associations and Institutions). 2012. *IFLA E-Lending Background Paper*. <https://www.ifla.org/news/ifla-releases-background-paper-on-e-lending/>.

IFLA (International Federation of Library Associations and Institutions). 2021. *IFLA Statement on Controlled Digital Lending*, 2 June. <https://repository.ifla.org/handle/123456789/1835>.

Kahle, Brewster. 2020a. "Internet Archive Staff and Covid-19: Work-at-Home for Most, Full-Pay Furlough & Medical for Scanners." *Internet Archive Blogs*. 25 March. <https://blog.archive.org>.

[org/2020/03/25/internet-archive-staff-and-covid-19-work-at-home-for-most-full-pay-furlough-with-medical-for-scanners](https://www.archive.org/2020/03/25/internet-archive-staff-and-covid-19-work-at-home-for-most-full-pay-furlough-with-medical-for-scanners).

Kahle, Brewster. 2020b. "Temporary National Emergency Library to close 2 weeks early, returning to traditional controlled digital lending." *Internet Archive Blogs*. 10 June. <https://blog.archive.org/2020/06/10/temporary-national-emergency-library-to-close-2-weeks-early-returning-to-traditional-controlled-digital-lending/>.

Lepore, Jill. 2020. "The National Emergency Library Is a Gift to Readers Everywhere." *The New Yorker*. 26 March. <https://www.newyorker.com/books/page-turner/the-national-emergency-library-is-a-gift-to-readers-everywhere>.

Mercanti, Fabio. 2022. *Prestito digitale*. Roma: Associazione italiana biblioteche.

Mercanti, Fabio. 2023. "Controlled digital lending e digital lending: un confronto sulla lettura digitale tramite le biblioteche." *AIB studi*. 62 (2):279-295. <https://doi.org/10.2426/aibstudi-13903>.

Milliot, Jim. 2020. "Publishers Sue the Internet Archive for Copyright Infringement." *Publishers weekly*. 1 June. <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/83472-publishers-charge-the-internet-archive-with-copyright-infringement.html>.

Mount, Dan. 2014. *A Review of Public Library E-Lending Models*, Civic Agenda European Unit.

Mount, Dan. 2016. *Research for CULT committee. E-lending: challenges and opportunities*, Brussels: European parliament.

National Writers Union. 2019. *NWU denounces "Controlled Digital Lending"*. 13 February. <https://nwu.org/nwu-denounces-cdl/>.

Pang, Xuan. 2022. "Controlled digital lending: Past emergency implementation led to future exploration." *Journal of Access Services*. 19 (4). <https://doi.org/10.1080/15367967.2023.2170803>.

Pohl, Adrian. 2012. "Building the Ecology of Libraries: An Interview with Brewster Kahle." *Open Knowledge Foundation*. 23 March. <https://blog.okfn.org/2012/03/23/building-the-ecology-of-libraries-an-interview-with-brewster-kahle>.

Roncaglia, Gino. 2010. *La quarta rivoluzione. Sei lezioni sul futuro del libro*. Roma-Bari: Laterza.

Roncaglia, Gino. 2015. "L'e-lending bibliotecario: alcune note introduttive." *Biblioteche oggi*, 33: 5-7.

Rosen, Judith. 2012. "ReDigi Plans to Sell Used E-books." *Publishers weekly*. 27 July. <https://www.publishersweekly.com/pw/by-topic/digital/retailing/article/53334-redigi-plans-to-sell-used-e-books.html>.

Rosen, Judith. 2013. "Sale of Used E-books Getting Closer." *Publishers weekly*. 16 February. <https://www.publishersweekly.com/pw/by-topic/digital/retailing/article/56004-sale-of-used-e-books-getting-closer.html>.

Schard, Robin. 2021. "Hachette Book Group v. Internet Archive: Is There a Better Way to Restore Balance in Copyright?." *Internet Reference Services Quarterly*, 24 (1-2): 53-58. <https://doi.org/10.1080/10875301.2021.1875100>.

Schwabach, Aaron. 2021. "The Internet Archive's National Emergency Library: is there an emergency fair use superpower?" *Northwestern Journal of Technology and Intellectual*, 18 (2): 187-216. <https://scholarlycommons.law.northwestern.edu/njtip/vol18/iss2/3>.

Sisto Michelle, 2022. "Publishing and library e-lending: an analysis of the decade before Covid-19." *Publishing research quarterly*, 38 (2): 405-422. <https://doi.org/10.1007/s12109-022-09880-7>.

Society of Authors. 2019. *Help us tackle e-book piracy!*. 18 January. <https://www.societyofauthors.org/News/News/2019/Help-us-tackle-e-book-piracy>.

Tondelli, Cal R. 2021. "Mass Digitization and the Consumer Book Market of the Future." *Loyola Consumer Law Review*, 33 (2): 420-440. <https://lawcommons.luc.edu/lclr/vol33/iss2/8>.

Wohlsen, Marcus. 2013. *Amazon Wants to Get Into the Used E-Book Business-Or Bury It*. 8 February. <https://www.wired.com/2013/02/amazon-used-e-book-patent>.

Wu, Michelle. 2011. "Building a Collaborative Digital Collection: A Necessary Evolution in Libraries." *Georgetown Law Faculty Publications and Other Works*. 699: 527-551. <https://scholarship.law.georgetown.edu/facpub/699>.

Wu, Michelle. 2017. "Piece by Piece Review of Digitize-and-Lend Projects Through the Lens of Copyright and Fair Use." *Georgetown Law Faculty Publications and Other Works*. <https://scholarship.law.georgetown.edu/facpub/1974>.

Wu, Michelle. 2019. "Revisiting Controlled Digital Lending Post-ReDigi." *Georgetown Law Faculty Publications and Other Works*, 24 (5). <https://scholarship.law.georgetown.edu/facpub/2134>.

Digital curation and long-term digital preservation in libraries

Valdo Pasqui^(a)

a) retired from University of Florence, <https://orcid.org/0000-0002-4105-2208>

Contact: Valdo Pasqui, valdopasqui@tiscali.it

Received: 22 June 2023; Accepted: 28 October 2023; First Published: 15 January 2024

ABSTRACT

For over twenty years public institutions and foundations committed in cultural heritage curation and preservation worked up methodologies, strategies and techniques for long-term digital preservation of digital resources. Thanks to a relevant number of initiatives and services carried on by international institutions and specialized agencies, public and private libraries reached a high level of awareness in this field which requires continuous adaptation due to the constant evolution of publishing industry and the rapid emergence of Open Science.

After a quick review of digital preservation milestones, this article looks at some emerging trends and issues which characterize the constant transformation of digital preservation landscape. Dynamic generated digital contents and Open Science to fulfil findability, accessibility, interoperability and reuse (FAIR) of digital assets are engaging new challenges for existing digital preservation services. Digital curation and digital preservation are deeply merged and proceed in parallel to assure long-term usage of digital resources and data. How to select and deploy digital preservation platforms is another complex area where Cloud services can offer the opportunity to reduce the costs of technical infrastructures maintenance and management. Policy and decision making institutional levels need more awareness to draw up digital preservation strategies and to activate services to assure digital contents and data long-term access and reuse. Finally the paper suggests to promote “digital preservation by design” as a new base rule to be applied since the conception of every project and initiative in the cultural heritage, scientific and academic contexts.

KEYWORDS

Digital preservation; Digital curation; Digital resources; Digital collections; Open Science; Cloud services.

Digital curation e long-term digital preservation nelle biblioteche

ABSTRACT

Per oltre venti anni le istituzioni e le fondazioni impegnate nella cura e nella conservazione del patrimonio culturale hanno sviluppato metodologie, strategie e tecniche per assicurare la conservazione a lungo termine delle risorse digitali. Grazie a un significativo numero di iniziative e di progetti svolti da istituzioni internazionali e da agenzie specializzate, le biblioteche pubbliche e private hanno acquisito un elevato livello di consapevolezza in questo campo che richiede continui adeguamenti a causa della costante evoluzione dell'editoria e della rapida affermazione della Scienza Aperta. Dopo una rapida rassegna dei capisaldi che caratterizzano la digital preservation, l'articolo rivolge la propria attenzione ad alcune tendenze e tematiche emergenti che contraddistinguono la costante trasformazione della conservazione delle risorse digitali in ambito bibliotecario. I contenuti digitali generati in modo dinamico e la Scienza Aperta volta a perseguire la reperibilità, l'accessibilità, l'interoperabilità e il riuso (principi FAIR) delle risorse digitali e dei dati stanno determinando nuove sfide per i servizi di digital preservation esistenti. La curatela e la conservazione delle risorse digitali sono tra loro profondamente intrecciate e i loro processi procedono in parallelo al fine di assicurare l'utilizzo delle risorse digitali e dei dati nel lungo periodo. Come selezionare e implementare le piattaforme per la conservazione digitale è un'altra area complessa nella quale

i servizi in Cloud possono offrire opportunità per ridurre i costi di gestione e manutenzione delle infrastrutture tecnologiche necessarie.

I livelli istituzionali politici e decisionali devono acquisire una maggiore conoscenza e consapevolezza per elaborare strategie e per attivare servizi che assicurino l'accesso e il riuso a lungo termine dei contenuti digitali e dei dati. Infine l'elaborato suggerisce di promuovere il principio della "digital preservation by design" come nuova regola base da adottare fin dalla fase del concepimento di ogni progetto e iniziativa negli ambiti del patrimonio culturale e della ricerca accademica e scientifica.

PAROLE CHIAVE

Conservazione digitale; Curatela e salvaguardia delle risorse digitali; Risorse digitali; Collezioni digitali; Scienza Aperta; Servizi in Cloud.

1. Introduzione

Le istituzioni preposte alla cura del patrimonio culturale e della conoscenza (archivi, musei, biblioteche accademiche, pubbliche e private) da oltre due decenni sono impegnate nello sviluppo di servizi per la conservazione delle risorse digitali. Nel periodo 1999-2007 i due progetti internazionali InterPARES 1 e InterPARES 2¹ hanno posto le basi concettuali e definito metodologie, strategie e standard per la conservazione permanente di documenti prodotti o mantenuti in forma digitale. La Prof.ssa Luciana Duranti, Direttrice di questa serie dei progetti, ha fornito un'efficace e ancora attualissima sintesi degli obiettivi e della metodologia di ricerca adottata nella prima fase di InterPARES (Duranti 2006) sottolineando tra i principi ispiratori il coinvolgimento di esperti e competenze appartenenti a molteplici aree disciplinari e l'apertura a mutuare da altri ambiti conoscenze per riconciliarle e svilupparle ai fini archivistici. Mentre il primo progetto ha rivolto la propria attenzione agli oggetti nativi digitali prodotti in ambito amministrativo e legale e organizzati in database o sistemi di gestione documentale, il secondo ha ampliato lo sguardo includendo entità dinamiche che possono avere manifestazioni variabili e entità interattive, tenendo conto del comportamento del sistema e delle interazioni da parte dell'utente che possono causare cambiamenti di contenuto e/o forma (es. possibilità di selezionare porzioni del contenuto). Con questi presupposti InterPARES 1 e 2 hanno sviluppato una serie di principi fondanti per la conservazione dei documenti digitali come le definizioni di documento digitale, di ciclo di vita del documento digitale, di affidabilità, di accuratezza (esattezza e correttezza del contenuto del documento la cui responsabilità appartiene all'autore), di autenticità (proprietà del documento formata dalla sua identità certa, costituita dai suoi attributi/metadati, e dalla sua integrità) e di autenticazione (dichiarazione di autenticità aggiunta in conformità a precise norme di legge, p.e. apponendo un firma digitale o un sigillo elettronico) e concetti quali la distinzione tra componenti intellettuali e componenti tecnologiche e la decomposizione del documento digitale nelle sue componenti digitali al momento dell'archiviazione. Significativa l'affermazione che "la conservazione di un documento digitale consiste della conservazione delle sue componenti digitali e delle loro connessioni, in modo tale che, ogni volta che il documento venga richiamato, le sue componenti si ricompongano in una riproduzione identica al documento inizialmente immagazzinato".

In Italia il Legislatore ha rivolto una particolare attenzione alla conservazione dei documenti digitali formati in ambito amministrativo. Il Codice per l'Amministrazione Digitale (CAD), fin dalla sua prima emanazione nel 2005 (Dlgs 7 marzo 2005, n. 82), ha fissato il quadro normativo di riferimento per la conservazione e per l'adozione di sistemi di conservazione capaci di garantire l'autenticità, l'integrità e l'affidabilità dei documenti informatici (cfr. nella versione vigente² gli Art.li 34 e da 40 a 44). Il DPCM 3 dicembre 2013 "Regole tecniche in materia di sistema di conservazione", unitamente al DPCM 13 novembre 2014 "Regole tecniche in materia di formazione, trasmissione, copia, duplicazione, riproduzione e validazione temporale dei documenti informatici", hanno regolato lo sviluppo e la fornitura dei servizi di conservazione fino al settembre 2020 quando l'Agenzia per l'Italia digitale (Agid) ha emesso le "Linee Guida sulla formazione, gestione

¹ L'acronimo sta per "International research on Permanent Authentic Records in Electronic Systems"- L'iniziativa è proseguita con ulteriori due fasi InterPARES 3 (2007-2012) e InterPARES (2012-2018) <http://www.interpares.org/>

² <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2005-03-07:82>.

e conservazione dei documenti informatici”³. A seguire Agid ha emesso anche il “Regolamento sui criteri per la fornitura dei servizi di conservazione dei documenti informatici” (Determinazione Agid n. 455/2021 del 25 giugno 2021)⁴ i cui allegati (A e B) formulano i requisiti generali, di qualità, di sicurezza e organizzazione necessari per la fornitura di un servizio di conservazione e le indicazioni per il piano di cessazione del servizio⁵. Grazie anche a questa sequenza di normative nel Paese si è sviluppata un’ampia offerta di servizi di conservazione gestiti da aziende private e sono stati sviluppate soluzioni anche da parte di Pubbliche Amministrazioni tra le quali a titolo esemplificativo si ricordano il Polo archivistico dell’Emilia-Romagna (ParER)⁶, il DAX-Servizio regionale di conservazione a norma della Regione Toscana⁷ e il sistema documentale eDocumento dell’Università di Napoli Federico II⁸.

In considerazione della molteplicità e diversità degli ambiti che si avvalgono dei servizi di digital preservation, dell’opportunità sorta informalmente in occasione di un evento seminariale di approfondire alcune tematiche relative alla conservazione digitale nell’ambito delle biblioteche e dei limiti di spazio a disposizione stabiliti dalla rivista l’articolo si pone l’obiettivo di esaminare la digital preservation nell’attuale contesto in trasformazione delle biblioteche. Tale scelta non è motivata da una visione settoriale ma trae origine dalla percezione che sia opportuno porre un’attenzione specifica sulle biblioteche le quali al fine di svolgere la propria missione sono al tempo stesso soggetti produttori di risorse digitali (es. le collezioni digitalizzate) e consumatori di risorse digitali prodotte e possedute da terze parti e che, grazie allo sviluppo di nuove forme di editoria e all’accentuata rilevanza dei dati scientifici derivante dai principi della scienza aperta, sono chiamate ad affrontare nuove opportunità e sfide relative alla conservazione nel lungo tempo delle risorse digitali.

Le biblioteche hanno maturato da molti anni la consapevolezza dell’importanza della digital preservation per garantire l’accesso nel tempo alle proprie collezioni digitali e alle riviste elettroniche. Le campagne di digitalizzazione di opere in formato analogico (es. manoscritti, incunaboli, testi rari, foto etc.) hanno prodotto collezioni digitali che evitano la manipolazione degli originali, garantiscono l’esistenza di copie digitali da consultare in caso di deterioramento o di perdita totale dei contenuti originali e consentono la diffusione del patrimonio culturale nazionale presso un pubblico più ampio rispetto a quello specialistico⁹. Il salvataggio dei siti web, dei file e delle basi dati che ospitano queste collezioni non garantiscono la conservazione nel lungo periodo e non

³ https://www.agid.gov.it/sites/default/files/repository_files/linee_guida_sul_documento_informatico.pdf.

⁴ <https://www.agid.gov.it/it/piattaforme/conservazione>.

⁵ Con successiva Determinazione Agid n. 629/2021 è stato modificato il testo del Regolamento per introdurre il Marketplace dei servizi di conservazione nel quale le PA possono individuare i servizi di conservazione e avviare la fase contrattuale, evitando l’obbligo di trasmettere entro trenta giorni dalla stipula i contratti per consentire all’Agenzia di verificare il soddisfacimento dei requisiti generali nonché dei requisiti di qualità, di sicurezza e organizzazione previsti dall’allegato A del Regolamento: https://conservatoriqualeficati.agid.gov.it/?page_id=276.

⁶ <https://poloarchivistico.regione.emilia-romagna.it/conservazione/conservazione-parer>.

⁷ <https://www.regione.toscana.it/-/dax-infrastruttura-per-la-conservazione-a-norma-dei-documenti-informatici>.

⁸ <http://www.praxis.unina.it/documentale>.

⁹ Vedere le gare avviate dal “Piano nazionale di digitalizzazione” (PND) del patrimonio culturale, nell’ambito del PNRR, “Misura 1 – Patrimonio culturale per la prossima generazione Componente 3 - Turismo e Cultura 4.0” <https://pnrr.cultura.gov.it/misura-1-patrimonio-culturale-per-la-prossima-generazione/1-1-piattaforme-e-strategie-digitali-per-laccesso-al-patrimonio-culturale/>.

prevedono i rischi cui sono esposti gli oggetti digitali: a) l'obsolescenza dei formati; b) il rapido cambiamento delle tecnologie; c) la vulnerabilità delle componenti infrastrutturali (repository, server, sistemi di archiviazione, software etc.) esposte agli attacchi di pirateria informatica volti ad alterare, distruggere o rendere inutilizzabili i contenuti digitali.

Per le riviste elettroniche si aggiunge il rischio di non poter accedere agli articoli a causa della cessazione degli abbonamenti, della prolungata indisponibilità dei siti degli editori, della discontinuità della rivista o dell'editore stesso. Un gruppo di esperti appartenenti a università, biblioteche e centri di ricerca americani già nel 2005 aveva elaborato un manifesto (Waters 2005) che sottolineava l'urgente necessità di avviare piani di conservazione delle riviste elettroniche attraverso la creazione di "trusted archives in which the published scholarly record in electronic form can persist outside of the exclusive control of publishers, and in the control of entities that value long-term persistence".

Anche i repository istituzionali ad accesso aperto, attivati per sostenere le politiche di Open Access, e le piattaforme digitali open source specializzate per le riviste devono confrontarsi con la necessità di garantire l'accesso nel tempo ai contenuti digitali e di proteggerli dai rischi citati¹⁰. La diffusione dell'editoria elettronica ha ulteriormente ampliato lo spettro di opere la cui unica manifestazione (secondo il modello FRBR) è nativa digitale: e-book, audio libri, podcast, video, pubblicazioni che includono dati e risorse digitali richiamati dinamicamente da sorgenti esterne. Il panorama sta subendo un'ulteriore evoluzione grazie alle iniziative che promuovono la scienza aperta e all'attuazione dei principi FAIR¹¹ (Wilkinson et al. 2016) le quali per consentire il riuso dei dati necessitano di programmi di conservazione che oltre ai dati includano anche i flussi di lavoro, i software utilizzati e tutte le componenti dell'attività di ricerca.

Siamo dunque di fronte a uno scenario in rapida trasformazione nel quale si aprono nuove sfide e nuove opportunità e cresce l'urgenza per le istituzioni culturali, accademiche e di ricerca di acquisire consapevolezza che la cura delle collezioni digitali, delle pubblicazioni e dei dati deve procedere in parallelo con la definizione di strategie e l'attivazione di servizi per la loro conservazione nel tempo.

2. Lo scenario consolidato

I numerosi progetti e i programmi sviluppati negli ultimi decenni hanno consolidato la digital preservation conferendole un elevato grado di maturità che si fonda su standard, strumenti, flussi di lavoro, istituzioni, progetti, iniziative e convegni.

La Digital Preservation Coalition (DPC) è una delle fonti più autorevoli per terminologia, metodologie, buone pratiche e linee guida¹² e il suo Digital Preservation Handbook¹³ definisce lo scopo e l'essenza della digital preservation affermando che: "refers to the series of managed activities

¹⁰ JLIS.it usa la piattaforma OJS e partecipa ad un programma di digital preservation; lo stato di conservazione dei numeri pubblicati sono consultabili sul Keepers Registry <https://keepers.issn.org/>.

¹¹ <https://www.go-fair.org/fair-principles/>.

¹² <https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports>.

¹³ <https://www.dpconline.org/handbook>. La traduzione italiana è stata realizzata da Assunta Caruso e Claudia Lanza, con la premessa di Roberto Guarasci.

necessary to ensure continued access to digital materials for as long as necessary [...] beyond the limits of media failure or technological and organisational change”.

Il modello di riferimento condiviso dai sistemi di conservazione è il Reference Model for an Open Archival Information System (OAIS) (CCSDS 2012) sviluppato e consolidato dalle agenzie facenti capo al Consultative Committee for Space Data Systems (CCSDS) per le esigenze di archiviazione e conservazione delle immagini e dei dati relativi alle osservazioni spaziali. OAIS, Standard ISO 14721:2012I, ha fissato alcune pietre miliari della digital preservation:

- il modello funzionale, i flussi di lavoro, le attività, i ruoli in cui si articolano i processi di conservazione (cfr. Fig. 1), i pacchetti di versamento (SIP), archiviazione (AIP) e disseminazione (DIP) composti dagli oggetti digitali e dai metadati;
- le tipologie dei metadati che devono accompagnare gli oggetti digitali nei flussi di conservazione, suddivisi nelle due macro categorie Preservation Description Information (reference, provenance, context, fixity, access rights) e Representation Information (structure e semantic);
- il trattamento dei formati degli oggetti conservati per garantirne la fruibilità nel tempo in caso di obsolescenza mediante tecniche di trasformazione o avvalendosi dell'emulazione (virtualizzazione) dei programmi necessari alla loro interpretazione;
- il concetto di Designated Community che identifica uno o più gruppi/comunità di potenziali fruitori (consumer) interessati a reperire e accedere ai contenuti conservati.

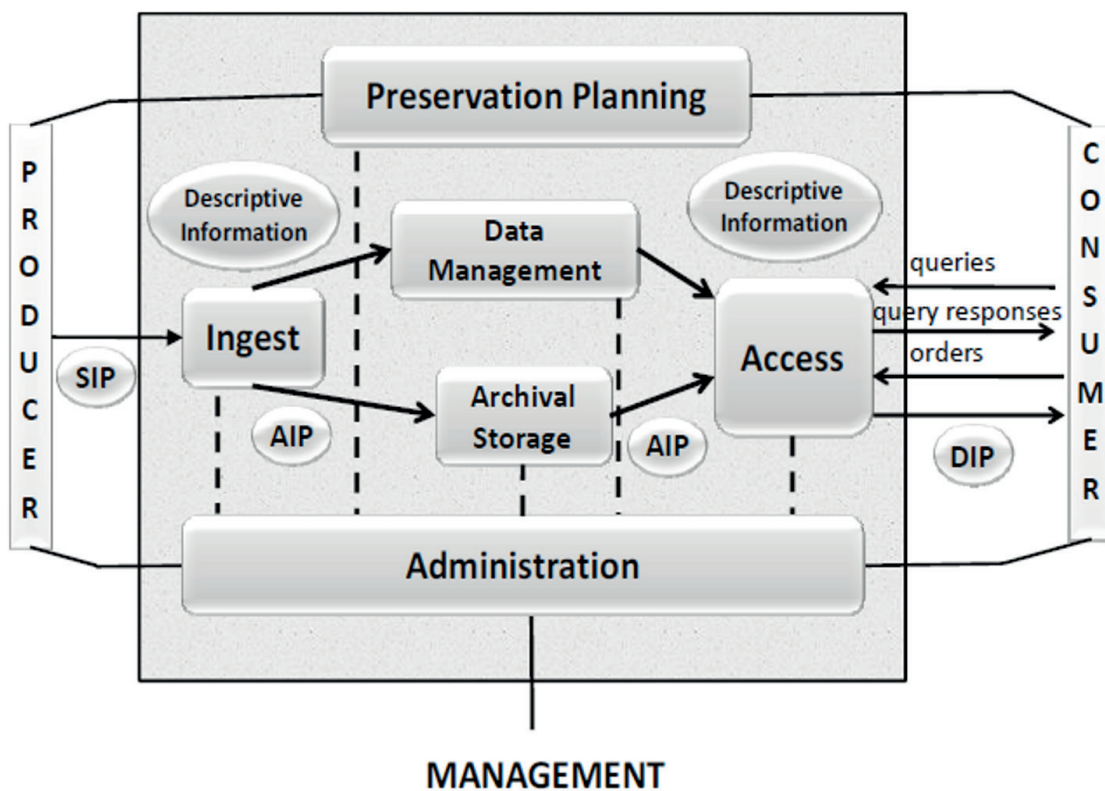


Fig. 1. Modello funzionale di OAIS.

I metadati previsti da OAIS sono formalizzati dal Data Dictionary for Preservation Metadata PREMIS¹⁴ mantenuto dalla Library of Congress che rimanda ad altri set di metadati standard (es. MARC, MODS, Dublin Core Metadata Element Set etc.).

Lo schema METS Metadata Encoding & Transmission Standard¹⁵ è lo standard per la codifica dei metadati descrittivi, amministrativi e strutturali¹⁶, espresso tramite XML schema language del W3C. Il BagIt File Packaging Format¹⁷ definisce come strutturare in modo gerarchico i file ai fini della loro memorizzazione e del trasferimento dei contenuti digitali.

Il Format Index della Library of Congress¹⁸ contiene le specifiche dei formati ritenuti correntemente standard per le varie tipologie di opere e il registro PRONOM¹⁹ classifica e descrive i formati di file usati per codificare gli oggetti digitali. Per l'identificazione, il riconoscimento e la validazione dei formati dei file sono disponibili vari strumenti open source:

- la Open Preservation Foundation sviluppa strumenti e librerie come JHOVE, VeraPDF, jpylyzer, fido;²⁰
- The National Archives ha realizzato il software DROID (Digital Record Object Identification)²¹ che permette di effettuare il riconoscimento dei formati dei file anche in modo massivo;
- l'International Image Interoperability Framework (IIIF)²² sviluppa un insieme di standard aperti per la distribuzione e la visualizzazione di oggetti digitali di elevata qualità (audio e video) attraverso una serie di API e di visualizzatori (Mirador, Universal Viewer etc.).

Agli standard e agli strumenti si affianca anche l'area delle metodologie per la verifica e la valutazione dell'affidabilità dei processi e dei repository. Il Center for Research Libraries (CLIR)²³ fin dal 2006 ha avviato un processo di certificazione di alcuni repository digitali tra cui CLOCKSS, Hathitrust e Portico. Nel 2007 il Center for Research Libraries (CRL), il National Archives and Records Administration Digital Repository e OCLC hanno prodotto la "Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist" (CRL e OCLC 2007) che valuta il livello di sicurezza e affidabilità dei repository sulla base di un insieme di criteri di riferimento organizzati in tre dimensioni: l'infrastruttura organizzativa, la gestione degli oggetti digitali, l'infrastruttura tecnologica e la sicurezza. Il CCSDS ha emesso una raccomandazione (CCSDS 2011), standard ISO 16363:2012, che elenca e descrive i requisiti e le pratiche cui attenersi per garantire l'affidabilità dei repository.

Il CoreTrustSeal è una certificazione rilasciata da una comunità internazionale, non governativa e no profit, sulla base di 15 requisiti,²⁴ aggiornati periodicamente e suddivisi nelle stesse tre cate-

¹⁴ <https://www.loc.gov/standards/premis/>.

¹⁵ <https://www.loc.gov/standards/mets/mets-home.html>.

¹⁶ <https://www.loc.gov/standards/mets/METSita.html> traduzione italiana a cura di Angela Di Iorio.

¹⁷ <https://github.com/jkunze/bagitspec> e IETF RFC8493 <https://datatracker.ietf.org/doc/html/rfc8493>.

¹⁸ <https://www.loc.gov/preservation/resources/rfs/index.html>.

¹⁹ <https://www.nationalarchives.gov.uk/PRONOM/>.

²⁰ <https://openpreservation.org/tools>.

²¹ <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>.

²² <https://iiif.io/>.

²³ <https://www.crl.edu/archiving-preservation/digital-archives/certification-assessment>.

²⁴ CoreTrustSeal Requirements 2023-2025 https://zenodo.org/record/7051012#.ZdhkI_ZbxEY.

rie di TRAC; dopo l'esame del Board l'esito della certificazione e le osservazioni dei revisori sono pubblicate in forma anonima nella lista dei repository/servizi certificati.²⁵

L'insieme di questi standard, metodologie, tecniche e pratiche forniscono un ampio insieme di punti di riferimento imprescindibili per chiunque voglia avviare nuovi programmi o progetti di conservazione delle risorse digitali.

3. Sfide, criticità ed evoluzioni

Nonostante questo quadro consolidato la digital preservation è caratterizzata da una grande dinamicità grazie alle interdipendenze con la filiera dell'editoria elettronica, con la ricerca e le istituzioni accademiche, oltre che con le tecnologie informatiche di cui si avvale. La continua trasformazione è riscontrabile nei ricorrenti appuntamenti internazionali come le conferenze iPRES²⁶ che fin da Pechino 2004 offrono alle varie comunità coinvolte nella digital preservation la possibilità di confrontarsi su strategie, metodologie ed evoluzioni tecnologiche. La National Digital Stewardship Alliance (NDSA)²⁷ e la Digital Library Federation (DLF)²⁸ dal 2011 organizzano annualmente convegni,²⁹ mentre la DPC e la Library of Congress promuovono frequenti eventi come il recente convegno dedicato alle tecnologie di storage.³⁰ Queste iniziative hanno prodotto una copiosa messe di articoli e ricerche tra i quali spiccano i rapporti di due indagini. Il primo (Rieger et al. 2022) fornisce valutazioni sulla efficacia e la longevità dei sistemi di digital preservation e curation, tratte da interviste con istituzioni e comunità che se ne avvalgono. Il secondo è lo studio della Open Preservation Foundation (OPF 2020) che riassume le risultanze dell'indagine condotta nel biennio 2019-2020 con 98 organizzazioni di 31 paesi diversi, di cui il 51% in Europa (non è compresa l'Italia) attraverso quesiti sulla numerosità e ruoli dello staff, tipologia e dimensione delle collezioni, politiche e risorse finanziarie impegnate, utilizzo di soluzioni e strumenti open source. Parte degli spunti offerti da questi rapporti sono ripresi nel seguito di questo paragrafo in cui sono esaminate quattro tematiche di attualità per la digital preservation dal punto di vista delle biblioteche.

3.1 La gestione dei contenuti digitali complessi

Gli articoli pubblicati online sulle principali riviste seguono schemi editoriali che impongono una struttura abbastanza rigida, prevedono una modalità di lettura lineare e adottano formati standard (EPUB, HTML, PDF/A, PDF). Tuttavia in ambito accademico si stanno sempre più diffondendo forme di pubblicazione che:

- includono audio, video, immagini ad alta risoluzione, dati, mappe e visualizzazioni generate dinamicamente mediante interfacce programmatiche (API) e web-services esposti da

²⁵ <https://amt.coretrustseal.org/certificates>.

²⁶ International Conference on Digital Preservation (iPRES) <https://ipres-conference.org/>.

²⁷ Consorzio di 276 organizzazioni impegnate nella gestione e nella conservazione a lungo termine delle risorse digitali.

²⁸ <https://www.diglib.org/>.

²⁹ Digital Preservation Conferences <https://ndsa.org/conference/digital-preservation/past/>.

³⁰ Designing Storage Architectures for Digital Collections. March 27-28, 2023 Library of Congress, Washington, DC <https://digitalpreservation.gov/meetings/storage23.html>.

applicazioni esterne (es. YouTube, Google Maps, Amazon WS, sistemi informativi geografici, software di elaborazione statistica etc.);

- consentono al fruitore percorsi di lettura non lineari e interattivi non sempre replicabili nelle copie conferite ai servizi di conservazione digitale con le attuali modalità (es. harvesting, acquisizione di pacchetti di archiviazione).³¹

Due esempi di queste tipologie di pubblicazioni sono lo studio sulla necropoli egiziana di Saqqara pubblicato dalla Stanford University Press (Sullivan 2020) e un lavoro della University of Minnesota Press sui video giochi (Boluk e Le Mieux 2017), entrambe citabili tramite un identificatore DOI. Il primo include alcune ricostruzioni dinamiche tridimensionali del sito archeologico. Il secondo consente di scaricare software dei video giochi descritti in apposte schede e offre agli utenti registrati la possibilità di aggiungere note e commenti per arricchire il testo.

Le implicazioni sugli attuali servizi di conservazione delle riviste elettroniche di queste nuove modalità di pubblicazione sono state oggetto di uno studio da parte di un gruppo di università e University Press (New York University Libraries, NYU Press, Michigan Publishing, the University of Minnesota Press, UBC Press and Stanford University Press) e servizi di digital preservation (CLOCKSS, Portico, the libraries of the University of Michigan and NYU) che ha elaborato un insieme di 68 linee guida rivolte agli editori e agli autori (Greenberg et al. 2021). L'impostazione teorico-metodologica del lavoro esamina le risorse digitali secondo due livelli che pongono in evidenza la stretta relazione esistente tra digital preservation e digital curation:

- I. *Publication resources* – il livello tecnico nel quale sono individuati i contenuti digitali veri e propri della pubblicazione, incluse le “risorse esterne” (video, audio, mappe etc.) che sono richiamate tramite API;
- II. *Core intellectual components* – il livello astratto di un'opera che individua le componenti necessarie per la sua comprensione, il punto di partenza per selezionare le componenti di primo livello indispensabili affinché il pacchetto predisposto per la conservazione possa restituire in modo attendibile il contenuto dell'opera e consentire al lettore di percepire il contributo “intellettuale e retorico” degli autori, “prendendo atto che la sua forma corrente può non essere disponibile in futuro a causa dei cambiamenti tecnologici e dei contesti sociali” (linea guida n. 10).

Tra le numerose indicazioni operative viene raccomandato di assegnare identificatori unici e persistenti (Handle, DOI o ARK ID) ad ogni risorsa esterna, di verificare che i diritti d'uso comprendano anche la conservazione, di curare la creazione di metadati che descrivano le relazioni e i diritti. È consigliato il formato EPUB3 per impacchettare le pubblicazioni che includono video e audio oppure di ricorrere all'archiviazione locale di tali oggetti. Per le pubblicazioni contenenti codice JavaScript, mappe dinamiche, visualizzatori di immagini IIIF o che recuperano dati esterni si suggerisce di creare una versione alternativa che pre-carichi tutte le risorse esterne (compresi codice, dati e documentazione) e associ una videoregistrazione del comportamento delle interazioni nel contesto originale. Il documento propone soluzioni volte ad armonizzare i sistemi di digital preservation esistenti e le nuove modalità di pubblicazione e di formazione dei contenuti digitali e promuove una stretta collaborazione tra editori, autori e servizi di conservazione.

³¹ Per esempio perché le interfacce programmatiche e i dati delle fonti esterne sono rilasciati secondo modalità tecniche o sono soggetti a diritti d'uso che limitano la capacità dei servizi di digital preservation di avvalersene per riprodurre il contenuto e il contesto originali.

3.2 Le sfide della scienza aperta

I principi FAIR per il reperimento, l'accessibilità, l'interoperabilità e il riuso dei dati prodotti nell'ambito della ricerca hanno dato un impulso significativo alla Scienza Aperta (Open Science). Il CERN ha sviluppato Zenodo³² che adotta i paradigmi di accesso aperto, dati aperti e software aperto per consentire ai ricercatori di curare, condividere e pubblicare dati e software prodotti nella loro attività scientifica. La Scienza Aperta necessita di programmi di digital preservation³³ e la collaborazione tra i due contesti è agevolata dal fatto che gli obiettivi delle agenzie e delle istituzioni che erogano servizi per la conservazione delle risorse digitali sono ampiamente allineati con i principi FAIR. In questo scenario le biblioteche possono dare un contributo significativo grazie all'esperienza maturata nella conservazione delle collezioni digitali e alla collaborazione fornita in ambito accademico ai ricercatori per la pubblicazione ad accesso aperto e l'alimentazione dei repository istituzionali.

Lo European Open Science Cloud (EOSC),³⁴ il progetto pan-europeo che intende creare un ambiente virtuale per la condivisione, l'elaborazione e l'accesso ai dati della ricerca, nel 2020 ha commissionato al DPC uno studio (Currie e Kilbride 2021) nel quale sono stati analizzati ruoli, responsabilità e rischi della long term data preservation nell'ambito della scienza aperta. Il rapporto finale ha posto in evidenza che nella visione EOSC la digital preservation non è esplicita, devono essere chiariti ruoli, responsabilità e compiti di tutti gli attori coinvolti, occorre definire strategie e adottare soluzioni di digital preservation per affrontare i rischi relativi alla fruizione dei dati. Nel contesto dei dati scientifici il focus si sposta dalle pubblicazioni ai dati raccolti, elaborati e prodotti e anche a tutto quanto è necessario per riutilizzarli e riprodurli, emergono così alcuni obiettivi sfidanti per la digital/data preservation:

- la maggior parte dei dati dell'attività scientifica non sono pubblicati e non sono corredati dai necessari metadati per consentire nuove analisi e elaborazioni;
- la mole dei dati coinvolti implica la necessità di sviluppare strumenti automatici per la generazione e il mantenimento dei metadati;
- i piani di gestione dei dati dei progetti, quando previsti, sono carenti nel garantire l'affidabilità e poco efficaci ai fini della digital preservation, intesa come tematica da affrontare solo alla fine dei progetti;
- oltre ai dati (data sets) occorre conservare la corrispondenza, i workflow di processo, i software (Benner et al. 2022), le applicazioni, le librerie, gli strumenti di analisi, gli ambienti di esecuzione e i sistemi operativi, la documentazione e le eventuali dipendenze da servizi esterni o da altri componenti necessari per riprodurre e riutilizzare i dati per ricreare il contesto dell'attività di ricerca e i casi d'uso che hanno generato i dati;
- l'accesso ai dati da parte degli utilizzatori richiede un'elevata larghezza di banda della rete e le infrastrutture hardware, in particolare le componenti di memorizzazione, devono ga-

³² <https://zenodo.org/> è stato attivato l'8 maggio 2013, dunque da poco ha compiuto dieci anni di vita.

³³ In ambito europeo l'importanza della "conservazione e riutilizzo dell'informazione scientifica" è stata esplicitata da due Raccomandazioni della Commissione UE, la 417 del 17 luglio 2012 e la 790 del 25 aprile 2018: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:194:0039:0043:IT:PDF> e <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32018H0790&from=EN>.

³⁴ <https://eosc-portal.eu/>.

rantire in modo flessibile la scalabilità per consentire l'archiviazione e il trasferimento di dati la cui dimensione tende a crescere in modo continuo e rilevante;

- l'esistenza di dati personali o dati che se opportunamente analizzati ed elaborati possono ricondurre all'identità di persone implicano la necessità di garantire modalità di trattamento che tutelino la riservatezza e la sicurezza in conformità alle normative vigenti nei paesi coinvolti.

Il rapporto prodotto dal DPC propone dodici raccomandazioni, alcune di tipo tecnico (es. l'uso di identificatori persistenti, l'adozione di standard dei metadati come PREMIS, l'integrazione con l'infrastruttura di autenticazione e autorizzazione federata), altre di tipo organizzativo e politico come l'invito a esplicitare la digital preservation nella direzione e nel controllo dei progetti, ad approfondire i costi attraverso l'elaborazione di opportuni modelli, a includere la digital preservation nei programmi di formazione e aggiornamento delle competenze per sviluppare "the next generation of Open Science and data professionals". Gran parte di queste risultanze sono state le premesse del progetto ARCHIVER³⁵ che ha selezionato servizi di long-term preservation in ambito EOSC, secondo una metodologia di Pre-Commercial Procurement attuata da un gruppo d'acquisto guidato dal CERN di Ginevra³⁶ che ha promosso la creazione di consorzi ai quali è stato proposto un modello architetturale articolato nei quattro livelli (Storage/Base Archiving/Secure backup; Preservation; Baseline user services; Advanced services) di cui solo i primi tre obbligatori. Grazie a una rigorosa metodologia di confronto su 15 consorzi candidati nella fase di progettazione ne sono stati selezionati solo 5 che nella seguente fase di sviluppo del prototipo si sono ridotti a 2 per intraprendere la fase pilota, Arkivum e LIBNOVA LABDRIVE³⁷, ora referenziati nel Marketplace European Open Science Cloud (EOSC)³⁸.

In Italia il Ministero dell'Università e della Ricerca (MUR) nell'ambito del Programma Nazionale per la Ricerca (2021-2027)³⁹ con il DM n. 268 del 28-02-2022 ha adottato il Piano Nazionale per la Scienza Aperta che tra le proprie azioni contempla anche "investire in pratiche di cura dei dati finalizzate alla conservazione sul lungo periodo".⁴⁰

Di recente il Consiglio Nazionale delle Ricerche (CNR) ha approvato la roadmap per la scienza aperta (Castelli et al. 2023) nella quale sono previsti l'adozione di una policy "per la gestione e la conservazione della produzione scientifica istituzionale" e la realizzazione di un ecosistema di repository affidabili "per il deposito, la condivisione e la conservazione dei prodotti della ricerca CNR" i cui primi componenti saranno: il CNR Open Access Archive/Repository per le pubblicazioni e i prodotti della letteratura scientifica, il CNR Data Archive/Repository per la conservazione dei dati e il CNR Software Archive/Repository per la conservazione del software prodotti da personale CNR.

³⁵ ARCHIVER Project <https://archiver-project.eu/>.

³⁶ Gli altri componenti erano EMBL's European Bioinformatics Institute, Deutsches Elektronen-Synchrotron (DESY) e Port d'Informació Científica (PIC) (Spagna).

³⁷ ARCHIVER Long-Term Data Preservation Solutions <https://archiver-project.eu/archiver-long-term-data-preservation-solutions>.

³⁸ <https://search.marketplace.eosc-portal.eu/search/service?q=Digital%20preservation>.

³⁹ <https://www.mur.gov.it/it/aree-tematiche/ricerca/programmazione/programma-nazionale-la-ricerca>.

⁴⁰ <https://www.mur.gov.it/it/atti-e-normativa/decreto-ministeriale-n-268-del-28-02-2022>.

3.3 Criticità e criteri per la scelta delle soluzioni

La scelta delle soluzioni di digital preservation deve essere guidata dall'analisi della tipologia, quantità, frequenza di produzione, ritmo di accrescimento e durata della conservazione delle risorse digitali, dalle modalità di accesso da parte della comunità degli utilizzatori, dal contesto organizzativo (persone, competenze etc.) e dalle risorse economiche disponibili. Occorre evitare l'errore di pensare che esistano soluzioni "chiavi in mano" in grado di gestire qualunque esigenza. Spesso è opportuno diversificare le soluzioni avvalendosi di più servizi o prodotti, scelti in funzione delle risorse digitali da conservare e del bacino di utenza. Per esempio un ateneo può avvalersi dei servizi specializzati (Portico, CLOCKS o LOCKSS Network) per le riviste elettroniche per le quali ha sottoscritto abbonamenti e per le pubblicazioni della propria university press e scegliere altre soluzioni per le collezioni digitali, per il repository istituzionale ad accesso aperto e per i dati della ricerca.

Un approccio per valutare e confrontare soluzioni diverse è proposto dalla matrice NDSA⁴¹ che sulle righe riporta cinque aree funzionali (memorizzazione, integrità, controllo, metadati, contenuto) e sulle colonne quattro diversi livelli di digital preservation (Conoscere, Proteggere, Controllare e Sostenere). Ciascuna cella della matrice contiene asserzioni sulle attività da compiere per soddisfare il relativo livello di conservazione. Un'altra metodologia sistematica è quella sviluppata dal progetto ARCHIVER (Fernandes et al. 2020) che ha individuato 47 use-case atomici e per ciascuno di essi ha prodotto la stima del valore aggiunto in caso di soddisfacimento, la misura della complessità e dell'impegno o rischio indotti dall'implementazione.

L'analisi dei criteri per il confronto tra soluzioni (sia no profit che commerciali) richiederebbe uno spazio molto più ampio di quello consentito pertanto si riportano solo alcuni requisiti imprescindibili:

- la definizione puntuale dei ruoli e delle responsabilità, cercando di evitare la completa esternalizzazione del controllo dei processi di digital preservation;
- la messa a punto di un Service Level Agreement (SLA) che dettagli in modo completo e trasparente i livelli dei servizi erogati;
- la verifica dell'esistenza di comunità di utilizzatori vitali e proattive in grado indirizzare gli investimenti e lo sviluppo del prodotto/servizio e di incidere nel medio e lungo periodo sulla sua pianificazione e sulle sue evoluzioni;
- il riscontro delle modalità di "uscita" previste per limitare i rischi di dipendenza dal fornitore o dal servizio, misurando la complessità del recupero e della migrazione degli oggetti digitali e dei metadati e della riattivazione dei processi di conservazione con un nuovo servizio.

Un ulteriore ambito di valutazione riguarda la modalità di installazione o erogazione (deployment) "in house" o "in hosting" o "in Cloud". Ospitare e gestire una piattaforma di digital preservation comporta la disponibilità di un'infrastruttura tecnologica complessa (server, sistemi di storage, apparati di rete, software, dispositivi di backup, servizi e strategie di cybersicurezza etc.) e di uno staff tecnico-informatico che richiedono continui adeguamenti e aggiornamenti e di conseguenti investimenti economici. Il Cloud è un insieme di tecnologie abilitanti per la realizzazione di

⁴¹ <https://ndsa.org/publications/levels-of-digital-preservation/>

piattaforme di digital preservation la cui adozione è prevista dal Piano triennale per l'informatica nella Pubblica Amministrazione⁴² che prescrive la dismissione dei data center che non soddisfano determinati requisiti (classificati nel gruppo A) e il passaggio in Cloud avvalendosi o del Polo Strategico Nazionale (PSN)⁴³ o dei servizi in Cloud qualificati⁴⁴ dall'Agenzia per la Cybersicurezza Nazionale (ACN)⁴⁵.

Le tipiche declinazioni del Cloud per l'implementazione delle soluzioni di digital preservation sono:

- servizi di tipo di IaaS (Infrastructure as a Service) su cui installare l'intero insieme dei componenti software della piattaforma di digital preservation (open source o proprietaria), che tuttavia richiedono la disponibilità di risorse umane con competenze sistemiche e gestionali;
- servizi di tipo di IaaS usati solo per la componente di memorizzazione avvalendosi di Cloud Storage erogato da un fornitore terzo (es. Amazon Web Services S3, Google Cloud Platform, Microsoft Azure a livello mondiale o altri a livello nazionale), anche in questo scenario sono richieste competenze tecniche;
- servizi di digital preservation erogati in modalità SaaS (Software as a Service), in questo caso occorre verificare l'esistenza di strumenti di monitoraggio dei flussi e dei processi di conservazione, il rispetto dei criteri di affidabilità e l'esistenza di eventuali certificazioni (trust service).

L'adozione di soluzioni Cloud, oltre ad un'accurata valutazione dei costi di esercizio, impone anche di verificare che la dislocazione geografica delle infrastrutture di memorizzazione degli oggetti digitali e dei dati sia conforme alle normative vigenti e implica una maggiore complessità delle modalità di uscita quando la soluzione adottata si avvale di uno o più fornitori Cloud terza parte. Ricorrere ai servizi Cloud consente di razionalizzare l'adozione di soluzioni di digital preservation e rappresenta un punto fermo per la costruzione di un'eventuale infrastruttura nazionale.

3.4 Sensibilizzare e rendere consapevoli le istituzioni

L'espansione dell'editoria digitale, la trasformazione delle modalità di pubblicazione in ambito accademico e scientifico, le nuove istanze indotte dalla scienza aperta determinano un quadro multiforme e in continua evoluzione nel quale si potrebbero includere altre tipologie di risorse digitali usate in questi contesti come i blog, i social e i siti web di tipo informativo e divulgativo. Cresce pertanto la necessità di sensibilizzare le istituzioni e soprattutto i livelli politici e decisionali verso l'adozione di programmi di digital preservation dei contenuti nativi digitali. Sensibilizzare significa anzitutto rimuovere l'errata convinzione (o speranza) che garantire l'accesso e il mantenimento nel tempo dei contenuti digitali e dei dati prodotti sia "compito di altri". Dai rapporti e dagli studi citati in precedenza emerge che il livello di consapevolezza e di coinvolgimento delle istituzioni che si occupano del patrimonio culturale e della ricerca scientifica non sono ancora soddisfacenti. Uno studio del

⁴² https://www.agid.gov.it/sites/default/files/repository_files/piano_triennale_per_linformatica_nella_pa_2022-2024.pdf.

⁴³ <https://www.polostrategiconazionale.it/>.

⁴⁴ Cloud Marketplace <https://catalogocloud.acn.gov.it/>.

⁴⁵ Vedere anche le Determinazioni ACN n. 306/2022 e n. 307/2022 e relativi allegati <https://www.acn.gov.it/notizie/contenuti/qualificazione-del-cloud-della-pa-al-via-gestione-acn>.

2016 (Regan et al. 2016) rilevava che solo il 30% delle riviste elettroniche aveva attuato strategie di digital preservation⁴⁶. La ricerca della Open Preservation Foundation mostra che quasi il 19,5% delle organizzazioni intervistate non ha elaborato o pianificato una policy di digital preservation.

Per quanto riguarda il contesto italiano, a differenza di altri Paesi, non esiste un'agenzia o un servizio nazionale per la digital preservation. Per le riviste elettroniche circa 35 istituzioni, in prevalenza università, hanno aderito a Portico⁴⁷ e una decina a CLOKSS.⁴⁸ Alcuni atenei e altre istituzioni hanno avviato servizi di digital preservation delle proprie collezioni digitali.⁴⁹ Magazzini Digitali è il servizio nazionale di conservazione e accesso ai documenti digitali di interesse culturale⁵⁰ che comprende il deposito legale delle tesi di dottorato e alcune riviste scientifiche ad accesso aperto e siti web consultabili attraverso la piattaforma Archive-it⁵¹. Il Piano Nazionale per la Digitalizzazione (PND) demanda la conservazione delle collezioni digitali che verranno prodotte al costituendo Polo di conservazione digitale⁵² ma non contempla la conservazione delle risorse digitali native⁵³. Ne risulta un quadro molto frammentato, privo di strategie e di linee programmatiche, salvo quanto proposto in (AIB 2023).

Uno strumento molto utile per far comprendere l'urgenza e l'importanza della digital preservation è la Global List of Digitally Endangered Species – The Bit List⁵⁴ attraverso la quale il DPC classifica il rischio a cui sono esposti i vari tipi di contenuti digitali per segnalare l'urgenza di azioni di conservazione, secondo una scala crescente che varia da basso rischio, a vulnerabile, minacciato, criticamente minacciato, fino a estinto cioè inaccessibile. Il livello di minaccia cresce (categorie minacciato e criticamente minacciato) soprattutto a causa della scarsa capacità o della mancanza di volontà o dell'assenza di agenzie/enti che dovrebbero essere responsabili della conservazione. Occorre promuovere la consapevolezza che lo sviluppo, la gestione e il mantenimento di servizi di long-term digital preservation, oltre agli investimenti nelle infrastrutture tecnologiche, richiedono sostegno organizzativo per assicurare: i) il costante sviluppo di competenze attraverso la formazione di nuove figure professionali, per esempio bibliotecari esperti di metadattazione, digital curation, web archiving e digital preservation (tematiche raramente contemplate nei curricula accademici); ii) la chiara definizione e assegnazione delle responsabilità e dei ruoli in funzione della

⁴⁶ La sezione statistiche di Keepers, gestito da ISSN, aggiornata in collaborazione con 17 agenzie, al 1 giugno 2023 riporta un totale di 83.884 titoli di riviste di cui 21.368 conservati da tre o più agenzie.

⁴⁷ [https://www.portico.org/coverage/libraries/?country=Italy&keyword=.](https://www.portico.org/coverage/libraries/?country=Italy&keyword=)

⁴⁸ [https://clockss.org/digital-archive-community/supporting-libraries/.](https://clockss.org/digital-archive-community/supporting-libraries/)

⁴⁹ In mancanza di un censimento si citano come esempi le Università di Padova e Ca' Foscari di Venezia che utilizzano la piattaforma open source Phaidra al cui sviluppo contribuiscono con l'Università di Vienna. La Sapienza di Roma e la Fondazione Biblioteca Europea di Informazione e Cultura (BEIC) a metà marzo 2023 hanno annunciato l'adozione del software proprietario Rosetta.

⁵⁰ Curato dalla Biblioteca nazionale centrale di Firenze (BNCF), in collaborazione con la Biblioteca nazionale centrale di Roma (BNCR) e la Biblioteca nazionale Marciana di Venezia (BNM) [https://www.bncf.firenze.sbn.it/biblioteca/magazzini-digitali/.](https://www.bncf.firenze.sbn.it/biblioteca/magazzini-digitali/)

⁵¹ <https://archive-it.org/home/BNCF>, Archive-It è un servizio di web archiving attivato da Internet Archive nel 2006 al quale la Biblioteca nazionale centrale di Firenze ha aderito dal 2018.

⁵² La realizzazione del Polo è affidata all'Archivio Centrale dello Stato: [https://acs.cultura.gov.it/piano-nazionale-di-ripresa-e-resilienza-del-ministero-della-cultura/polo-di-conservazione-digitale/.](https://acs.cultura.gov.it/piano-nazionale-di-ripresa-e-resilienza-del-ministero-della-cultura/polo-di-conservazione-digitale/)

⁵³ Sull'argomento vedere anche il recente articolo del Direttore generale – Sovrintendente dell'Archivio centrale dello Stato (ACS) (De Pasquale 2023)

⁵⁴ [https://www.dpconline.org/digipres/champion-digital-preservation/bit-list.](https://www.dpconline.org/digipres/champion-digital-preservation/bit-list)

tipologia delle soluzioni e dei servizi di digital preservation; iii) l'attuazione di processi di verifica e revisione per valutarne la qualità e l'affidabilità.

L'efficienza e l'efficacia organizzativa dei programmi di digital preservation incidono in modo determinante sulla loro sostenibilità e al tempo stesso dipendono dalla consapevolezza e dalla volontà dei vertici istituzionali nell'esprimere chiare linee d'indirizzo strategico e definire obiettivi concreti e raggiungibili. Nelle dinamiche che legano tra loro le tre dimensioni, politica, organizzativa e sostenibilità, è fondamentale far riferimento alle esperienze consolidate, aderire alle comunità già attive (fornitori di soluzioni e servizi di digital preservation, fornitori di servizi Cloud, editori, distributori, produttori di contenuti digitali) e promuovere la collaborazione tra le istituzioni preposte alla cura del patrimonio culturale e della conoscenza (accademiche, enti di ricerca, biblioteche pubbliche e private, archivi e musei) per condividere gli investimenti e le competenze evitando di agire in modo isolato e senza coordinamento.

4. Digital preservation by design

La strategia adottata dal CNR citata in precedenza dimostra come l'assunzione di responsabilità al livello politico-decisionale delle istituzioni culturali, accademiche e di ricerca sia la condizione imprescindibile per definire strategie di digital preservation efficaci e sostenibili. Le esperienze di molti progetti internazionali, nonché i servizi attivati in molti Paesi, indicano all'Italia la necessità di rendere strutturale la digital preservation, iniziando per esempio con l'includerla come argomento di studio nei corsi universitari per la formazione dei bibliotecari e degli archivisti e creando curricula misti tra l'ambito umanistico e le discipline tecnico-informatiche.

La trasformazione digitale ci ha ormai reso familiare il concetto "by design" che significa "a partire dalla progettazione", un principio declinato nelle tre accezioni di security by design, interoperability by design, privacy by design. Lo scenario delle risorse digitali native induce a coniare anche la declinazione "digital preservation by design".⁵⁵ Ciò significa che negli ambiti del patrimonio culturale, della produzione accademica e scientifica, ogni progetto o iniziativa che produce dati e contenuti digitali dovrebbe includere fin dal suo concepimento l'adozione di strategie volte ad assicurare l'accesso, la fruizione e la conservazione nel tempo delle risorse digitali e dei dati attraverso l'attivazione di soluzioni e servizi di long term digital preservation affidabili. Da questa impostazione discende il legame sempre più stretto tra digital preservation e digital curation. La prima inizia nel momento in cui si comincia a curare la produzione, l'organizzazione e la metadattazione dei contenuti digitali e dei dati. Viceversa il processo di raccolta, selezione, organizzazione e diffusione dei contenuti e dei dati (curation) prosegue durante l'intero ciclo di vita della conservazione digitale al fine di garantire nel tempo l'accesso e la fruibilità dei contenuti e dei dati digitali. Affinché l'Italia non resti al traino rispetto agli altri paesi dell'area europea e nord-americana è auspicabile che la politica e le istituzioni, a livello locale e nazionale, percepiscano la necessità di elaborare strategie e attivare programmi di digital preservation, investendo nella formazione delle competenze e allocando le risorse per creare un'infrastruttura nazionale diffusa e condivisa.

⁵⁵ La Roadmap Scienza Aperta del CNR propone il principio "FAIR by design". Nel contesto di questo articolo si ritiene che il principio "digital preservation by design" sia inclusiva dell'intero patrimonio (digitale) culturale.

Riferimenti bibliografici

Ultimo accesso di tutti i riferimenti online, compresi quelli citati nelle note, il 22 giugno 2023.

AIB (Associazione italiana biblioteche). Gruppo di lavoro sulle biblioteche digitali. 2023. *Piano d'azione per l'infrastruttura nazionale della conoscenza*. Roma: Associazione italiana biblioteche. <https://www.aib.it/prodotto/piano-dazione-per-linfrastruttura-nazionale-della-conoscenza/>.

Benner, Jessica G., Seth Erickson, Wendy Hagenmaier, Monique Lassere, Christa Williford, e Lauren Work. 2022. *Supporting Software Preservation Services in Research and Memory Organizations: A White Paper from the Software Preservation Network's Research-in-Practice Working Group*. Software Preservation Network and Council on Library and Information Resources. <https://doi.org/10.5281/zenodo.7272726>.

Boluk, Stephanie, e Patrick LeMieux. 2017. *Metagaming: Playing, Competing, Spectating, Cheating, Trading, Making, and Breaking Videogames*. University of Minnesota Press. <https://doi.org/10.5749/9781452958354>.

Castelli, Donatella, Giovanni De Simone, Flavia Cancedda, Leonardo Candela, Valentina Colcelli, Raffaele Conte, Francesca Di Donato, Silvia Giannini, Emma Lazzeri, Silvana Mangiaracina, Roberto Puccinelli, e Maria Adelaide Ranchino. 2023. *Roadmap Scienza Aperta*. Consiglio Nazionale delle Ricerche. https://doi.org/10.57665/BICE_ROADMAP2023.

CCSDS (Consultative Committee for Space Data Systems). 2011. *Audit and Certification of Trustworthy Digital Repositories. Recommended Practice CCSDS 652.0-M-1 Magenta Book September 2011*. Washington, DC, USA. <https://public.ccsds.org/pubs/652x0m1.pdf>.

CCSDS (Consultative Committee for Space Data Systems). 2012. *Reference Model for an Open Archival Information System (OAIS) Recommendation CCSDS 650.0-B-2 MAGENTA BOOK June 2012*. Washington, DC, USA. <https://public.ccsds.org/pubs/650x0m2.pdf>.

CRL (The Center for Research Libraries) e OCLC (Online Computer Library Center). 2022. *Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist*. https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf.

Currie, Amy, e William Kilbride. 2021. *FAIR Forever? Long Term Data Preservation Roles and Responsibilities, Final Report (Version 7)*. Zenodo. <https://doi.org/10.5281/zenodo.4574234>.

De Pasquale, Andrea. 2023. *L'Archivio Centrale Dello Stato E Il Suo Ruolo Di Guida Per La Conservazione Della Memoria Digitale Del Paese. Dal Progetto Repository Degli Archivi Digitali Degli Organi Centrali Dello Stato Al Progetto Polo Di Conservazione Digitale*. DigItalia 18 (1):30-37. <https://doi.org/10.36181/digitalia-00058>.

Duranti, Luciana. 2006. *Un quadro teorico per le politiche, le strategie e gli standards di conservazione digitale: la prospettiva concettuale di InterPARES*. *Bibliotime*, anno IX, numero 1 (marzo 2006). Atti del Seminario "Digitali si diventa. Presupposti teorici e conseguenze culturali della digitalizzazione in biblioteca" Modena, 12 dicembre 2005. <https://www.aib.it/aib/sezioni/emr/bibtime/num-ix-1/duranti.htm>.

Fernandes, João, Bob Jones, Marion Devouassoux, Dominique Buyse, Ruben Van Calenberg,

Vaggelis Motsenitalis, e Jakub Urban. 2020. *ARCHIVER D2.1- State of the Art, Community Requirements and OMC Results (1.0)*. <https://doi.org/10.5281/zenodo.3618215>.

Greenberg, Jonathan, Karen Hanson, e Deb Verhoff. 2021. *Guidelines for Preserving New Forms of Scholarship*. NYU Libraries. <https://doi.org/10.33682/221c-b2xj>.

OPF (Open Preservation Foundation).2020. *Open Preservation Foundation 2019-2020 Digital Preservation Community Survey*. <https://zenodo.org/record/4066912#.Y9lRlq3MJEZ>.

Regan, Shannon, Joyce McDonough, Bob Wolven, e Oya Y. Rieger. 2016. *Strategies for Expanding e-Journal Preservation*. <https://confluence.cornell.edu/display/culpublic/Strategies+for+Expanding+E-Journal+Preservation>.

Rieger, Oya Y., Roger C. Schonfeld, e Liam Sweeney. 2022. *The Effectiveness and Durability of Digital Preservation and Curation Systems*. <https://sr.ithaka.org/publications/the-effectiveness-and-durability-of-digital-preservation-and-curation-systems/>.

Sullivan, Elaine A. 2020. *Constructing the Sacred Visibility and Ritual Landscape at the Egyptian Necropolis of Saqqara*. Stanford University Press. <https://doi.org/10.21627/2020cts>.

Waters, Donald J, ed. 2005. *Urgent Action Needed to Preserve Scholarly Electronic Journals*". The Andrew W. Mellon Foundation. <https://www.arl.org/wp-content/uploads/2005/10/ejournal-preservation-15oct05.pdf>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, e Barend Mons. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Sci Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

Quality Assessment of Research Comparisons in the Open Research Knowledge Graph: A Case Study

Jennifer D'Souza^(a), Hassan Hussein^(b), Julia Evans^(c), Lars Vogt^(d),
Oliver Karras^(e), Vinodh Ilangovan^(f), Anna-Lena Lorenz^(g), Sören Auer^(h)

a-h) TIB Leibniz Information Centre for Science and Technology

Contact: Jennifer D'Souza, jennifer.dsouza@tib.eu; Hassan Hussein, hassan.hussein@tib.eu;
Julia Evans, julia.evans@tib.eu; Lars Vogt, lars.vogt@tib.eu; Oliver Karras, oliver.karras@tib.eu;
Vinodh Ilangovan, vinodh.ilangovan@tib.eu; Anna-Lena Lorenz, anna.lorenz@tib.eu; Sören Auer, auer@tib.eu

Received: 04 April 2023; **Accepted:** 21 June 2023; **First Published:** 15 January 2024

ABSTRACT

The Open Research Knowledge Graph (ORKG) is a digital library for machine-actionable scholarly knowledge, with a focus on structured research comparisons obtained through expert crowdsourcing. While the ORKG has attracted a community of more than 1,000 users, the curated data has not been subject to an in-depth quality assessment so far. Here, proposed as a first exemplary step, within a team of domain experts, we evaluate the quality of six selected ORKG Comparisons based on three criteria, namely: 1) the quality of semantic modelling, 2) the maturity of the Comparisons in terms of their completeness, syntactic representation, identifier stability, and their linkability mechanisms ensuring the interoperability and discoverability. Finally, 3) the informative usefulness of the Comparisons to expert and lay users. We have found that each criterion addresses a unique and independent aspect of quality. Backed by the observations of our quality evaluations presented in this paper, a fitting model of knowledge graph quality appears one that is indeed multidimensional as ours.

KEYWORDS

Knowledge Graph; Open Research Knowledge Graph; Linked Open Data (LOD); Human-Computer Interaction; Survey.

1. Introduction

Scholarly communication has long relied on discourse-based methods, but the sheer volume of new publications each year can be overwhelming for researchers. The Open Research Knowledge Graph (ORKG) project (Auer et al. 2020) advocates for a structured format to represent scholarly contributions as a knowledge graph (KG) readable by machines and humans. With its next-generation semantic publishing model (Berners-Lee, Hendler, and Lassila 2001; Shotton 2009), it claims that only with a combination of human- and machine-actionability can we allow for novel exploration and assistance services for scholars, thereby strengthening research even in the face of voluminous publication cycles worldwide (Bornmann and Mutz 2015, 2217; Johnson, Watkinson, and Mabe 2018, 6).

The ORKG simplifies access to scholarly knowledge by explicitly representing a machine-actionable set of property-value pairs to describe scholarly contributions. With such a structured description format, several works of research can be compared by their similar properties using the ORKG Research Comparisons feature (Oelen et al. 2019; 2020). In keeping with standards, it adheres to the FAIR (Findable, Accessible, Interoperable, and Reusable) Data Principles (Wilkinson et al. 2016). ORKG Comparisons are a way of making science more accessible to all researchers, thereby a concrete step toward improved scientific communication and also encouraging scientific interdisciplinarity. However, given that the ORKG employs crowdsourcing to populate the KG, data quality unavoidably varies. Maintaining high standards thus necessitates effective evaluation and management of the information quality of the graph.

Information quality is generally defined as “fitness for use” (Zaveri et al. 2016, 2). Embedded within this definition is a question: use by *whom* and for *what purpose*? As alluded to earlier, the ORKG is intended to be both human-readable and machine-actionable – two very disparate user groups with equally disparate (and sometimes competing) desiderata. A graph with high machine-actionability may be overly complex and confusing for human readers, while a graph with high human-readability may lack the explicit conceptual relationships necessary for machines. In other words, there is no singular definition of quality suitable to all users and use cases.

The ORKG knowledge curation is based on the wisdom-of-the-crowd philosophy¹ concretely implemented via the crowdsourcing methodology. Crowdsourcing is “the outsourcing of a piece of work to a crowd of people via an open call for contributions” rather than hiring or contracting employees for the task (Daniel et al. 2018, 2). It has become an indispensable tool for data acquisition in many fields (Zhang 2022, 749), but ensuring the quality of the obtained data remains a complex open issue (Daniel et al. 2018, 2).

This work presents a qualitative case study evaluating six crowdsourced ORKG Comparisons based on three criteria: 1) semantic modelling quality; 2) the maturity of the Comparisons in terms of their completeness, syntactic representation, identifier stability, and their linkability mechanisms ensuring the interoperability and discoverability; and finally 3) the informative usefulness of the Comparisons to expert and lay users. Each of these quality dimensions corresponds to an ORKG usage goal for humans and machines – as already noted, these goals are not always complementary. Therefore, each criterion is addressed with its own quality evaluation approach.

¹ https://en.wikipedia.org/wiki/Wisdom_of_the_crowd

The paper is structured as follows. Section 2 discusses related work on KG quality evaluations and introduces the Knowledge Graph Maturity Model (KGMM) (Hussein et al. 2022) used for criterion 2. Section 3 introduces the ORKG and its features, in particular the ORKG Comparisons analysed in this work. Section 4 details the evaluation criteria, namely semantic modelling, maturity, and informativeness. Section 5 presents the quality evaluations of the six selected ORKG Comparisons for each criterion. Section 6 discusses the obtained results, possible conclusions, and limitations of the study, and Section 7 concludes the paper.

2. Related Work

Data quality is a multidimensional, inherently contextual, and task-dependent concept. To evaluate data quality, researchers advise combining multiple metrics adapted to each individual use case, such as surveys of stakeholders' perceptions, assessments of ease of access, and statistics of accuracy and completeness (Pipino, Lee, and Wang 2002; Bizer and Cyganiak 2009; Zaveri et al. 2016; Färber et al. 2018). The following subsections present two perspectives on knowledge graph quality evaluation that are particularly relevant to the ORKG: the generic frameworks of Linked Open Data and FAIR Guiding Principles, and the more specific Knowledge Graph Maturity Model (KGMM).

2.1 Linked Open Data and FAIR Guiding Principles

The Five-Star Linked Open Data (LOD) system was introduced by Tim Berners-Lee in 2010 to encourage publishing data on the Web in machine-readable, non-proprietary formats following the Resource Description Framework (RDF) standards and containing links to other data.² Another prominent approach to linked data management is FAIR – Findable, Accessible, Interoperable, and Reusable – which defines the “characteristics that contemporary data resources, tools, vocabularies and infrastructures should exhibit to assist discovery and reuse” by both humans and machines (Wilkinson et al. 2016). Zaveri et al. (2016) surveyed thirty quality measures for linked data quality. They compiled eighteen “core” data quality dimensions divided amongst four classifications and encompassing both qualitative and quantitative metrics for assessment. However, they find that most metrics were not explicitly defined or were not statistically precise, and none were formally validated. The resulting framework they present is a starting point for further development of more defined metrics.

2.2 Knowledge Graph Maturity Model

The Knowledge Graph Maturity Model (KGMM) offers an assessment of data in KGs with the goal of offering it in “the most mature, complete, representable, stable, and linkable shape” (Hussein et al. 2022). The model itself distinguishes five maturity levels, with each level possessing its own set of quality measures that are ranked in priority as either essential, important, or useful.

² <https://www.w3.org/DesignIssues/LinkedData.html>

To pass to the next maturity level, all measures rated essential and at least half of those rated as important must be satisfied. Whether a measure is satisfied or not is a binary choice. Below, we briefly describe the five levels of the KGMM in the context of the specific application scenario of ORKG Comparisons.

- **Level 1: Published.** Passing the quality measures defined for this maturity level indicates that the KG is accessible on the web with an open licence. ORKG Comparisons satisfy this level by default, due to the ORKG publishing model.
- **Level 2: Completeness.** Passing this level implies that the ORKG Comparison has sufficient information scope and density to meet the requirements of the given task (Wang and Strong 1996). This encompasses the accuracy, timeliness, and trustworthiness of the data, as well as a clear record of its provenance. It additionally includes whether there are missing values, whether all necessary statements are included, and whether there is documentation of the KG's contents.
- **Level 3: Representation.** Clearing this level assures that the Comparison is free from redundant or duplicated entities; its format and metadata ensure it is reusable by others. Where possible, data might be presented in different formats (e.g., tables, charts, etc.) while still maintaining full backward compatibility.
- **Level 4: Stability.** Stability implies that the KG's data is trackable w.r.t. provenance and version history. Furthermore, it enables mechanisms to leverage globally unique IDs such as DOIs and ORCID. It provides a means for querying the data in the KG, such as SPARQL or an API endpoint. Based on the features of the ORKG, Comparisons satisfy this level by default.
- **Level 5: Linkability.** The KG can be dereferenced using URIs and an HTTP call retrieves the graph in RDF syntax. The KG is well-linked to internal and external resources. The ORKG satisfies dereferenceability by default even though they may or may not be well-linked. Based on the minimal dereferenceability criteria, Comparisons automatically pass this level.

In general, the KGMM can be applied in a crowdsourcing setting involving multiple reviewers of the quality of the KG. Survey questions addressing many of the quality measures have been developed to appear alongside a KG, and crowdsourced reviewers are asked to reply with their opinion.

3. Open Research Knowledge Graph

In this section, we introduce two relevant facets of the Open Research Knowledge Graph (ORKG).

Paper Contributions. Figure 1 illustrates a structured description of a paper's contribution in the ORKG. The ORKG represents papers as research contribution(s) in a structured and semantic way using (subject, predicate, object) triple statements,³ where the predicates are salient properties of a given work's contribution. The subject and object positions are generally filled by resources.⁴

³ <https://www.w3.org/TR/rdf11-concepts/#data-model>

⁴ <https://www.w3.org/TR/rdf11-concepts/#resources-and-statements>

ORKG resources are resolved to the ORKG namespace and constitute the ORKG scholarly knowledge vocabulary.⁵ Mapping between an ORKG resource and an equivalent concept resource in another ontology is done using the OWL sameAs property.⁶ ORKG objects may also be filled by a literal such as a text, number, or date. The ORKG also stores basic metadata of the original paper including the title, publication date, and list of authors. Papers may also be assigned to a research field within the ORKG's 700-fields taxonomy.⁷

The screenshot displays the ORKG user interface for a specific contribution. At the top, the title of the paper is shown: "Multiscale deformations lead to high toughness and circularly polarized emission in helical nacre-like fibres". Below the title, there is a list of authors: February 2016, Materials Science and Engineering, Jia Zhang, Wenchun Feng, Huangxi Zhang, Zhenlong Wang, Heather A. Calcaterra, Bongjun Yeom, Ping An Hu, and Nicholas A. Kotov. The DOI is provided as <https://doi.org/10.1038/ncomms10701>. The main content area is titled "GO-PVA" and shows a structured contribution with the following properties and values:

Property	Value
Instance of	R198658, Contribution
has material	Graphene oxide NaOH poly(vinyl alcohol)
method	wet spinning
research problem	Mechanical properties of nacre-inspired materials
result	Strength Toughness Young modulus
Sample shape and dimensions	1D fiber

On the right side, there is a sidebar with a "Provenance" tab selected, showing the contribution was added on 09 Sep 2022 by Volodymyrddk. There is also an "Add to comparison" button at the top right of the sidebar.

Figure 1. The ORKG user interface of an example scholarly paper w.r.t. its structured contribution content as (property, value) pairs, where the properties are the contribution relevant aspects.

ORKG Research Comparisons. The ORKG platform supports downstream services such as the creation of tabular Comparisons of contributions which constitute ORKG subgraphs. Figure 2 shows an example of one such Comparison. The ORKG Comparisons feature creates one aggregated view of the values of several structured contributions with more-or-less similar sets of predicates. Comparisons can encompass multiple contributions from the same article (e.g., a Comparison of AI benchmark characteristics introduced as several contributions in a single paper); or contributions with similar predicates from different articles (e.g., a Comparison of the Covid-19

⁵ <https://www.w3.org/TR/rdf11-concepts/#vocabularies>

⁶ https://orkg.org/property/SAME_AS

⁷ https://orkg.org/help-center/article/20/ORKG_Research_fields_taxonomy

reproductive number (R0) estimate studies undertaken by different research groups in different countries).

1D fiber based on graphene oxide

September 2022 Volodymyrddk

Mechanical properties of nacre-inspired 1D fiber based on graphene oxide.

Properties	Scalable One-Step Wet-Spinning of Graphene Fibers and Yarns from Liquid Crystalline Dispersions of Graphene Oxide: Towards Multifunctional Textiles <i>GO fibers coagulated by CaCl2 - 2013</i>	Scalable One-Step Wet-Spinning of Graphene Fibers and Yarns from Liquid Crystalline Dispersions of Graphene Oxide: Towards Multifunctional Textiles <i>GO fibers coagulated by chitosan - 2013</i>	Liquid crystal self-templating approach to ultrastrong and tough biomimic composites <i>GGO-HPG - 2013</i>	Ultrastrong Fibers Assembled from Giant Graphene Oxide Sheets <i>GO-ca2+ - 2012</i>
has material	calcium(2+) Graphene oxide	chitosan Graphene oxide	Ca2+ Graphene oxide hyperbranched polyglycerol	Ca2+ Graphene oxide
has method	wet spinning	wet spinning	wet spinning	wet spinning
has result	Strength Toughness Young modulus	Strength Toughness Young modulus	Strength Toughness Young modulus	Strength Toughness Young modulus
has result/strength				
↳ has measurement value*	412	442	652	364.4
↳ has unit*	megapascal	megapascal	megapascal	megapascal
↳ has unit/megapascal/same as*	https://www.wikidata.org/entity/Q21062777	https://www.wikidata.org/entity/Q21062777	https://www.wikidata.org/entity/Q21062777	https://www.wikidata.org/entity/Q21062777
↳ same as*	strength	strength	strength	strength
↳ same as/strength/same as*	https://www.wikidata.org/entity/Q605035	https://www.wikidata.org/entity/Q605035	https://www.wikidata.org/entity/Q605035	https://www.wikidata.org/entity/Q605035
has result/toughness				
↳ has measurement value*	4.8	4.8	14	6.8
↳ has unit*	megajoule per cubic metre	MJ/m3	MJ/m3	%
↳ has unit/mj/m3/same as*		megajoule per cubic metre	megajoule per cubic metre	
↳ same as*	toughness	toughness	toughness	toughness

Figure 2. An example ORKG Comparison based on a common set of properties (depicted in the grey column) used to semantically describe four different contributions of scholarly articles (depicted in the columns with the red header).

3.1 Six ORKG Research Comparisons – the Qualitative Analysis Corpus of this Work

The ORKG partners with researchers (senior PhD students, postdocs, and other advanced scholars) from diverse domains to curate contributions and Comparisons that address key research problems in their field. Grantees receive up to €400 per month as an incentive as well as mentoring from the ORKG team. Of the 1,134 total Comparisons in the ORKG, 348 were created by curation grantees. Six of these Comparisons were selected based on the quality of their semantic modelling, as defined by experts (see Section 4.1 for a detailed explanation of the semantic modelling criteria). Table 1 provides overview statistics of these Comparisons.

Comparison name	Contributions (Papers)	Research field	Research problem
1D Fiber ⁸	8 (5)	Materials Science and Engineering	Mechanical properties of nacre-inspired materials
Microneedle Tech. ⁹	7 (7)	Medicinal Chemistry and Pharmaceuticals	Microneedle technology
Supply Chain ¹⁰	16 (16)	Operations Research, Systems Engineering and Industrial Engineering	Supply chain
CO ₂ Gas Flux ¹¹	11 (11)	Oceanography	CO ₂ flux estimation of the ocean
Glucose Sensors ¹²	7 (7)	Nanoscience and Nanotechnology	Development of Glucose Sensors Based on Nanomaterials
Smart Cities ¹³	10 (9)	Information Systems, Process and Knowledge Management	Smart cities

Table 1. Overview of the six ORKG Comparisons qualitatively analysed in this work. A research paper can have more than one contribution hence the count of papers and contributions are not always equal.

4. Six ORKG Research Comparisons – a Qualitative Analysis

The six selected ORKG Comparisons were qualitatively analysed by the following three criteria: 1) semantic modelling, 2) maturity ratings based on the KGMM, and 3) the usefulness of their content for researchers and laypeople. These criteria address different user groups and use cases, with semantic modelling focusing on machine-actionability, content usefulness for human readers, and the maturity model blending both with the principles of LOD and FAIR.

4.1 Qualitative Evaluation Criterion 1 – Semantic Modelling

The evaluation dataset in this study was initially selected based on the semantic modelling quality determined by two semantic modelling experts in the ORKG team. Good semantic modelling criteria, especially for crowdsourced knowledge, are defined based on observations made on the data in terms of which types of modelling decisions create a human-readable and machine-actionable graph. However, the ORKG recommended guidelines suggest best practices, but are not mandated. For this evaluation, we selected three Comparisons which satisfy all or many of the semantic modelling guidelines and can be deemed good examples to learn from; and three Comparisons that are examples of pitfalls to avoid in semantic modelling. It should be noted that *only* the quality of the modelling is judged here – the content itself is not considered.

⁸ <https://orkg.org/comparison/R215963/>

⁹ <https://orkg.org/comparison/R161079/>

¹⁰ <https://orkg.org/comparison/R212576/>

¹¹ <https://orkg.org/comparison/R160742/>

¹² <https://orkg.org/comparison/R143853/>

¹³ <https://orkg.org/comparison/R140131/>

4.1.1 Semantic Modelling Principles

- I. *Use of ontology resources.* A marker of a high-quality ORKG Comparison is in its use of resources rather than literals for object nodes wherever possible. Resources can be reused across statements, thereby establishing connections between papers. At best, resources should have a user-supplied definition and additional description statements, which we identify as *ontology resources*. However, this requires more input work on the part of the creator, and it is often the case that resources are not described or reused within the ORKG. We identify these as *ad hoc resources*. Use of ontology resources is always preferable to use of ad hoc resources.
- II. *Establishing mappings between resources.* Where appropriate, ORKG resources may be linked to external ontologies, such as Wikidata and GeoNames. Such resources are also considered ontology resources. This results in ORKG resources situated in the linked data cloud as well.
- III. *Modelling best practices for resources that are constants.* A prerequisite to following these recommended practices is to determine whether a triple object should indeed be modelled as a literal value and not a resource. A hint for which values to model as literals can be gleaned from the available RDF datatypes for literals – e.g., boolean, string, integer, decimal, date, etc.¹⁴ However, there still remains significant subjective decision making on the part of the modeller. To support this, the following sub-list offers some recommendations for common use cases, although it is by no means exhaustive.
 - A. *Use the QUDT specification to model measurements.* A resource representing a measurement may be used in the object position in one statement and then reused in the subject position of additional statements for each facet of information such as measurement value (specified with a literal) and measurement unit (specified with a resource).
 - B. *Use the international standard YYYY-MM-DD date format.* Building upon XML Schema, RDF specification supports the date datatype for literals and specifies the format.
 - C. *Express time period nodes as a two-statement resource in the graph, one for start and one for end.* The literals for date and time must not constitute more than one date or time reference point. Ranges can be expressed as a resource node in the graph which is further specified via begin and end predicates leading to literal value specifications of the begin and end date/time.
 - D. *Maintaining consistency in modelling.* Maintaining consistency – of terms, labels, modelling strategies, and so on – is also a hallmark of high-quality modelling. The modelling choices should follow a consistent pattern of expression.
- IV. *Select relevant properties which have values across multiple contributions.* An ORKG Comparison with many empty cells affects human-readability. It is best to minimise the number of empty cells as far as possible by careful selection of properties with values across most if not all contributions in the Comparison.
- V. *Use phrase-based resource names.* Avoid communicating too much information (e.g., complete sentences) in individual literals or labels of resources. This hinders machine-actionability. However, the use of sentences is encouraged for the description predicate.

¹⁴ <https://www.w3.org/TR/rdf11-concepts/#section-Datatypes>

4.2 Qualitative Evaluation Criterion 2 – Knowledge Graph Maturity Model

KGMM (Hussein et al. 2022) was discussed in theory to apply to KGs generally, but also included a proposed implementation specifically for ORKG Comparisons. All measures that were not either directly satisfied by the ORKG publishing model or which did not have a mechanism in the ORKG based on which they could be satisfied, were proposed as a survey questionnaire. A survey question asked the user to offer a yes-or-no answer based on whether the measure was deemed satisfied or not. The cumulative responses to the survey, in combination with those measures which are automatically satisfied, determine the maturity of the ORKG Comparison.

4.2.1 Maturity Evaluation as a Questionnaire

- **Pilot evaluations.** The proposed KGMM questionnaire was pilot tested to assess the clarity and difficulty of the questions. Specialists from Biology, Physics, and Meteorology were involved in the pilot study, and based on their feedback, four questions were rephrased. The final survey questions and the corresponding KGMM maturity measure are presented in Table 2, and the survey itself is published online.
- **Final evaluations.** The survey was conducted with six subject librarians from TIB. They were selected based on their one-to-one domain expertise mapping against the selected Comparisons, where in one case the subject librarian was a specialist in the broader domain of the Comparison but not the specific research problem addressed.¹⁵ Each subject librarian was presented with a Comparison from their domain to rate. A summary of their responses is provided in Table 4 (see Section 5).

A requirement of the current KGMM is that the participant taking it would also need to have a basic understanding of semantic modelling to accurately address its questions. E.g., to correctly respond to the usefulness of the Linkability measure in the KGMM Level 5 would mean the participant knows what ontologies are – a concept from the field of semantic modelling – and has an idea of the ones that are suitable. Here, we make the following disclaimer: our six participants, while they are domain experts, may not necessarily be aware of the field of semantic modelling. Thus, to obtain a balanced view, the maturity evaluation survey was also run with one additional participant: the developer of the KGMM model. His responses are recorded in parentheses in Table 4 alongside the maturity evaluations from the subject librarians. Thus, to get an accurate estimation of the maturity ratings, the reader of the table can see the response from the domain-expert who may not be aware of semantic modelling juxtaposed against the response of one with sufficient proficiency in semantic modelling and advanced proficiency in science but who is not a domain-expert.

¹⁵ Specifically, *Several works on the dynamic redesign of a closed-loop supply chain network using accelerated Benders decomposition and robust optimisation models.*

	Measure	Survey Question
Level 1: Published	Responsiveness**	Satisfied by ORKG publishing model.
	Licence**	
	Syntactic Accuracy*	
	Easiness*	
Level 2: Completeness	Provenance**	Satisfied by ORKG publishing model.
	Timeliness*	Do you think that the comparison is outdated?
	Instance Completeness*	Do you think the properties of this comparison can generalize as a template to capture new contributions from additional scholarly articles on this research theme?
	Property Completeness*	Do you think that the comparison has enough properties?
	Documentation Completeness*	Do you think the comparison's description is descriptive?
Level 3: Representation	Reusability**	Satisfied by ORKG publishing model.
	Conciseness**	Do you think that the comparison has adequate resource/property values?
	Data Representation	If the comparison doesn't have a visualisation, do you think it needs one? If the comparison has a visualisation, do you think it is a fitting representation of the results?
Level 4: Stability	Trackability**	Satisfied by ORKG publishing model.
	Queryability*	
	Identifier Stability*	
Level 5: Linkability	Dereferencability	Satisfied by ORKG publishing model.
	Linkability	Would you have suggestions for ontologies to link/resolve the predicates and resources of the selected comparison?

Table 2. Questionnaire methodology used to evaluate the ORKG Comparisons' maturity based on KGMM. Measures marked with double asterisks are essential, those with one asterisk are important, and those without an asterisk are useful. The evaluation for each measure is formulated as a corresponding question to which the responses are recorded as a yes/no.

4.3 Qualitative Evaluation Criterion 3 – Usefulness

Another measure of quality is how well a given KG serves its intended information purpose. For the ORKG, one intended purpose of Comparisons is to offer readers an overview of the scientific progress on a research problem or similar. To assess this, we asked subject librarians from each domain: *Is this comparison useful to a specialist in this field/layperson trying to understand this topic?* The evaluations are described in Section 5 and summarised in Table 5.

5. Evaluations

5.1 1D fiber based on graphene oxide

This Comparison contains 9 predicates, 50 resources, and 32 literals and has a maximum sub-graph depth of 4 nodes. It presents studies measuring the strength, toughness, and compressive stiffness of composite fibres made of graphene oxide in combination with different materials.

- **Semantic modelling.** This Comparison makes good use of ontology resources, including establishing mappings from ORKG resources to external ontologies. Modelling best practice standards are followed and the representations are clear and consistent. The selected properties contain values across multiple contributions. Resource names are succinct.
- **Maturity rating.** This Comparison has a maturity level of 2. It satisfies Timeliness and Instance Completeness, but not Property Completeness or Documentation Completeness (Level 2). It does not satisfy Conciseness or Data Representation (Level 3). It also does not satisfy Linkability (Level 5).
- **Usefulness.** In the opinion of the subject librarian, this Comparison is lacking critical information for specialists, namely details about the process such as sample size and mean variation. Moreover, the names used for materials are inconsistent (and therefore confusing) and in some cases unclear. As for the suitability of the Comparison for laypeople, the subject librarian was unsure, believing that more information about the samples and materials, as well as the testing method, would probably make it more useful for laypeople.

5.2 Microneedle Technology as Insulin Delivery Systems

This Comparison contains 19 predicates, 99 resources, and 133 literals and has a maximum sub-graph depth of 3. It surveys contributions describing microneedles constructed from six different materials and of varying sizes for the purpose of subcutaneous insulin injection.

- **Semantic modelling.** This Comparison uses ontology resources very well, although at the time it was created the ORKG did not have the functionality for linking to external ontologies. Modelling best practices are followed and the overall structure is consistent. Almost all properties have values across all contributions. Resources have short phrase-based names.
- **Maturity rating.** This Comparison is still Level 1. It satisfies Property Completeness, but not Timeliness, Instance Completeness, or Documentation Completeness (Level 2). It satisfies Conciseness and Data Representation (Level 3). It does not satisfy Linkability (Level 5).
- **Usefulness.** This Comparison was deemed not useful to a specialist because it only evaluates one review article, despite a considerable output of additional research on this topic. It is also not suitable for laypeople because it is too niche.

5.3. Several works on the dynamic redesign of a closed-loop supply chain network using accelerated Benders decomposition and robust optimisation models

This Comparison contains 15 predicates, 45 resources, and 83 literals and has a maximum sub-graph depth of 2 nodes. It presents research on problems arising in supply chain reconfiguration

according to a number of variables, such as whether the entire network is expanding, the future facility plans, and the number of products.

- **Semantic modelling.** While this Comparison does use resources, most are ad hoc rather than ontology resources. The Comparison contains no measurements but still follows good modelling practices and maintains consistency. Almost all properties have values across all Comparisons. Resource names are succinct.
- **Maturity rating.** This Comparison is also still Level 1. It satisfies Property Completeness, but not Timeliness, Instance Completeness, or Documentation Completeness (Level 2). It satisfies Conciseness and Data Representation (Level 3). It also satisfies Linkability (Level 5).
- **Usefulness.** The subject librarian described the Comparison as providing an expert-level overview of the differences and similarities of the contributions. While it serves this purpose well, it was unclear how useful this overview would be for those already experts in the field. On the other hand, there is too little pedagogical exposition for a layperson. A particular stumbling point for this Comparison is that the topic has connections to multiple fields, none of which are explained. Additionally, the property names are confusing and lacking context.

5.4 CO₂ gas flux assessment of the various oceanic regimes

This Comparison contains 20 predicates, 12 resources, and 116 literals and has a maximum subgraph depth of 1 node. It gives an overview of studies measuring changes in atmospheric CO₂ concentration over bodies of water depending on the time and location.

- **Semantic modelling.** It has an overabundance of literals, and only uses ad hoc resources. The modelling does not follow best practices (e.g. dates are not expressed in the proper format, time periods are expressed using a single statement) and is inconsistent. Many of the selected properties are not relevant to all contributions. However, it does use appropriate phrase-based naming.
- **Maturity rating.** This Comparison has a maturity level of 5. It satisfies all measures except for Data Representation (Level 3).
- **Usefulness.** The subject librarian was unable to judge if this Comparison would be useful for specialists. The Comparison was judged not suitable for laypeople.

5.5 Comparison among glucose sensors based on different nanomaterials

This Comparison contains 9 predicates, 8 resources, and 80 literals and has a maximum subgraph depth of 1 node. It includes the sensitivity and measurements of glucose sensors produced from seven different nanomaterial constructions.

- **Semantic modelling.** Similar to the previous Comparison, there are a multitude of literals which would be more appropriate as resources, and only a single ad hoc resource. The modelling does not follow best practices (e.g., measurements are modelled as single statements, and include text preceding the numeric value), nor is it consistent. While some of the values follow phrase-based naming conventions, the values for sensor measurements are given as full sentences.

- **Maturity rating.** This Comparison also has a maturity level of 5. All measures are met, except for Instance Completeness (Level 2).
- **Usefulness.** The subject librarian determined that this Comparison would be useful for a specialist in the field, but would not be suitable for a layperson as it is too specific.

5.6 Smart cities and cultural heritage

This Comparison contains 26 predicates, 104 resources, and 5 literals and has a maximum sub-graph depth of 2 nodes. It contains surveys, frameworks, and methodology publications reporting broad insights on modelling cultural heritage in smart cities, according to varied definitions of smart urban development.

- **Semantic modelling.** This Comparison contains many ad hoc resources instead of ontology resources. There are no measurements to evaluate, but the modelling is inconsistent and confusing. Several properties apply to only a handful of contributions, leading to a Comparison with many empty values. Full-sentences are given as values for multiple properties, including those describing the research problem and results.
- **Maturity rating.** This Comparison has a maturity level of 2. It satisfies Timeliness and Documentation Completeness, but not Instance Completeness or Property Completeness (Level 2). It does not satisfy Conciseness, but it does satisfy Data Representation (Level 3). It satisfies Linkability (Level 5).
- **Usefulness.** For specialists in the field, the subject librarian was unable to judge if this Comparison would be useful. They deemed it useful for laypeople.

	1D Fiber	Micro-needle Tech.	Supply Chain	CO ₂ Gas Flux	Glucose Sensors	Smart Cities
Ontology resources	✓	✓	✓	✗	✗	✗
External links	✓	✗	✓	✗	✗	✗
Modelling best practices	✓	✓	○	✗	✗	○
Consistency	✓	✓	✓	✗	✗	✗
Relevant properties	✓	✓	✓	✗	✓	✗
Phrase-based names	✓	✓	✓	✓	✗	✗

Table 3: The results of the semantic modelling evaluation. The ✓ symbol indicates the presence of the given element while the ✗ symbol indicates its absence. The ○ symbol indicates this element is not applicable to the given Comparison.

	1D Fiber	Micro-needle Tech.	Supply Chain	CO ₂ Gas Flux	Glucose Sensors	Smart Cities
Responsiveness**	★	★	★	★	★	★
Licence**	★	★	★	★	★	★
Syntactic Accuracy*	★	★	★	★	★	★
Easiness*	★	★	★	★	★	★
Provenance**	★	★	★	★	★	★
Timeliness*	✓ (✓)	✗ (✗)	✗ (✓)	✓ (✗)	✓ (✓)	✓ (✓)
Instance Completeness*	✓ (✓)	✗ (✓)	✗ (✓)	✓ (✓)	✗ (✓)	✗ (✗)
Property Completeness*	✗ (✓)	✓ (✓)	✓ (✓)	✓ (✓)	✓ (✓)	✗ (✓)
Documentation Comp.*	✗ (✗)	✗ (✗)	✗ (✓)	✓ (✓)	✓ (✗)	✓ (✓)
Reusability**	★	★	★	★	★	★
Conciseness**	✗ (✓)	✓ (✗)	✓ (✓)	✓ (✗)	✓ (✓)	✗ (✗)
Data Representation	✗ (✗)	✓ (✓)	✓ (✓)	✗ (✓)	✓ (✓)	✓ (✓)
Trackability**	★	★	★	★	★	★
Queryability*	★	★	★	★	★	★
Identifier Stability*	☆	☆	☆	☆	☆	☆
Dereferencability	★	★	★	★	★	★
Linkability	✗ (✗)	✗ (✗)	✓ (✗)	✓ (✗)	✓ (✓)	✓ (✗)
Maturity Level	2 (5)	1 (2)	1 (5)	5 (2)	5 (5)	2 (2)

Table 4. Whether a Comparison satisfies the surveyed measures, according to a subject librarian (or in the opinion of the KGMM developer), and the maturity level resulting from these judgments. The ✓ symbol indicates that it satisfies the given measure, while the ✗ symbol indicates that it does not. The ★ symbol indicates measures which are automatically satisfied by the ORKG publishing model, while ☆ means the mechanism for satisfying this measure is present but not guaranteed to be applied. Measures marked with double asterisks are essential, those with one asterisk are important, and those without an asterisk are useful.

	1D Fiber	Micro-needle Tech.	Supply Chain	CO ₂ Gas Flux	Glucose Sensors	Smart Cities
Expert	✗	✗	○	○	✓	○
Layperson	○	✗	✗	✗	✗	✓

Table 5. Whether a Comparison is useful to an expert or layperson, in the opinion of a subject librarian. The ✓ symbol indicates it is useful, the ✗ symbol indicates not useful, and the ○ symbol indicates the subject librarian was unsure or felt unable to judge.

6. Discussion

6.1 Evaluation Results

The KGMM is still in its early development stages. Many of the measures are inherently difficult to evaluate, in that they are both subjective and challenging for even domain experts to assess. As some measures lack proposed survey questions and are not addressed by the ORKG, it was necessary to exclude four measures from Level 2 – of which two were classified as essential – from our implementation. Additionally, we classify Timeliness as important and not essential because we observe it is encumbered by the research field productivity artefact. In a highly productive field,

e.g., Computer Science, where the publication cycle is rapid, if Timeliness were essential a Comparison for that field would never cross the next maturity stage; on the other hand, a Comparison for a low productive field, e.g., Mathematics, would face no such hindrance. This insight about an essential measure halting progression through the KGMM relates more broadly to one of our findings. Though the KGMM is a hierarchical framework, its stagewise measures can be entirely independent between stages. Therefore, it is possible for a KG to receive a low maturity rating, even while satisfying the essential measures of higher levels. This may give the impression that the amount of improvement or development necessary for a KG to attain a high maturity level is much greater than it actually is.

The KGMM takes a holistic approach to quality evaluation which incorporates a broad range of criteria and standards, some of which are also influenced by external factors such as UI. While care has been taken in the development of the KGMM to ensure a minimum level of clean and accurate data modelling, it cannot replace an assessment focused on and specific to semantic modelling. Thus, we observe no correspondence between the semantic modelling evaluation and the KGMM rankings.

With regard to the usefulness evaluation, we note that the results are heavily influenced by the subjectivity of the domain experts consulted. Moreover, ORKG Comparisons may be used for a variety of independent purposes (e.g., understanding a research problem, finding datasets, identifying potential collaborators, etc), and whether a Comparison is useful for one purpose does not determine its usefulness for another. Nonetheless, two interesting points of consideration emerge from this assessment. First, none of the Comparisons were judged to be useful for both specialists and laypeople, and in the case of those that were not useful for either, it was for different reasons. Second, semantic modelling which is highly machine-actionable is often not highly human-actionable and may be difficult for readers to interpret. Similarly, flawed semantic modelling often simplifies the KG and presents information in a more intuitive structure, so reduced machine-actionability may result in increased human-actionability.

6.2 Differing Information Objectives of ORKG Comparisons

A good quality evaluation of ORKG Comparisons should consider the Comparison's information objective as seen by its creator. For instance, Comparison of Water Domain Ontologies provides an overview of multiple ontologies developed for describing water and water-related systems, such as water quality management and solar-powered water heating. The Comparison shows generic ontology aspects such as number of classes, number of data properties, number of object properties, and links to the ontologies themselves. However, if someone were seeking qualitative information on the theme of water domain ontologies showing the reader how the knowledge capture objective of each of the ontologies differ from or are similar to each other, this Comparison would not suffice. While it is briefly modelled in terms of the 'ontology component' property in the given Comparison, a new Comparison that directly addresses the new information objective sought would be ideally befitting.

6.3 Limitations of Our Survey Methodology

Using a survey to solicit feedback from KG users has two notable advantages: it is easy to source the feedback, and the relevant users are consulted directly. However, this approach also has some limitations. One limitation is that the survey takers must be knowledgeable about both semantic modelling and the relevant field of study. As shown in Table 4, the results of the KGMM survey are dramatically different when taken by the developer of the KGMM, who is not a domain expert in any of the Comparison fields but is an expert in KGs, compared with the subject librarians. In five cases the developer identified additional external ontologies which ought to be linked within the given Comparison, which were only identified by subject librarians in two cases. Conversely, when asked if the properties generalise well enough for additional contributions to be added, the developer answered affirmatively in five cases, where subject librarians answered affirmatively in only two cases. Determining the generalizability of the properties requires an understanding of the associated research landscape, which a domain expert is best equipped to assess. Another limitation is that a binary response does not benefit all question types. For instance, a question asking about the descriptiveness of the Comparison documentation would best be addressed by maybe a response continuum bar rather than a simple yes or no.

7. Conclusion

This work presents a case study evaluating ORKG Comparisons based on three quality criteria. First, the quality of the semantic modelling in each KG was assessed by an expert in this field. Second, subject librarians completed a survey evaluating the KGs on a research topic from within their domain of expertise. This survey elicited feedback relevant to establishing a KG's maturity level, as defined in Hussein et al. (2022). Third, the same subject librarians assessed whether they believed the given KG would be useful for informing either a domain expert or a layperson on the research theme. No Comparison was highly rated in all approaches, nor were any judged useful for both domain experts and laypeople. While the study's small sample size of only six subgraphs from a single KG and the subjective nature of the evaluation criteria limit the statistical rigour and generalizability of the conclusions, they are nonetheless suggestive. These results highlight that quality criteria must be defined independently according to each use case, and assessing multiple quality dimensions in aggregate measures is challenging.

Some avenues for future work include the possibility of incorporating explicit semantic modelling evaluation into the maturity model, reformulating the maturity model to allow for answers other than a binary yes or no, and developing objective measures for assessing maturity. More generally, for any quality evaluation framework, it is necessary to assess quality by identifying independent quality criteria and reflecting this independence in the overall quality assessment.

References

- Auer, Sören, Allard Oelen, Muhammad Haris, Markus Stocker, Jennifer D'Souza, Kheir Eddine Farfar, Lars Vogt, Manuel Prinz, Vitalis Wiens, and Mohamad Yaser Jaradeh. 2020. "Improving Access to Scientific Literature with Knowledge Graphs." *Bibliothek Forschung Und Praxis* 44 (3): 516–29. <https://doi.org/10.1515/bfp-2020-2042>.
- Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. "The Semantic Web." *Scientific American* 284 (5): 34–43. <https://www.lassila.org/publications/2001/SciAm.pdf>.
- Bizer, Christian, and Richard Cyganiak. 2009. "Quality-Driven Information Filtering Using the WIQA Policy Framework." *Journal of Web Semantics* 7 (1): 1–10. <https://doi.org/10.1016/j.websem.2008.02.005>.
- Bornmann, Lutz, and Rüdiger Mutz. 2015. "Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References: Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References." *Journal of the Association for Information Science and Technology* 66 (11): 2215–22. <https://doi.org/10.1002/asi.23329>.
- Daniel, Florian, Pavel Kucherbaev, Cinzia Cappiello, Boualem Benatallah, and Mohammad Allahbakhsh. 2018. "Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions." *ACM Computing Surveys* 51 (1). <https://doi.org/10.1145/3148148>.
- Färber, Michael, Basil Ell, Carsten Menne, Achim Rettinger, and Frederic Bartscherer. 2018. "Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO." *Semantic Web Journal* 9 (2): 77–129. Accessed March 15, 2023. <https://doi.org/10.3233/SW-170275>.
- Hussein, Hassan, Allard Oelen, Oliver Karras, and Sören Auer. 2022. "KGMM - A Maturity Model for Scholarly Knowledge Graphs Based on Intertwined Human-Machine Collaboration." In *From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries: 24th International Conference on Asian Digital Libraries, ICADL 2022*, 253–69. Hanoi, Vietnam: Springer-Verlag. https://doi.org/10.1007/978-3-031-21756-2_21.
- Johnson, Rob, Anthony Watkinson, and Michael Mabe. 2018. "The STM Report: An Overview of Scientific and Scholarly Publishing. 5th Edition." https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf.
- Oelen, Allard, Mohamad Yaser Jaradeh, Kheir Eddine Farfar, Markus Stocker, and Sören Auer. 2019. "Comparing Research Contributions in a Scholarly Knowledge Graph." In *Sci-Know 2019: Third International Workshop on Capturing Scientific Knowledge*, 21-26. <https://doi.org/10.15488/9388>.
- Oelen, Allard, Mohamad Yaser Jaradeh, Markus Stocker, and Sören Auer. 2020. "Generate FAIR Literature Surveys with Scholarly Knowledge Graphs." In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 97–106. <https://doi.org/10.34657/5212>.
- Pipino, Leo L., Yang W. Lee, and Richard Y. Wang. 2002. "Data Quality Assessment." *Communications of the ACM* 45 (4): 211–18. <https://doi.org/10.1145/505248.506010>.

Shotton, David. 2009. "Semantic Publishing: The Coming Revolution in Scientific Journal Publishing." *Learned Publishing* 22 (2): 85–94. <https://doi.org/10.1087/2009202>.

Wang, Richard Y., and Diane M. Strong. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems* 12 (4): 5–33. <https://doi.org/10.1080/07421222.1996.11518099>.

Wilkinson, Mark D, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 1–9. <https://doi.org/10.1038/sdata.2016.18>.

Zaveri, Amrapali, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. 2016. "Quality Assessment for Linked Data: A Survey." *Semantic Web* 7 (1): 63–93. <https://doi.org/10.3233/SW-150175>.

Zhang, Jing. 2022. "Knowledge Learning With Crowdsourcing: A Brief Review and Systematic Perspective." *IEEE/CAA Journal of Automatica Sinica* 9 (5): 749–62. <https://doi.org/10.1109/jas.2022.105434>.

Citation metrics and strategic mutations of scientific research: narratives and evidence

Luca Guerra^(a)

a) University of Brescia, <https://orcid.org/0000-0002-8903-7744>

Contact: Luca Guerra, luca.guerra@unibs.it

Received: 16 February 2023; Accepted: 30 September 2023; First Published: 15 January 2024

ABSTRACT

Following the diffusion of the management model promoted by the New Public Management also within universities, sectors until then endowed with particular protection, such as the academic one, were progressively subjected to new controls and constraints, which gradually took the form of quantitative surveys, with a growing role of citation metrics. These evaluation processes have given rise to various important critical positions at an international level.

If it is true that, according to Goodhart's law, "when a measure becomes an objective it ceases to be a good measure", it is even more significant to note that when a measure becomes an objective, what is measured ceases to be what it was before. The quantitative measurement of academic performance has in fact triggered forms of gaming such as to alter the very game of scientific research, its purposes and the forms of its sharing. In the article we will carry out a comparison between the forms of mutation and the narratives that accompany them to see to what extent we are legitimate today to talk about the transformation of scientific research into *strategic scientific research*.

KEYWORDS

Scientific misconducts; Gaming metrics; Scholarly communication; Research assesment; Bibliometrics.

Metriche citazionali e mutazioni strategiche della ricerca scientifica: narrative ed evidenze

ABSTRACT

A seguito della diffusione anche all'interno delle università del modello gestionale promosso dal New public management, settori fino ad allora dotati di una particolare protezione, come quello accademico, si sono progressivamente visti sottoposti a nuovi controlli e vincoli, che hanno progressivamente assunto la forma di rilevazioni quantitative, con un ruolo crescente delle metriche citazionali. Questi processi di valutazione hanno suscitato diverse e importanti prese di posizione critiche sul piano internazionale.

Se è vero che, secondo la "legge" di Goodhart "quando una misura diviene un obiettivo smette di essere una buona misura", ancora più significativo è rilevare che quando una misura diviene un obiettivo, il misurato smette di essere quello che era prima. La misurazione quantitativa delle performance accademiche ha innescato infatti forme di *gaming* tali da alterare il gioco stesso della ricerca scientifica, dei suoi scopi e delle forme della sua condivisione. Nell'articolo svolgeremo un confronto tra le forme della mutazione e le narrative che le accompagnano per vedere fino a quale punto siamo legittimati oggi a parlare della trasformazione della ricerca scientifica in *ricerca scientifica strategica*.

PAROLE CHIAVE

Mutazioni della ricerca scientifica; Metriche citazionali; Comunicazione scientifica; Valutazione della ricerca; Bibliometria.

Insorgenze

Da alcuni anni assistiamo a insorgenze planetarie avverso le forme della misurazione della ricerca, dei ricercatori e le correlate politiche di assegnazione dei fondi, di promozione nelle carriere accademiche e delle abilitazioni scientifiche. Queste mobilitazioni sono cresciute in particolare in seguito all'affermazione del new public management che, dall'inizio degli anni Novanta, dopo un contesto di crisi economica, ha cominciato a investire le prassi manageriali delle pubbliche amministrazioni e ha posto al proprio centro il tema dell'*accountability*. Proprio attorno a questa esigenza di valutazione e misurazione, da qualche tempo entrata nel cuore delle università, sono sorti fenomeni impreveduti, controfinalità e conflitti. Le forme del disagio sono molto cresciute negli ultimi anni all'interno della comunità dei ricercatori e hanno portato per l'appunto a forme organizzate di critica e di opposizione.

Una delle più importanti e influenti prese di posizione in particolare contro la valutazione centrata sugli indici bibliometrici è quella conosciuta come dichiarazione DORA¹, ovvero *la San Francisco Declaration on research assessment*, proposta durante il convegno annuale della Società Americana di Biologia Cellulare nel dicembre 2012. Nella dichiarazione si prende specificamente di mira l'*Impact factor* (IF) come fonte di perniciose conseguenze ad ampio raggio nel mondo della ricerca scientifica. In questa dichiarazione non si contesta tanto, su un piano generale, l'uso di metodi quantitativi² nella valutazione della ricerca, ma specificamente l'impiego distorto dell'IF, il primo e il più noto indicatore elaborato dall'Institut for Scientific Information (ISI) nel 1961 e impiegato su base sistematica nel Journal of citation report (JCR) dal 1974. Si tratta effettivamente di un indice per il quale – nota Biagioli (2020) – un articolo paradossalmente “nasce valutato”, già dotato di impatto, ancor prima che qualcuno l'abbia letto e ammesso che qualcuno mai lo leggerà. Eugene Garfield (1998) stesso, fondatore di ISI, ha ritenuto di assumere una posizione critica³ rispetto a come viene mal impiegata la sua creatura. Quell'indice – spiega – l'aveva inventato per aiutare i bibliotecari nella selezione dei titoli di periodici da sottoscrivere. In termini generali – potremmo pure considerare – è plausibile pensare a periodici molto citati come fonti scientifiche di alta qualità anche se questo non ci dà, per converso, la sicurezza che, per esempio, periodici orientati allo sviluppo di nuovi promettenti filoni di ricerca ancora poco frequentati e dunque con basso IF, contengano articoli di scarsa qualità.

Dunque, gli estensori di DORA, dopo avere specificato alcune ragioni di contrarietà al modo in cui viene impiegato l'IF, si rivolgono a cinque interlocutori: agenzie di finanziamento, istituzio-

¹ <https://sfdora.org>. Il sito consente la sottoscrizione online della Dichiarazione ed è diventato anche un riferimento nella raccolta di *case studies* relativi alla sperimentazione di nuove prassi di valutazione della ricerca.

² Si suggerisce per esempio di adottare al posto del solo IF, un set di indicatori numerici quali: impact factor a cinque anni, EigenFactor, SCImago, h-index (combinati con la valutazione della non meglio precisata “influenza su politiche e prassi”).

³ Per esempio, pur all'interno di una difesa del complessivo lavoro di ISI contro le critiche di due professori tedeschi, nel 1998 Garfield scriveva “The source of much anxiety about Journal Impact Factors comes from their misuse in evaluating individuals, e.g. during the Habilitation process”. Qui il testo integrale della lettera all'editore: [http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101\(6\)p413y1998english.html](http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101(6)p413y1998english.html).

ni, editori, organizzazioni che forniscono metriche, ricercatori affinché non facciano affidamento nelle loro rispettive attività all'IF e invitano tutti a esplorare nuovi indicatori quantitativi, a dare misura di impatto del singolo *articolo* (anziché del suo contenitore) e ad assegnare una maggiore importanza alla valutazione qualitativa. Da quel manifesto è nato il progetto operativo, svolto in collaborazione con il *Centre for Science and Technology Studies (CWTS)* di Leida, una delle più note agenzie di elaborazione di indici, chiamato *Tools to advance research metrics*⁴ (TARA).

Come opportunamente segnalava già Figà-Talamanca (2000), in un'aspra critica dell'IF (qualificato come "virus"), non solo è assai poco sensato utilizzare le citazioni di un ente astratto come l'"articolo medio"⁵ da nessuno mai scritto, ma deve anche essere preso in conto l'effetto trasformativo che il fatto della diffusione dell'IF ha prodotto e produce nel sistema editoriale e della distribuzione. L'orientamento preferenziale e concorrenziale a pubblicare sul nucleo delle riviste a maggiore IF determinava e determina infatti la concentrazione degli articoli più interessanti in quel nucleo, facendo avverare la previsione a mezzo della previsione stessa. Questo spostamento ha così determinato un *focus sempre maggiore sulle riviste*, ha cioè amplificato enormemente l'importanza del "contenitore" e ha portato, in fatto, a ridurre relativamente il *focus* sull'articolo. Questo accadimento ha generato, in breve tempo, una rigida gerarchizzazione delle testate e ha posto le basi per una *crescita sempre più poderosa del potere nelle mani degli editori*. Ulteriormente, quella che Robert Merton⁶ (1998) qualificava come "moneta" della scienza, cioè le citazioni che si impiegano per pagare tributo intellettuale, entra in una deriva iperinflattiva e perde valore. Non solo il numero delle citazioni-moneta esplose parallelamente all'impiego sempre più diffuso dell'IF, ma inizia a circolare moneta falsa⁷: ci si accorda tra autori, riviste, gruppi per citarsi a vicenda con abbondanza. Se riconosciamo che nell'atto stesso di citare già intervengono motivi strategici (per appoggiare certi autori e correnti o al contrario per cercare di emarginarli, pure talvolta per questioni di rivalità personale), dobbiamo riconoscere che attualmente la componente strategica delle citazioni ne risulta enormemente accresciuta, laddove esse non mirano più soltanto al riconoscimento intellettuale ma a condizionare incarichi, carriere e finanziamenti⁸.

Nella letteratura sulla valutazione della ricerca viene molto citata la legge di Goodhart che recita "quando una misura diviene un obiettivo, cessa di essere una buona misura"⁹. Il punto più impor-

⁴ <https://www.cwts.nl/projects/current-projects/tools-to-advance-research-assessment-tara>. Il Progetto TARA prevede la ricognizione di politiche e prassi in tema di valutazione della ricerca, in particolare, di area europea e statunitense, e di sviluppare approcci innovativi, in linea con le raccomandazioni DORA.

⁵ Ancor più in un contesto di asimmetria. Garfield, a sostegno della significatività dei suoi indici citazionali pur basati su un ristretto nucleo di riviste, poteva citare Bradford e il suo "moltiplicatore" (di applicazione settoriale) per cui risultava che gli articoli più "rilevanti" interni ad una disciplina sono concentrati in poche riviste. Similmente Lotka aveva rilevato che, all'interno dei settori scientifici, pochi autori contano per la gran parte degli articoli "rilevanti" di quei settori. Anche Garfield ha proposto la sua "legge di concentrazione" che garantirebbe significatività scientifica anche solo mediante l'indicizzazione di un numero ristretto di riviste selezionate.

⁶ In un contesto in cui il sociologo della scienza parla di riconoscimento simbolico tra pari e di citazioni, viene usata l'espressione "coin of the realm".

⁷ Mario Biagioli, in relazione a questo traffico di citazioni, parla di "post-production misconduct" (Biagioli 2016).

⁸ Si vedano ad esempio le considerazioni di Paul Wouters (2020).

⁹ L'espressione viene attribuita all'economista Charles Goodhart, che la riferiva agli obiettivi di politica monetaria del governo Thatcher. Si veda il libro di Muller (2018).

tante però da considerare è che *quando una misura diviene un obiettivo, il misurato cessa di essere quel che era prima*. L'IF ha infatti generato e diffuso comportamenti adattivi che gli anglosassoni chiamano *gaming behaviours*. Secondo la classica distinzione tra *play* e *game*, quest'ultimo è un gioco autotrasformativo, nel quale le regole divengono esse stesse oggetto del gioco. Le regole che governano i giochi della ricerca scientifica e della *scholarly communication* sono entrate così in una fase di mutazione, i cui contorni è opportuno cominciare a delineare.

A distanza di pochi anni, in seguito all'*International conference on science and technology indicators*, svoltosi a Leida nel settembre 2014, viene pubblicato nel 2015 su Nature il *Manifesto di Leida per le metriche della ricerca*¹⁰, articolato in dieci principi. La constatazione in premessa fa riferimento alla prassi ormai divenuta prevalente e "ossessiva" di valutare la ricerca a mezzo di dati, spesso male raccolti e male interpretati. I principi esortano a: utilizzare le metriche solo in funzione ausiliaria rispetto alla valutazione qualitativa condotta da esperti; correlare le prestazioni valutate alle finalità specifiche dei vari progetti di ricerca (di base, frontiera, sviluppo); non penalizzare le ricerche di interesse locale e quelle sociali ed umanistiche; rendere trasparenti indici e loro applicazioni; consentire ai valutati di verificare dati e analisi; utilizzare una *suit* di indicatori diversificati per ambito disciplinare e utilizzare normalizzazioni¹¹ di collocazione delle riviste per settore; valutare i ricercatori secondo più dimensioni qualitative del lavoro di ricerca; non fingere precisioni numeriche che sarebbero solo millantate; utilizzare *suit* di indicatori anziché uno solo, che rischia di condurre a comportamenti "sistemici"; esaminare e aggiornare regolarmente gli indici che devono correlarsi alle trasformazioni delle prassi di ricerca. In conclusione, gli estensori riconoscono che "research metrics can provide crucial information that would be difficult to gather or understand by means of individual expertise" e invitano i valutatori ad un saggio mix di dati quantitativi e valutazione qualitativa.

Alla metà del 2015 viene pubblicato il rapporto *The metric tide: report of the independent review of the role of metrics in research assesment and management* (Wilsdon et al. 2015). Si tratta di uno studio che sviluppa considerazioni sotto molteplici prospettive a partire dai risultati di una "call of evidence" nella quale la comunità accademica del Regno Unito era chiamata ad esprimere le proprie posizioni, in particolare in relazione al Research excellence framework (il processo di valutazione della ricerca accademica finalizzato all'efficiente allocazione delle risorse per la ricerca). Il gruppo di lavoro fa proprie le raccomandazioni di DORA, prende atto della "marea" crescente costituita dall'impiego sempre più diffuso e sistematico delle metriche (degli "indici", come preferiscono chiamarli) e, dopo avere evocato il rischio di una "new barbary", auspica l'adozione di un approccio pragmatico che conduca alla realizzazione di "metriche responsabili"¹². "Proceeding cautiously, in an exploratory way, seems an appropriate approach" (Wilsdon et al. 2015, 29) – scrivono gli autori. Poiché il rapporto tra metriche citazionali e giudizi di qualità è problematico in

¹⁰ <https://www.nature.com/articles/520429a>. In questo caso i redattori del Manifesto appartengono perlopiù a istituzioni di produzione di metriche o ad agenzie di valutazione e management della ricerca, i quali riconoscono il verificarsi di pre-occupanti "unintended consequences".

¹¹ In realtà, indici normalizzati per quartili nei vari settori disciplinari al fine della loro comparabilità già esistono sia in Web of science che in Scopus (naturalmente ogni indice è una "costruzione" e dunque la delimitazione dei *confini* delle discipline tra riviste e l'afferenza disciplinare degli articoli è un'opzione esercitata dai compilatori).

¹² È disponibile un blog per favorire il confronto all'indirizzo www.ResponsibleMetrics.org.

quanto assai variabile all'interno delle diverse discipline¹³, nel rapporto viene anche presentata l'"informed peer review" "the idea that the judicious application of specific bibliometric data and indicators may inform the process of peer review, depending on the exact goal and context of the assessment" (Wilsdon et al. 2015, 64). Ma anche questa proposta di valutazione deve essere assunta con cautela poiché – suggerisce il rapporto – sono varie le interpretazioni circa la sua concreta attuazione ed inoltre è difficile escludere che la *peer review* non sia condizionata, a sua volta, dalle metriche relative a prodotti scientifici oggetto di valutazione.

Anziché opporsi alla marea, quasi fosse un'evenienza naturale, allo scopo di evitare o minimizzare "strategic behaviours and gaming" e, ancor peggio, forme di "goal displacement" (formula mertoniana per cui, semplicemente, il fine della ricerca scientifica di produrre nuova conoscenza viene sostituito dal fine di produrre impatto bibliometrico), nel rapporto vengono definiti cinque principi che dovrebbero informare queste un po' misteriose entità che si chiamano "metriche responsabili". I principi proposti si articolano in: "robustness" (le metriche devono basarsi sulla migliore raccolta possibile di dati in termini di accuratezza e di ambito); "humility" (va riconosciuto che la valutazione quantitativa dovrebbe essere di supporto alla valutazione qualitativa condotta dagli esperti, e non sostitutiva); "transparency" (raccolta dei dati e processi di analisi dovrebbero essere mantenuti aperti e trasparenti, affinché i valutati possano controllare e verificare i risultati); "diversity" (va tenuto conto delle diversità tra ambiti e andrebbe utilizzato un range di indicatori che riflettano e sostengano una pluralità di percorsi di ricerca e di carriera); "reflexivity" (vanno riconosciuti e anticipati gli effetti potenziali e sistemici degli indicatori, che vanno aggiornati in risposta a quegli effetti).

Da allora si sono moltiplicati manifesti e pubbliche prese di posizione da parte di comunità scientifiche disciplinari, finanziatori, editori e vari altri soggetti interessati alla produzione e alla circolazione delle conoscenze scientifiche che, con accenti più o meno convergenti, prospettano i rischi dell'utilizzo delle metriche citazionali, sottolineano i pericoli di una crescente pressione all'impatto e alla produttività e auspicano l'adozione di diversi criteri di valutazione dei prodotti scientifici e in generale delle performance dei ricercatori, al fine di preservare l'integrità della scienza.

Un passo importante dal piano della denuncia a quello della proposta è stato recentemente compiuto dall'internazionale *Coalition for advancing research assesment* (CoARA)¹⁴ con l'*Agreement on reforming research assesment*¹⁵ (sottoscritto anche dall'Anvur alla fine del 2022) mediante il quale si apre infine un cantiere operativo che impegna i membri a condividere entro un anno dalla sottoscrizione le loro soluzioni operative e a fissare come data la fine del 2027 per verificarne idoneità ed efficacia. Articolato per principi e impegni, l'*agreement* invita a partire dai cantieri di TARA per le STEMM e da Humetrics¹⁶ (progetto finanziato dalla Mellon Foundation) per le HSS e in cui si combinano l'affermazione di principi generali con l'invito alla *rule of thumb* (ad un approccio

¹³ "The correlation between bibliometrics and peer review is weaker in most fields in the humanities, the applied fields, the technical sciences, and the social sciences. This is partly caused by lower coverage in the citation databases, but also by varying citation and publication cultures" (Wilsdon et al. 2015, 64).

¹⁴ Benché tra i fondatori vi siano la Commissione europea e la European university association, la Coalizione è internazionalmente aperta a tutte le organizzazioni impegnate nell'ambito della ricerca scientifica che ne sottoscrivono l'Agreement. Le decisioni vengono prese, su base paritaria, dai membri della General assembly che nomina Presidente e Steering board.

¹⁵ <https://coara.eu/agreement/the-agreement-full-text>.

¹⁶ <https://humetricshss.org/about>.

pragmatico). Il passaggio ad un piano operativo è certamente da giudicare positivamente e quasi sicuramente tale dispiegamento di uomini e mezzi porterà a nuove proposte, tuttavia almeno rileviamo come già ora la bibliometria¹⁷ metta a disposizione una vasta gamma di indici (variamente “responsabili”), che il riferimento allo SCOPE framework¹⁸ (estremamente impegnativo e laborioso per gli istituti che lo vorranno adottare) rischia di creare una congerie di ulteriori indici, e che il *peer reviewing* vi compaia come una sorta di panacea, senza ravvisarne limiti e costi¹⁹.

Mutazioni e casi teratologici

Nel Manifesto di Leida si fa riferimento a forti pressioni sui dottorandi perché pubblichino in gran fretta e su riviste ad alto IF al fine di attirare finanziamenti, si cita anche il caso di sociologi spagnoli che per spuntare indici più elevati di IF si concentrano su modelli astratti o su dati statunitensi.

Quanto in generale l’obiettivo dell’impatto orienta e plasma la ricerca scientifica stessa, nei suoi contenuti e nei suoi stili? In che misura muta il processo di produzione e condivisione della conoscenza, nelle sue prassi, nei suoi oggetti e nelle sue finalità? Fino a quale punto si può parlare di una *trasformazione della ricerca scientifica in ricerca scientifica strategica*?

Vediamo alcune mutazioni o almeno apparenti tali. Retraction watch, che cura un database di articoli ritirati (e anche solo corretti) dopo la pubblicazione, ci aggiorna di un numero crescente di pubblicazioni scientifiche che sono state ritirate. Oransky, co-fondatore di Retraction watch, si riferisce a un “retraction boom” (Oransky 2022) registrato negli ultimi anni. L’autore ci dice che nel 2010 gli articoli ritirati erano in media 45 al mese e che nel 2021 sono stati quasi 300. Il tasso complessivo di ritiro è un po’ inferiore allo 0,1% ma aggiunge che, sulla base di alcuni rapporti e indagini condotti dai “cacciatori” di pubblicazioni errate o fraudolente, il tasso di articoli che andrebbero ritirati sarebbe di circa il 2%. A cosa è dovuta questa crescita dei ritiri? Ad un progressivo peggioramento del livello qualitativo della ricerca? Non parrebbe: per Oransky la comunità dei ricercatori è diventata sempre più attenta alle ricerche falsificate o semplicemente erranee e dunque è stata in grado di attivare un più efficiente sistema di autocontrollo. Dunque, se il numero delle pubblicazioni ritirate cresce, non è tanto perché crescono quelle fasulle ma soprattutto perché vengono più efficacemente intercettate. Resta tuttavia da registrare il fatto inquietante che un numero niente affatto trascurabile di ricercatori falsifica scientemente e antiscientificamente i dati per poter pubblicare e guadagnare impatto. La cosa ulteriormente preoccupante è che spesso “researches continue to cite retracted papers” e che il processo di ritiro “remains comically clumsy, slow and opaque – often taking years, if it ever happens at all” (Oransky 2022, 9). Può essere

¹⁷ Per un’introduzione in lingua italiana si possono vedere Nicola De Bellis (2014) e Simona Turbanti (2018).

¹⁸ <https://inorms.net/scope-framework-for-research-evaluation>. L’International network of research management societies è un’organizzazione internazionale che intende promuovere iniziative e progetti, condivisione di buone prassi e conoscenze tra gli istituti che si occupano di management della ricerca.

¹⁹ Si vedano per esempio: Brezis e Birukou (2020); Dondio et al. (2019); Richard Smith (2006). Biagioli parla di “fake peer-review rings” a indicare la permeabilità del modello da parte di condotte fraudolente miranti allo scambio “collaborativo” di citazioni. Restano ancora assai istruttive le considerazioni di Figà-Talamanca che ricostruisce per l’ambito biomedico italiano l’alternativa *peer reviewing*/metriche citazionali in termini di conflitto di potere intergenerazionale interno all’accademia.

di interesse sapere che esiste una correlazione positiva (Fang and Casadevall, 2011) tra frequenza di ritiro e IF, interpretabile con il fatto che su quelle riviste si pubblica più facilmente quando in questione ci siano scoperte di grande rilevanza (e dunque gli autori che vi vogliono pubblicare sentono l'incentivo a gonfiare indebitamente o anche fraudolentemente la portata dei loro risultati) e anche perché esiste un tasso di controllo e sorveglianza sugli articoli, dopo la pubblicazione, molto più elevato rispetto agli altri periodici. Sul JAMA è apparso uno studio (Gaudino et al. 2021) specifico per l'ambito biomedico relativo al periodo 1971-2020 da cui è risultato che il tasso delle pubblicazioni ritirate è crescente, specialmente dal 1980 al 2014 (in particolare dopo il 2004) ed inizia a calare a partire dal 2015, laddove il numero assoluto è in continua crescita almeno fino a circa il 2018. Nell'ambito degli studi oncologici Pantziarka e Meheus (2019), oltre a confermare l'andamento crescente dei ritiri per questa specializzazione, evidenziano come gli articoli esaminati avessero grande impatto potenziale anche su pazienti colti in cerca di trattamenti alternativi a quelli convenzionali, con connesso aumento del rischio di mortalità. La buona notizia è che il tempo intercorrente tra la pubblicazione di un articolo di ambito oncologico e la sua identificazione come erroneo o fraudolento è in significativo costante calo, con riduzione del rischio che questi articoli entrino nel circolo delle citazioni. Come risulta da diverse note sul blog di Retraction watch, in più di un caso, le industrie farmaceutiche avevano preso in considerazione per la produzione di farmaci alcune ricerche smascherate poi come erranee o fraudolente.

La fantasia dei ricercatori impegnati nel *goal displacement* è molto prolifica. Tra le modalità entrate maggiormente in voga, per esempio, c'è il "p-hacking" (conosciuto anche come "dredging" o "snooping") che consiste in almeno due varianti: operare molti test statistici sui dati e riportare solo quelli che danno risultati significativi di correlazione (i falsi positivi) oppure operare test statistici su dati presentati come casuali e invece raccolti sulla base di alcuni criteri non dichiarati. Un altro modo scorretto di produrre risultati è lo HARKing (*hypothesizing after the results are known*) consistente nel presentare come ipotesi di correlazione successiva alla raccolta e ai test sui dati quanto invece è ipotesi precedente, assecondando così il "publication bias", laddove risulterebbe meno profittevole pubblicare risultati che diano nullità di rilevanza statistica. Un altro preoccupante fenomeno all'interno del mondo della ricerca è quanto viene definita "crisi della riproducibilità" che mette per l'appunto in questione uno dei principi della scientificità stessa e dovuto a diversi fattori, di cui è difficile dire quanti siano intenzionali e malevoli e quali no. La mancanza o la parzialità dei dati messi a disposizione, l'impiego di licenze che inibiscono od ostacolano il riuso, l'applicazione di software proprietari o con versione di rilascio non dichiarata: tutto ciò incide negativamente sulla possibilità del controllo e della validazione dei risultati da parte di altri gruppi di ricerca. Altre forme perniciose per l'integrità della scienza derivano dalla diffusione dei "predatory journals"²⁰, ovvero di quelle riviste che millantano crediti scientifici, comitati editoriali illustri (i cui membri spesso ignorano di farne parte), rigore nel *peer reviewing*, livelli di IF inverosimili o altri indici fantasmagorici interamente inventati, riconoscimenti inter-

²⁰ In realtà, pochi sono i casi in cui esistono sentenze definitive di condanna per falso nei confronti di editori di riviste scientifiche. Esiste una zona grigia assai ampia entro cui gli editori possono adottare condotte predatorie o di spamming e ugualmente pubblicare anche articoli scientifici validati. Verrebbe però forse più facile identificare quali siano gli editori *non predator*, date le esose politiche di prezzo (ed ora, in seguito all'affermazione degli "accordi trasformativi", politiche di APC) da parte di editori blasonati che legalmente minano il processo di condivisione della conoscenza scientifica.

nazionali etc., laddove si tratta in realtà di riviste che, in diversi casi, neppure svolgono il lavoro di controllo e selezione delle pubblicazioni ma mirano semplicemente ad incassare quanti più introiti dall'*article processing charge* (APC) secondo il modello pay-to-play. Naturalmente ci si può legittimamente chiedere: chi sarebbero queste prede così indifese e sprovvedute? Forse è più verosimile immaginare che tali periodici siano, non di rado, opportunisticamente utilizzati dai ricercatori non tanto per pubblicare articoli finalizzati alla lettura ma al mero incremento del punteggio bibliometrico. Sul web viene comunque aggiornata la straordinariamente lunga Bealls'list²¹ dei "potential predatory journals and editors". Esiste poi quell'emergente multiforme insieme di "post-production misconducts" (ipoteticamente senza pregiudizio della qualità intrinseca della ricerca) che spazia dalla vera e propria frode al comportamento adattivo e che muta al variare delle metriche, a cui è associabile quella esplosione citazionale a cui abbiamo già accennato. Si tratta del circolo del "never ending tuning" (degli indici) e "hacking" (degli stessi) sul quale Biagioli (2016) insiste molto, rappresentandolo però quasi come un processo che si autogenera, senza – a nostro giudizio – richiamare adeguatamente il fatto che quel circolo si salda ogni volta perché un ricercatore viene meno ai principi fondativi dell'integrità della ricerca e della pubblica condivisione dei suoi risultati (poiché, come noto, non esiste virtualmente nulla che non possa essere manipolato).

Frequenti, ancora, sono state le denunce di casi per i quali l'autore aveva suggerito come valutatori alcuni reviewer e comunicato indirizzi mail di cui questi stesso era titolare.

Resta poi il caso, quasi incredibile, del sw SCIGen, inventato da alcuni goliardi studenti del MIT per creare pseudo articoli di computer science dall'aspetto a prima vista credibile, e che è stato utilizzato in particolare da alcuni accademici cinesi per pubblicare articoli "a getto" (pubblicati anche da editori blasonati). Il più recente e sofisticato ChatGPT, bot basato sull'intelligenza artificiale, promette nuove chicche pseudoscientifiche essendo in grado, oltre a rispondere a domande, di produrre testi, codici sw, contratti, immagini etc. Che il sistema dei controlli e delle metriche possa infatti essere ampiamente giocato, è testimoniato anche dal signor Ike Antkare (I can't care) dotato di un h-index superiore a quello di Einstein, pur senza essere mai esistito, se non nella fantasia dell'impertinente Cyril Labbé.

Meno preoccupante ma significativa è la pratica denunciata del "salami-slicing" consistente nel frazionare in più articoli la comunicazione dei risultati al fine di aumentare la produttività ma al costo di rendere più laborioso e farraginoso il lavoro di ricostruzione da parte del lettore. Va tuttavia evidenziato come statisticamente risulti un trend di lunga lena nella crescita media della lunghezza delle pubblicazioni, talché dal *salami-slicing* possiamo inferire un'"inflazione" retorica e un impoverimento del contenuto (e non una riduzione della lunghezza degli articoli). La pratica crescente della *co-authorship*²², ulteriormente, per cui "the average number of authors of publications in the scholarly literature continues to increase" (Wilsdon et al. 2015, 32) è ancora più ambigua da interpretare perché può costituire risposta all'incentivo di mostrare alti livelli individuali di prolificità scientifica aumentando strategicamente i gruppi di ricerca con cui si collabora, oppure può indicare la scelta di una collaborazione meramente figurativa e negoziata come risposta

²¹ <https://beallslist.net>.

²² Probabilmente il record di co-autorialità spetta ad un articolo pubblicato nel 2014 dalla *Physical review letters* firmato da 5.154 ricercatori.

alla pressione alla pubblicazione oppure ancora indicare un aumento del livello di collaborazione proprio della ricerca attuale (multidisciplinare e transdisciplinare).

Alcuni autori ritengono che, complessivamente, le mutazioni occorse al mondo della ricerca abbiano profondamente alterato la fisionomia del processo della *scholarly communication* nell'arco di pochi decenni, più di quanto sia accaduto dalla metà del Seicento. Mark Edwards e Siddharta Rey, in un articolo dal titolo suggestivo “Science is broken”, dopo avere considerato che “Competition among researchers for funding has never been more intense, entering an era with the worst funding environment in half a century”, scrivono (riferendosi prudentemente però ad un ipotetico futuro): “We believe that this system presents a real threat to the future of science. If immediate action is not taken, we risk creating a corrupt professional culture akin to that revealed in professional cycling (ie, 20 out of 21 Tour de France podium finishers during 1999-2005 were conclusively tied to doping), where an uncontrolled perverse-incentive system created an environment in which athletes felt that they had to cheat to compete” (Edwards and Rey, 2017). I risultati di un questionario sottoposto ad alcune migliaia di ricercatori negli Stati Uniti portavano gli autori a ipotizzare che “certain features of the research working environment may have unexpected and potentially detrimental effects on the ethical dimensions of scientists’ work” (Martinson 2005). Ancora – scrivono altri autori – “some violations are considered to be related to the increased pressures to publish” (Hayer 2013).

Parallelamente a queste denunce allarmate e allarmanti, sono sorte, accanto ad istituzioni consolidate come per esempio l’US Office of Research Integrity, iniziative spesso informali gestite da gruppi di ricercatori (talvolta anonimi e con il gusto dello sberleffo) come Retraction watch, PubPeer e altri blog in funzione di “watchdogs” proprio per intercettare queste inedite forme di *gaming*. È tornata conseguentemente di grande attualità anche la questione dell’*etica* della scienza, della responsabilità e dell’integrità di coloro che vi partecipano. Un recente documento particolarmente significativo è costituito dai cinque Principi di Hong Kong (Moher et al. 2020) come guida per i ricercatori e per i valutatori al fine di preservare l’“integrità” della scienza. L’ “Allea European code of conduct for research integrity” propone principi e buone prassi per la preservazione della libertà della ricerca e la promozione di forme di autoregolazione e autocorrezione²³. Lo “Slow science manifesto”²⁴ denuncia i rischi per la qualità della ricerca e richiama alla natura di *common* della conoscenza, oggi considerata a rischio a causa delle recenti forme di science management. Le linee guida del Committee on publication ethics²⁵ (COPE) stabiliscono alcuni principi etici a cui dovrebbe inderogabilmente informarsi la *scholarly communication*.

Il tema dell’etica della scienza è probabilmente oggi più radicale di quello che generalmente si ritiene e – a nostro giudizio – costituisce il nucleo più importante e ancora in gran parte opaco interno a quanto in Italia oggi si definisce “terza missione” delle università. Sottratta alla retorica

²³ <https://allea.org/code-of-conduct>. L’ALLEA (All European academies) è una federazione di una cinquantina di istituti accademici europei, particolarmente impegnata nel miglioramento delle politiche di gestione della ricerca e del public engagement degli scienziati, sia in ambito strettamente scientifico che in quello umanistico.

²⁴ <https://slowscience.be/the-slow-science-manifesto-2>. Il movimento, che ha tra i propri fondatori l’epistemologa Isabelle Stengers, sostiene che una scienza governata dalle misurazioni di performance non solo rischia di smarrire la propria vocazione alla scoperta non strettamente finalizzata, ma anche di trasformarsi in una macchina finanziaria preminentemente finalizzata alla raccolta di fondi.

²⁵ <https://publicationethics.org>.

e a certo paternalismo, che spesso accompagnano le riflessioni sulla terza missione e anche sulla *citizen science*²⁶, agli scienziati si chiede un'interrogazione radicale sul proprio esercizio, sulle conseguenze che esso produce nella società, nel pianeta e in generale su tutti gli oggetti di significato, fino a interrogare in radice il metodo scientifico stesso, nelle sue secolari e congiunte strutture epistemologiche del disinteresse e della natura meramente osservativa.

Produzione di dati e cultura *publish or perish*

Abbiamo detto, all'inizio, che l'applicazione dei principi del new public management e in particolare dell'*accountability* ha avuto un ruolo sostanziale nel riplasmare strutture e prassi proprie delle pubbliche amministrazioni. Almeno un altro aspetto va però messo in linea di conto in questa nostra ricostruzione che vede l'impiego sempre crescente di metodi quantitativi nella raccolta, elaborazione e valutazione dei dati. E cioè il fatto che, per l'appunto, mai come oggi ci siamo ritrovati sommersi in un gigantesco mare di dati o, meglio, mai come oggi le società del terziario avanzato hanno *prodotto* tale quantità industriale di dati. Non è la sede per entrare nel merito della nozione di "dato", che per lo più viene inteso in forma oggettivistica, ma è bene almeno avere contezza che il dato non è mai un oggetto, bensì una *relazione*, si trova cioè collocato nel punto di precipitazione di un complesso intreccio di pratiche, operazioni e tracciamenti che lo mettono e lo tengono all'opera. Il dato non sta, per richiamarci ad una celebre affermazione di Foucault, in un limbo, in attesa di un atto di attenzione e dunque la nota formuletta "dato-informazione-conoscenza" è futile, poiché intende ingenuamente la relazione come "sopravveniente" al dato e neppure ravvisa il fatto che il dato già incorpora molta conoscenza. Questa crescita esponenziale di dati è dovuta principalmente al fatto che ai nostri gesti sono state connesse scritture elettroniche automatiche e dunque, nell'esplicarsi, essi producono registrazioni sempre più fitte e "granulari", molto spesso senza che ci sia né intenzione né consapevolezza. Oltre ad essere automatiche, tali scritture avvengono a costi molto bassi e dunque si espandono con grande facilità a formare quella società che variamente è definita "audit society" (Power 1997) o "società della sorveglianza" e che – ancora per richiamarci alla prospettiva foucaultiana, vede *sovrapporsi* razionalità disciplinare (basata sulla registrazione e sul controllo del dettaglio e che si avvale di "quadrettature" (*quadrillages*) epistemologiche, organizzative, spaziali, architettoniche, elettroniche) e razionalità biopolitica che individua e plasma normali e trend statistici di ordine sociale, performativo-ergonomico e biologico). Le nuove tecnologie rendono possibile *contestualmente* il dettaglio e il massivo e dunque offrono al computo (agli algoritmi) un raggio di azione sociale inedito e crescente²⁷. Dunque, per inquadrare la questione della valutazione, non ci si può limitare alle metriche citazionali e alle performance dei ricercatori, solo perché investono gruppi sociali perlopiù privilegiati e maggiormente protetti come quelli dell'accademia.

²⁶ Nelle dichiarazioni di principio della *European citizen science association* e dell'americana *Citizen science association* si trovano esortazioni a considerare la ricerca a cui partecipano gli appassionati di scienza come di pari livello della ricerca condotta internamente agli istituti di ricerca. Basta però leggersi gli statuti di alcune iniziative specifiche, e si ritrova pressoché sistematicamente (e – a nostro giudizio – inevitabilmente) una netta distinzione tra *researchers* e *practitioners*. È bastata una fetta del salame piccante *chorizo*, recentemente spacciata per un'immagine di Proxima Centauri, a raffreddare gli entusiasmi sulla *crowd science*.

²⁷ Quanto nel marketing viene definito *massive customization*

Tale processo capillare ed invasivo, nel mondo della ricerca, ha iniziato ad innescare condotte elusive, resistenze, giochi.

Da tempo si sostiene che nell'accademia si sia introdotta una cultura ipercompetitiva legittimata dall'assunto che quanto più intensa è la competizione tra ricercatori e tanto maggiori saranno i risultati in termini di produttività e miglioramento della qualità della ricerca. Molta letteratura – viceversa – sostiene che quanto essa definisce *cultura del publish or perish* (a cui andrebbe aggiunto *impact or perish*) produce effetti opposti a quelli attesi e genera incentivi perversi che, anziché migliorare l'impresa scientifica, la mina nei fondamentali, innesca mutazioni e degenerazioni. Il rischio – come si è detto – è che dal *publish or perish* esca una *broken science*. Anche sul piano del vissuto si raccolgono convergenti testimonianze. Per esempio, i risultati di un questionario sul tema sottoposto a ricercatori olandesi di ambito biomedico porta gli autori della ricognizione a considerare che “substantial proportion of the responding medical professors judge publication pressure as having become excessive, and a substantial part believes that this affects the validity and credibility of medical science” (Tijdink et al. 2013). Secondo le testimonianze, il problema si sposta anche sul piano del proprio livello di benessere e di salute, che ha ancora, a sua volta, conseguenze sul piano dell'efficienza scientifica, clinica e didattica. Scrivono infatti ancora gli autori “Apart from general effects on medical science and practice, our study suggests that excessive publication pressure has detrimental effects on personal well-being. [...] publication pressure causes them to develop a cynical view on medical science, and may be associated with increased risk of developing burn-out” (Tijdink et al. 2013).

Non c'è ragione di dubitare che quelle testimonianze circa un crescente disagio personale e collettivo non siano autentiche e conseguenti a trasformazioni effettive. Eppure non è affatto facile riconoscere un sicuro e diretto legame tra uso delle metriche citazionali, cultura POP e condotte scientifiche discutibili o fraudolente. In un modello comportamentale molto complesso (e molto astratto), già nel 2011 i ricercatori Lacetera e Zirulia mettevano in questione il legame causale tra cultura POP e condotte manipolatorie o fraudolente. Anzi, le pressioni all'impatto e alla produttività e l'alto livello di competizione possono sortire – ad avviso degli autori – un effetto dissuasivo rispetto alle condotte scorrette. Leggiamo: “This pressure can actually serve as a powerful mechanism to deter fraud as it increases the incentives of peers to scrutinize each other's work. This is due both to the increase in the reward from discovering fraud and to the anticipation of the author's higher temptation to cheat” (Lacetera e Zirulia 2011, 588). Più recentemente, Fanelli, sulla base di un ampio studio²⁸ statistico che si concentra sui primi 15 anni di carriera di un ampio numero di ricercatori, tende ad escludere alcuni effetti del POP e conclude “In sum, according to our findings, scientists today are not publishing, on an individual basis, at higher rates than their colleagues in the 1950s” (Fanelli 2020, 117). È ulteriormente difficile discriminare tra comportamenti (subiti) indotti dalla pressione istituzionale del POP (per cui o hai elevato impatto e alta produttività o sei marginalizzato) e quelli invece (cercati) motivati soggettivamente dal careerismo e dalla volontà di incamerare bonus (opzione lasciata alla libertà dei singoli). Scrive Fanelli “The pressure to publish hypothesis is not supported because the only significant patterns we detected in these and previous analyses support a role for cash incentive policies, which do not constitute

²⁸ In relazione a questi studi si è impiegato il pomposo nome di *metascience*, laddove più prosaicamente si tratta dell'applicazione di metodi statistici alla *scholarly communication*.

real ‘pressures’ on researchers, but rather an incentive that is misaligned with the scientific norms of disinterestedness”²⁶ (Fanelli et al. 2022, 15). È vero che le rilevazioni di Fanelli riguardano un limitato campione di riviste, si concentrano su casi specifici di cattive condotte, concernono circoscritti ambiti disciplinari, selezionano (necessariamente) un numero limitato di variabili, ipotizzano “intenzioni” e dunque non sono né interamente generalizzabili né conclusive, tuttavia ci ammoniscono a non accedere a facili semplificazioni sui fenomeni in atto nell’ambito della ricerca e della comunicazione scientifica.

Aggiungiamo che quanto abbiamo definito insorgenze riguarda piccole *minoranze* di ricercatori: per esempio la dichiarazione DORA, tra i documenti più conosciuti, ad oggi risulta sottoscritta da circa 20.000 singoli ricercatori e 2.700 organizzazioni, numeri niente affatto trascurabili, laddove però il numero globale dei ricercatori è di poco inferiore ai nove milioni²⁹. Al netto delle ingenuità di alcuni commentatori paladini del *peer reviewing* e fieramente ostili alle metriche, è bene sapere che questa è anche una partita *interna* alla comunità scientifica, non solo una battaglia di puri principi. La misura del reciproco peso tra le due modalità di valutazione si polarizza entro un campo di forze e può determinare mutamenti per l’appunto in termini di carriere, distribuzione di fondi, assegnazione di incarichi etc.

Non possiamo pertanto semplicisticamente dire che competizione, misurazione delle performance, impiego delle metriche citazionali abbiano in sé effetti deleteri. Quanto possiamo ipotizzare, dai preoccupanti segnali di talune mutazioni, dalle molteplici insorgenze e dalle testimonianze di disagio, è che ci stiamo verosimilmente appressando ad una *soglia* (forse già in più ambiti oltrepassata) che può indurre la trasformazione della ricerca scientifica in ricerca scientifica strategica. È allora oggi il tempo dello studio attento e critico di *scholarly production* e *scholarly communication* e non delle dichiarazioni caricaturali su burocrati trinariciuti in cospirazione contro la scienza.

²⁹ Dato ricavabile da Science business del 2021 <https://sciencebusiness.net/news/number-scientists-worldwide-reaches-88m-global-research-spending-grows-faster-economy>.

Riferimenti bibliografici

- Biagioli, Mario. 2016. “Watch out for cheats in citation game”. *Nature* 535: 201. <https://doi.org/10.1038/535201a>.
- Biagioli, Mario. 2020. “Introduction”. In *Gaming the metrics: misconduct and manipulation in academic research*, edited by Mario Biagioli and Alexandra Lippman, 1-24. Cambridge, The Mit press. <https://doi.org/10.7551/mitpress/11087.003.0006>.
- Biagioli, Mario and Alexandra Lippman, ed. 2020. *Gaming the metrics: misconduct and manipulation in academic research*. Cambridge: The Mit press. <https://mitpress.mit.edu/9780262537933/gaming-the-metrics>.
- Brezis, Elise S., and Aliaksandr Birukou. 2020. “Arbitrariness in the Peer Review Process”. *Scientometrics*, 2123: 393–411. <https://doi.org/10.1007/s11192-020-03348-1>.
- Cari-Ann, Hayer, Mark A. Kaemingk, Jason J. Breeggemann, Daniel Dembkowski, and David Deslauriers. 2013. “Pressures to publish: catalysts for the loss of scientific writing integrity?”. *Papers in Natural Resources* 689. <https://doi.org/10.1080/03632415.2013.813845>. <https://digitalcommons.unl.edu/natrespapers/689>.
- Council of the European Union. 2022. “Council conclusions on research assessment and implementations of Open science. Annex 10126/22.” <https://data.consilium.europa.eu/doc/document/ST-9515-2022-INIT/en/pdf>.
- De Bellis, Nicola. 2014. *Introduzione alla bibliometria*. Roma: AIB.
- Dondio, Pierpaolo, Niccolò Casnici, Francisco Grimaldo, Nigel Gilbert, and Flaminio Squazoni. 2019. “The ‘Invisible hand’ of peer review: the implications of author-referee networks on peer review in a scholarly journal”. *Journal of Informetrics* 13: 708–716. <https://doi.org/10.1016/J.JOI.2019.03.018>.
- Edwards, A. Mark, and Roy Siddhartha. 2017. “Science is broken. Perverse incentives and the misuse of quantitative metrics have undermined the integrity of scientific research”. *Aeon*. <https://aeon.co/essays/science-is-a-public-good-in-peril-heres-how-to-fix-it>.
- Fanelli, Daniele. 2020. “Pressures to publish: what effects do we see? In *Gaming the metrics: misconduct and manipulation in academic research*, edited by Mario Biagioli and Alexandra Lippman, 111-122. Cambridge: The Mit press. <https://doi.org/10.7551/mitpress/11087.003.0011>.
- Fanelli, Daniele, Matteo Schleicher, Ferric C. Fang, Arturo Casadevall, and Elisabeth M. Bik. 2022. “Do individual and institutional predictors of misconduct vary by country? Results of a matched-control analysis of problematic image duplications”. *PLoS One* 17 (3) e0255334. <https://doi.org/10.1371/journal.pone.0255334>.
- Fang C., Ferric, and Arturo Casadevall. 2011. “Retracted science and the retraction index”. *Infective Immunology* 79 (10): 3855-9. <https://doi.org/10.1128/IAI.05661-11>.
- Figà-Talamanca, Alessandro. 2000. “L’impact factor nella valutazione della ricerca” [Bozza dell’intervento, Bologna 27 giugno 2000]. http://siba.unipv.it/biblioteche/info/SISSCOWEB_A_Fig_Talamanca_L%27impact_factor_nella_valutazione_della_ricerca_SISSCOWEB.htm.

- Garfield, Eugene. 1998. "The impact factor and using it correctly". *De Unfallchirurg*, 48 (2): 424. http://garfield.library.upenn.edu/papers/derunfallchirurg_v101%286%29p413y1998english.html.
- Gaudino Mario, Brice N. Robinson N., Katia Audisio, Mohamed Rahouma, Umberto Benedetto, Paul Kurlansky, and Stephen E. Fremes. 2021. "Trends and characteristics of retracted articles in the biomedical literature, 1971 to 2020". *JAMA international medicine* 181 (8): 1118-1121. <https://doi.org/10.1001/jamainternmed.2021.1807>.
- Hicks Diana, Wouters Paul, Waltman Ludo, De Rijcke Sarah, Rafols Ismael. 2015. "Bibliometrics: The Leiden Manifesto for research metrics". *Nature*. 520: 429-431.
- Lacetera, Nicola, and Lorenzo Zirulia. 2011. "The economics of scientific misconduct". *The Journal of Law, Economics, and Organization*. 27: 568–603. <https://doi.org/10.1093/jleo/ewp031>.
- Martinson, C. Brian, Melissa S. Anderson, and Raymond de Vries. 2005. "Scientists behaving badly". *Nature*. 435: 737-738.
- Merton, K. Robert. 1998. "The Matthew effect in science, II: cumulative advantage and the symbolism of intellectual property". *Isis*. 79, 4: 606-623.
- Moher, David, Lex Bouter, Sabine Kleinert, Paul Glasziou, Sham Mai Har, Virginia Barbour, Anne-Marie Coriat, Nicole Foeger, and Ulrich Dirnagl. 2020. "The Hong Kong Principles for assessing researchers: fostering research integrity". *PLoS Biology* 18 (7). <https://doi.org/10.1371/journal.pbio.3000737>.
- Muller, Z. Jerry. 218. *The tyranny of metrics*. Princeton: Princeton university press.
- Naujokaitytė, Goda. 2021. "Number of scientists worldwide reaches 8.8M, as global research spending grows faster than the economy" in *Science business*. <https://sciencebusiness.net/news/number-scientists-worldwide-reaches-88m-global-research-spending-grows-faster-economy>.
- Oransky, Ivan. 2022. "Retractions are increasing, but not enough". *Nature* 608: 9. <https://doi.org/10.1038/d41586-022-02071-6>.
- Pantziarka, Pan, and Lydie Meheus. 2019. "Journal retractions in oncology: a bibliometric study". *Future oncology* 15 (31): 3597-3608. <https://doi.org/10.2217/fon-2019-0233>.
- Power, Michael. 1997. *The Audit Society*. Oxford: Oxford University Press.
- Smith, Richard. 2006. "Peer review: a flawed process at the heart of science and Journals". *Journal of the Royal society of medicine* 99 (4): 178-182.
- Tijdink, K. Joeri, Anton C.M. Vergouwen, and Yvo M. Smulders. 2013. "Publication pressure and burn out among Dutch medical professors: a nationwide survey". *PLoS One* 8 (9): e7338. <https://doi.org/10.1371/journal.pone.0073381>.
- Turbanti, Simona. 2018. *Strumenti di misurazione della ricerca*. Milano: Editrice Bibliografica.
- Wilsdon, James, ed. 2015. *The metric tide: report of the independent review of the role of metrics in research assessment and management*. Thousand Oaks: SAGE publications. <https://doi.org/10.13140/RG.2.1.4929.1363>. <https://sk.sagepub.com/books/the-metric-tide>.

Wouters, Paul. 2020. "The mismeasurement of quality and impact". In *Gaming the metrics: misconduct and manipulation in academic research*, edited by Mario Biagioli and Alexandra Lippman, 66-75. Cambridge, The Mit press. <https://doi.org/10.7551/mitpress/11087.003.0006>.

The Return of the Panther. A university laboratory for the construction of an oral archive

Sara Manali^(a)

a) University of Palermo, <https://orcid.org/0000-0001-5289-0269>

Contact: Sara Manali, sara.manali@unipa.it

Received: 29 June 2023; Accepted: 03 November 2023; First Published: 15 January 2024

ABSTRACT

The article aims to present the activities of the Archival Laboratory at the University of Palermo, which involved the creation of an oral archive about the university movement known as “Pantera” (1989-90). Starting from some general observations about sound documents, oral sources, and sources for studying movements, the contribution reaches the broader project within which the Laboratory’s activities unfold: “Limone Lunare,” the archive of oral sources about political and social movements in contemporary Sicily. The article is completed with descriptions of the collected interviews.

KEYWORDS

Oral Archives; Invented Archives; Pantera movement; University teaching.

Il ritorno della Pantera. Un laboratorio universitario per la costruzione di un archivio orale

ABSTRACT

L'articolo intende presentare le attività del Laboratorio di Archivistica dell'Università di Palermo che hanno previsto la costruzione di un archivio orale sul movimento universitario della Pantera (1989-90). Partendo da alcune osservazioni generali in ordine ai documenti sonori e alle fonti orali, il contributo passa in rapida rassegna le fonti per lo studio dei movimenti fino ad approdare al più ampio progetto entro il quale si sviluppano le attività del Laboratorio: Limone Lunare, l'archivio di fonti orali sui movimenti politici e sociali siciliani di età contemporanea. Completano l'articolo le descrizioni delle interviste raccolte.

PAROLE CHIAVE

Archivi orali; Archivi inventati; La Pantera; Università.

Introduzione

Negli ultimi decenni i documenti sonori hanno fatto il loro ingresso nel vasto panorama delle tipologie documentarie da gestire, usare, tutelare. Se è vero che l'approccio pluridisciplinare agli archivi è quello che più restituisce la natura di essi, questo è vero anche – o forse maggiormente – per i documenti che mettono il suono e/o la voce al proprio centro. L'universo degli archivi sonori e delle fonti orali¹ (Stamuli 2019) è necessariamente e fortunatamente multidisciplinare e transdisciplinare poiché al centro e a cavallo tra diverse scienze. E in questa direzione vanno anche alcuni prodotti editoriali che ne trattano il tema, perché sarebbe epistemologicamente insufficiente guardare il prisma da un'unica sua faccia².

Le diverse ragioni che sottendono alla nascita di archivi sonori, per motivi che la disciplina archivistica rintraccia nella natura giuridica dei soggetti produttori, determinano una parcellizzazione anche del sistema conservativo in aderenza a quel pluralismo che contraddistingue la prassi italiana (Giuva 2014). La differenza sostanziale, ad ogni modo, tra gli archivi sonori e quelli che semplicemente possiamo definire tradizionali, sta nell'intenzionalità con cui il documento sonoro viene prodotto, già *fonte* sin da momento della sua produzione perché inserito nell'ambito di ricerche, in contrapposizione all'esigenza giuridico-amministrativa per cui vengono posti in essere gli archivi in genere. Questo concetto di volontarietà cozza con la definizione classica di archivio in senso proprio. Certamente, tutti gli archivi si *fanno* e non esistono a prescindere da un'azione, ma la ragione della genesi e della conservazione dell'archivio sonoro non si direbbe sia legata a ragioni pratiche, all'autodocumentazione, alla certezza del diritto e alla memoria delle attività di chi quell'archivio l'ha prodotto. Al contrario, il motore dell'azione sono la ricerca e al contempo la volontà di lasciare traccia di eventi e circostanze che altrimenti non troverebbero modo di residuarsi in un archivio e dunque nella storia, se non ricorrendo alla sola memoria di chi li ha vissuti e ne ha preso parte.

Per queste ragioni, le fonti orali sono parziali e di posizione, non sono né asettiche, né obiettive né tantomeno il racconto acritico di un fenomeno, piuttosto rivelano ragionamenti e considerazioni sul fatto stesso e in quanto tali si dimostrano utili per indagare il rapporto tra le esperienze personali e la storia collettiva: sono testimonianze e pezzi di vita ascrivibili in parte alla categoria degli *ego-documents* creati con intenzionalità in cui, per mezzo del racconto, il soggetto rappresenta sé stesso e il suo punto di vista, utili prospettive per indagare i più ampi aspetti di identità collettive³.

Ora, dal punto di vista prettamente archivistico, gli archivi cui si dà vita con un'operazione di costruzione di fonti orali e di fruizione in digitale, come quello che si sta per presentare, potrebbero essere ascrivibili al campo degli *invented archive*⁴ (Valacchi 2023), poiché si tratta di aggregazioni

¹ Pur abbracciando la tesi di Maria Francesca Stamuli che ritiene che *documento sonoro* e *fonte orale* possano essere semanticamente sovrapponibili, in questa sede si vorrà limitare il grado di complessità terminologico-espressiva evitando, quando possibile, di utilizzarle in maniera interscambiabile.

² Si veda ad esempio il recente *Documenti sonori. Voce, suono, musica in archivi e raccolte* (Brunetti, Robotti, e Salvalaggio 2021).

³ Si riprendono e si condividono i concetti presenti sul portale *Ti racconto la Storia* (<http://www.tiraccontolastoria.san.beniculturali.it/>), su cui brevemente si tornerà più avanti. (ultimo accesso 2023, ottobre 29).

⁴ Si ringrazia l'autore per aver consentito la lettura del contributo prima della sua pubblicazione.

per materia, digitali nel nostro caso, che liquefanno e mettono in crisi i tradizionali concetti sia di archivio che di documento e lasciano spazio anche alla performance creativa e artistica (Valacchi 2010) di chi li forma, oltre a tradire ineluttabilmente il suo punto di vista.

Naturalmente, esistono documenti sonori che nascono anche con altre finalità. Basti pensare alle registrazioni radiofoniche o ad altre registrazioni di eventi in presa diretta (di spettacoli teatrali, di comizi, di conferenze, di riunioni), agli audiolibri, o ancora alle deposizioni nei processi, alle intercettazioni telefoniche conservate all'interno delle procure, alle registrazioni audio generate mediante i servizi di messaggistica istantanea, alla produzione musicale e tanto altro.

La gestione di questi documenti ha posto – e pone – una serie di questioni tecnico-scientifiche all'archivista di un sicuro interesse, relativamente alle quali forse ancora troppo poco si è indagato. Uno dei problemi risiede nella scarsa possibilità di utilizzo di applicativi informatici per la loro descrizione e per la loro gestione in generale; il maggiore, ad avviso di chi scrive, lo riscontra chi ricorre all'utilizzo di software *open source* che molto spesso non supportano il materiale solamente sonoro, a favore invece di quello audiovisivo il quale, se non più diffuso, ha ricevuto trasversalmente maggiore interesse e ha condotto allo sviluppo di ambienti adatti per la loro gestione.

Il presente contributo vuole presentare gli esiti delle recenti attività del Laboratorio di Archivistica dell'Università di Palermo che hanno previsto la creazione, per mezzo di interviste, di un archivio di fonti orali sul movimento universitario della Pantera del 1989-90. È opportuno sottolineare che gli archivi dei movimenti politici e sociali, a prescindere da forma, tipologia e supporto – benché nell'opinione comune sia facile associarli – sono sostanzialmente diversi da quelli dei partiti politici e dei sindacati. Questa differenza deriva sicuramente dall'assenza, nel mondo movimentista, di formalizzazione istituzionale, una diversa 'struttura' organizzativa che determina differenze tanto in termini di produzione quanto di conservazione (Procaccia 2009). La questione strutturale legata tanto ai soggetti produttori quanto ai complessi documentari, che di riflesso si creano, è proprio quella su cui porre l'accento, poiché a strutture *semplici*, meno specchio di prassi burocratiche, meno inclini a produzione e conservazione, corrispondono archivi – mi si consenta l'espressione – *semplici*, in cui l'esigenza dell'autodocumentazione costituisce fattore determinante nei confronti del trattamento largamente inteso della propria memoria.

Tutto questo, in maniera molto generica, determina delle assenze in seno alla documentazione relativa ai movimenti per le quali, in un quadro come quello appena descritto, sarebbe pure improprio esprimersi in termini di *lacune* poiché anche queste legate a fattori 'strutturali' del modello dell'attività movimentista. La fonte orale, tuttavia, può colmare per mezzo dell'intervista gli eventuali vuoti. Se le precise istanze di ricerca e di narrazione, in virtù di assenza di altre fonti, sono quindi ciò che sta alla base della formazione delle fonti orali e della storia orale (Casellato 2021), questo è valido particolarmente per gli archivi dei movimenti sociali e politici la cui mancata formalizzazione condiziona la nascita di un archivio strutturato, unico, a fronte della presenza di tanti più o meno piccoli archivi personali di chi ai movimenti ha preso parte⁵.

Tutto questo, tuttavia, non sta a significare che le fonti orali siano il mezzo privilegiato per lo studio dei movimenti in generale e studenteschi in particolare. Esistono, naturalmente e in prima

⁵ In direzione della raccolta di documentazione utile alla ricostruzione dei movimenti studenteschi va, a titolo d'esempio, il progetto francese *Conservatoire des mémoires étudiantes*, raggiungibile all'indirizzo <https://www.cme-u.fr/> (ultimo accesso 2023, ottobre 29).

istanza, le carte degli archivi pubblici, degli organi centrali, del Ministero dell'Istruzione e del Ministero dell'Interno, ma anche gli occhi periferici di questi dicasteri (Questura, Prefettura, etc.), senza tralasciare gli archivi storici degli stessi atenei. Accanto ad essi, si pone la miriade di archivi privati, di persone, di partiti politici, di associazioni che, insieme ai resoconti giornalistici e filmici, contribuiscono alla ricostruzione del contesto di azione (Giorgi e Mineo 2012, 233-34)⁶.

L'archivio di fonti orali sul movimento della Pantera a Palermo e l'archivio *Limone Lunare*, di cui poco più avanti si parlerà, è stato inteso dalle curatrici, sin dal primo momento, come un processo culturale aperto nato con l'intenzione di raccogliere testimonianze che altrimenti sarebbero mute. Il fine è quello di contribuire alla ri-costruzione di una storia comune e di fornire strumenti di ricerca a varie discipline e anche – ovviamente – alla storiografia, che ne può e ne deve attingere, sia pure con tutte le criticità che una fonte orale può presentare, nonostante le quali però essa si costituisce come documento *tout court* con assoluta capacità di testimonianza di azioni, fatti, processi.

Il Progetto *Limone Lunare*: l'archivio digitale dei movimenti politici e sociali siciliani

Nel corso del 2021, con decreto n. 371 la Direzione Generale Archivi del Ministero della Cultura l'a.p.s. booq ha ricevuto un contributo finanziario per la realizzazione del progetto archivistico *Limone Lunare. Archivio di fonti orali dei movimenti politici siciliani negli anni '60 e '70 del Novecento. 1970-2020*, finanziato nell'ambito della concessione di contributi da effettuarsi su archivi dei movimenti politici e degli organismi di rappresentanza dei lavoratori.

Era da pochi mesi nato, infatti, un gruppo di lavoro e ricerca costituito, oltre che dalla scrivente, dalle dott.sse Giulia Crisci (Iuav-Università di Venezia), Francesca Di Pasquale (Soprintendenza Archivistica per la Sicilia-Archivio di Stato di Palermo) e Michela Nalbone (Università di Palermo), in collaborazione con booq, con l'obiettivo di costituire un archivio di testimonianze orali sui movimenti politici e sociali siciliani di età contemporanea e di disseminare i risultati delle ricerche tanto a livello locale quanto nazionale. Per questa ultima ragione, nel dicembre del 2022 è stato organizzato il convegno "Fare Archivio. Archivi dei movimenti sociali e delle lotte in Sicilia. Pratiche per la memoria", nel corso del quale si è presentato il progetto, si è ragionato sul ruolo delle Soprintendenze archivistiche nella tutela degli archivi sonori e si è riflettuto sull'opportunità di costituzione di una rete di tutti i soggetti, singoli o associati, che detengono patrimoni archivistici prodotti dai movimenti sociali siciliani e/o testimonianze delle varie stagioni di lotte, con l'intento di mettere a sistema il lavoro per la tutela e la valorizzazione di questi patrimoni.

Limone Lunare è costituito al presente da tre complessi documentari il cui fondo di partenza – al quale si deve anche la genesi del progetto – prende le mosse dall'attività di ricerca di Giulia Crisci e del ricercatore francese Camille Pageard; tale fondo è costituito da un estratto della trasmissione di Radio Libera (1970), prima radio pirata italiana, da interviste a protagonisti/e a vario titolo di

⁶ Si veda anche, per le fonti utili alla ricostruzione della «stagione dei movimenti» Grispigni e Musci 2003, 13-47. Non si vuole, in questa sede, presentare alcuna guida alle fonti per lo studio dei movimenti per il cui approfondimento si rimanda ai titoli sul tema presenti in bibliografia. Un lavoro di questo tipo, peraltro, abbraccia sia il tema delle fonti orali relative ai movimenti che, in parte, quello della storia delle università. Un'ottima sintesi fra queste due prospettive è rappresentata da Agostini, Giorgi, e Mineo 2014.

movimenti sociali e di lotte nati nel contesto di azione di Danilo Dolci⁷ e da due raccolte di tracce del gruppo musicale militante Canzoniere Popolare di Palermo. Il secondo fondo, anch'esso realizzato grazie alla medesima linea di finanziamento in collaborazione con UDIPALERMO, è costituito da un *corpus* di interviste realizzate da Marcella Geraci⁸ a donne che hanno militato a fianco di Letizia Colajanni⁹. L'ultimo fondo che compone al presente *Limone Lunare* è costituito, infine, dal complesso documentario composto dalle interviste ad alcuni/e protagonisti/e del movimento della Pantera a Palermo.

L'archivio *Limone Lunare* è descritto e accessibile on-line¹⁰ dalla fine del 2021. Per ciascuna fase di trattamento, di definizione di criteri, di metodologie standardizzate di comunicazione e descrizione del patrimonio si è seguito il recente Vademecum per il trattamento delle fonti orali (Coordinamento per le fonti orali 2021).

Limone Lunare è ricerca, principalmente, e ha inteso restituire testimonianze dei movimenti politici e sociali più rilevanti tra gli anni Sessanta del secolo scorso e il presente in Sicilia, evidenziando le strategie di emancipazione attraverso forme di auto-organizzazione collettive per mezzo della protesta, della lotta e anche dell'espressione artistica; non si è preteso di voler fornire tutti gli strumenti per raccontare la storia o *le storie* nel loro complesso, ma di restituire un pezzo di quelle vicende, quelle vissute dalle persone coinvolte.

Limone Lunare tuttavia è anche un esperimento e i processi di formazione dei tre fondi archivistici che lo compongono ne sono una conferma.

Il movimento della Pantera (1989-90). L'archivio di un sussulto trasversale

Sulla genesi dei movimenti studenteschi e/o universitari interviene sempre una riforma o un tentativo di riforma. Questo è stato valido per il Sessantotto, per il Settantasette, per l'Onda e, naturalmente, per la Pantera.

Il movimento della Pantera si sviluppò a cavallo tra il 1989 e il 1990 in opposizione alla riforma di Antonio Ruberti, Ministro dell'Università e della Ricerca Scientifica in quota socialista, ed ebbe il suo motore d'azione nell'allora facoltà di Lettere e Filosofia di Palermo: un movimento di occupazione di facoltà iniziato nel dicembre 1989 che dalla Sicilia poi si è espanso, a catena, in tutti gli atenei italiani fino alla smobilitazione avvenuta nel marzo/aprile del 1990.

I motivi della contestazione si situarono a più livelli. Nel mirino principale della protesta stava la nuova impostazione universitaria proposta dalla riforma Ruberti, che prevedeva l'autonomia «didattica, scientifica, organizzativa, finanziaria e contabile» degli atenei¹¹, «riconoscendo alle università la possibilità di intercettare risorse esterne per l'esercizio delle proprie funzioni» (Legrotta-

⁷ Non è possibile riassumere in maniera esaustiva un profilo su Danilo Dolci (1924-1997), sociologo, educatore, attivista, poeta, in poche righe. Per un approfondimento sulla sua attività si rimanda, tra i tanti, a Grifo 2021.

⁸ Marcella Geraci, ricercatrice, è giornalista pubblicista dal 2002.

⁹ Letizia Novella Colajanni (1914-2005) è stata una politica e attivista italiana, proveniente da una famiglia di lunga tradizione politica. Per una sua biografia si veda Geraci 2019.

¹⁰ L'archivio è consultabile all'indirizzo <https://booq.mrcrud.it/> (ultimo accesso 2023, agosto 19).

¹¹ *Legge 9 maggio 1989, n. 168, in materia di Istituzione del Ministero dell'Università e della Ricerca scientifica e tecnologica*, art.6, comma I.

glie 2019). Tuttavia, come emerge a chiare lettere dalle interviste, la contestazione aveva una forte connotazione locale legata ai problemi delle rispettive città e dei rispettivi atenei

In questi anni abbiamo visto sgretolarsi alcune pur insufficienti conquiste ottenute dagli studenti nel corso degli anni settanta, abbiamo visto ridursi le sessioni d'esami, aumentare la selezione e sancito il principio che più una facoltà è selettiva più è da considerarsi efficiente, abbiamo visto tornare in auge vecchi gruppi di potere che governano in modo clientelare e mafioso l'Università, abbiamo visto ridursi la quantità e la qualità dei servizi e di assistenza fornita agli studenti ed accresciuta attraverso l'aumento delle tasse l'esclusione dall'Università delle classi sociali subalterne. In questo clima di restaurazione e di neoconservatorismo si inserisce il progetto di legge Ruberti che con l'autonomia finanziaria e l'ingresso trionfale dell'industria privata nell'Università, con l'accentramento dei poteri nelle mani del Senato Accademico e l'ulteriore svilimento delle strutture di rappresentanza degli studenti, vuole imprimere un'accelerazione ai processi di riforma autoritaria nell'Università. [...] Rivolgiamo un appello alle componenti universitarie e agli studenti in particolare ad estendere l'agitazione e la protesta entro più breve tempo a tutti gli Atenei d'Italia ai fini della rifondazione di un movimento studentesco attivo capace di intervenire rispetto ai problemi dell'Università e della società in generale¹²

Di lì a poche settimane vennero occupate tutte le facoltà dell'ateneo palermitano ed entro la fine del gennaio del 1990 erano state occupate buona parte delle facoltà italiane, i cui studenti durante le festività natalizie avevano scelto di trascorrere le vacanze nel capoluogo siciliano per studiare i caratteri dell'occupazione e, al ritorno, *fare come a Palermo* (Maltese 2021, 93).

Alcuni movimenti studenteschi e universitari hanno avuto rilevanza anche per lo studio dei movimenti sociali (Caruso et al. 2010, 9). Questo punto, che si potrebbe collegare alla dimensione transgenerazionale o, nel caso della Pantera, *trasversale* (Maltese 2021, 107-115), può aver ostacolato il formarsi di una memoria condivisa e collettiva che presumibilmente ha avuto ricadute anche sul piano documentale. Questa logica della *trasversalità*, tuttavia, che a Palermo ha consentito la trasformazione del movimento da studentesco a universitario con l'adesione alla contestazione di alcune altre componenti universitarie, fu un elemento di grande innovazione della Pantera rispetto agli altri movimenti e apriva la strada auspicabilmente a un'ulteriore trasformazione in senso sociale del movimento, che invece non avvenne¹³.

La smobilitazione, a seguito di mesi di numerose assemblee nazionali, di cortei, di contestazione e di massiccia partecipazione, avvenne tra il marzo e l'aprile del 1990 decretando la fine di quella stagione movimentista. Sugli esiti successivi della Pantera si rimanda alle opinioni dei suoi animatori emerse nel corso delle interviste.

L'archivio costituito dalle interviste a protagonisti del movimento della Pantera a Palermo è stato

¹² Documento dell'Assemblea della facoltà di Lettere e Filosofia di Palermo contenuto nel CD-ROM *I movimenti a Palermo -1989/1990 La Pantera*, la raccolta di immagini, documenti, rassegna stampa, interviste, disegni a cura di Davide Ficarra e Vittorio Vizzini e promosso da ARCI Sicilia, 2000.

¹³ Il movimento della Pantera, poi, dal punto di vista delle analisi più propriamente ricadenti nell'ambito dei cultural studies, va ricondotto nell'alveo della subcultura intesa come stile di vita che esprime pratiche antagoniste in contrapposizione alle culture dominanti; è proprio con la Pantera peraltro che si apre una finestra di condivisione tra studenti universitari e centri sociali, come nel caso del rapporto tra la Pantera palermitana e il CSOA Montevergini (Maltese 2021, 79), in cui l'hip hop è tra gli anelli di congiunzione (Bonatesta 2022, 259).

realizzato nell'ambito delle attività del Laboratorio di Archivistica, tenuto dalla scrivente, all'interno del Corso di laurea in Beni Culturali dell'Università di Palermo nel corso degli anni accademici 2021/22 e 2022/23. Gli studenti e le studentesse ignoravano l'esistenza della Pantera; è emerso sin da subito che un evento di quella portata non aveva lasciato traccia nella memoria storica del corpo studentesco. Conoscevano il Sessantotto per la sua eco travolgente, ma non conoscevano nessun altro movimento studentesco, universitario, giovanile, neppure la recente Onda del 2008¹⁴. La temperie socio-politica del periodo, le ragioni che hanno portato all'occupazione, il significato della riforma Ruberti, l'evoluzione del movimento sono stati presentati agli studenti in occasione di un seminario tenuto da Pietro Maltese, docente di pedagogia all'Università di Palermo, che nei suoi studi si è occupato, tra le altre cose, del rapporto tra istruzione, politica e cambiamenti economico-produttivi e che si è dedicato al tema in un suo recente lavoro (Maltese 2021). Dall'incontro con Maltese sono emersi temi quali la 'fininvestizzazione'/'berlusconizzazione' della società; la 'brandizzazione' di un movimento universitario, che ha conosciuto l'attribuzione di un nome, di un logo e che si è avvalso di campagne pubblicitarie; la contestualizzazione del movimento nel più ampio quadro geopolitico determinato dalla fine del socialismo reale; l'adozione della strage di piazza Tienanmen e dell'Intifada palestinese come modello iconografico; i presunti temi meridionalisti nel movimento; la logica della *trasversalità*.

Un ulteriore passo in direzione della formazione di un'opinione sul movimento da parte degli studenti e delle studentesse è stato compiuto dedicando ore di ricerca e studio in emeroteca presso l'Istituto Gramsci Siciliano, con lo spoglio sistematico dei quotidiani Il Manifesto, L'Ora, La Repubblica, L'Unità e il Giornale di Sicilia, di fatto raccogliendo la totalità degli articoli pubblicati sulla Pantera dal dicembre del 1989 all'aprile del 1990, materiale che poi hanno utilizzato come studio utile alla formulazione del questionario e presente nel campo bibliografia della scheda descrittiva. Infine, si è arrivati alla strutturazione dei questionari, formulati in autonomia dagli studenti e dalle studentesse. Ciascun gruppo di lavoro ha articolato la scaletta di domande da porre, di temi da sottolineare, di questioni da sollevare sulla base delle proprie sensibilità, lasciando, in alcuni casi, anche spazio al fluire dei ricordi degli/le intervistati/e. Pertanto, ogni conversazione risulta essere strutturalmente diversa dalle altre.

Le interviste complessive, tra quelle realizzate e quelle raccolte, al presente sono ventidue. Le prime sei sono state realizzate nel 2022 (a Barbara Amodeo, Giuseppe Burgio, Annibale Cangemi, Andrea Cozzo, Davide Ficarra, Gilda Terranova) e altre otto nel 2023 (agli Avion Travel, Maurizio Carta, Antonio Caruso, Alessandro Crociata, V.C., Simone Lucido, Ottavio Navarra, Vincenzo Puleo, Antonio Rampolla). A queste, si sono aggiunte ulteriori sette interviste raccolte nel 2000 per il primo decennale del movimento da ARCI Sicilia, di cui ci è stato concesso l'utilizzo¹⁵.

Il campione delle/degli intervistate/i è stato individuato sulla base delle conoscenze personali della docente, degli studenti e delle studentesse e, in ultimo, grazie alla catena di segnalazioni che l'avvio del progetto ha generato. Il primo blocco di interviste ha riguardato solamente studenti allora iscritti alla facoltà di Lettere e Filosofia di Palermo; il secondo blocco, così come le interviste del decennale, ha considerato studenti di altre facoltà (Architettura, Giurisprudenza, Psicologia, Scienze, Scienze Politiche).

¹⁴ Sul movimento dell'Onda si veda Caruso, Giorgi, Mattoni e Piazza 2010.

¹⁵ Si ringraziano vivamente tutti/e gli/le intervistati/e per l'adesione entusiasta al progetto di raccolta di interviste.

Dalle interviste è emersa quale sia stata la percezione della riforma tra i protagonisti, i quali hanno provato a ragionare sui motivi che hanno visto Palermo come centro propulsore del movimento, contestualizzando la contestazione nella città che negli '80, quegli anni, era stata teatro delle guerre di mafia ma che stava conoscendo la sua Primavera con la prima giunta Orlando. Fra gli altri temi emersi si segnala l'eredità dei movimenti precedenti, il Sessantotto e il Settantasette; i vantaggi della menzionata logica della *trasversalità*; il rapporto con la stampa; il rapporto con gli altri atenei e le modalità di comunicazione, fra cui il massiccio utilizzo del fax; lo svolgimento delle assemblee nazionali; i rapporti coi partiti politici e con le istituzioni locali e no; l'andamento della quotidianità all'interno delle facoltà occupate; le eventuali azioni di smobilitazione; le opposizioni interne; gli endorsement pubblici (da parte di politici, esponenti della cultura, intellettuali, artisti, registi, musicisti, giornalisti, personaggi dello spettacolo); l'eredità della Pantera nelle università e nella società in generale.

Per la descrizione delle interviste è stato utilizzato un apposito ambiente di lavoro su xDams, per la cui realizzazione si ringrazia la società regista.exe che aveva già in precedenza realizzato altri ambienti per *Limone Lunare*. Nell'ambiente di lavoro è presente la struttura del fondo, una struttura semplice fondo – unità archivistiche. Ciascuna unità è stata descritta allo stesso modo in cui sono stati descritti i due fondi precedenti di *Limone Lunare*, con la rilevazione di alcuni dati fra cui le responsabilità, le persone coinvolte, le date (di intervista e dell'evento rappresentati), le chiavi di accesso persone/luoghi/soggetto-argomento che vanno a creare un *thesaurus* trasversale a tutti i documenti sonori trattati. È presente il file digitale dell'intervista, corredato dalla liberatoria al suo utilizzo, un breve abstract e le notizie storiche di contesto e in ultimo gli articoli di giornale utilizzati dagli studenti e dalle studentesse per la formulazione del questionario, inseriti come bibliografia a disposizione e a corredo delle schede archivistiche delle interviste.

L'idea finale era quella della creazione di un archivio online, opportunamente descritto, ad accesso completamente aperto e costantemente incrementabile (con ulteriori documenti sonori, ma non solo), da agganciare sia a *Limone Lunare*, come si è fatto, sia a esperienze simili già in atto e presenti in rete¹⁶. Il lavoro non è solo una riflessione e un racconto su un movimento universitario, ma vuole considerare la volontà da parte degli studenti di porsi come soggetto politico. È quindi il racconto di una mobilitazione, dei suoi caratteri e dei suoi protagonisti, ed è sembrato che fosse quantomai opportuno coinvolgere proprio gli studenti universitari in questo processo.

Il lavoro di descrizione del patrimonio oggetto del progetto ha fatto emergere alcune criticità legate ai software di descrizione archivistica *open-source*: tra quelli disponibili, la maggior parte non consentiva né la descrizione né la riproduzione di fonti orali. Si è scelto di utilizzare xDams, già utilizzato in altri analoghi lavori, rivelatosi il più idoneo poiché offre la possibilità di gestire e descrivere una molteplicità di tipologie documentarie. Tra queste, quella maggiormente adattabile alle esigenze progettuali risultava quella, seppur diversa per tante ragioni, di gestione del materiale audiovisivo, che regista.exe ha provveduto appositamente a trasformare in ambiente 'archivio interviste' considerando la specificità dei documenti sonori, con un modello descrittivo analitico che consente non solo di operare una schedatura di ciascuna registrazione, descritta a

¹⁶ Tra queste, la digital library *Ti racconto la Storia* promossa nell'ambito delle attività del SAN, Sistema Archivistico Nazionale (<http://www.tiraccontolastoria.san.beniculturali.it/>), e delle attività dell'associazione 'Pantera 90 Archivio' di recente costituzione con il fine di raccogliere materiali cartacei e audiovisivi sul movimento del 1990 per la costituzione di un database aperto alla libera consultazione (<http://www.lapantera.org>) (ultimo accesso 2023, agosto 19).

livello di ‘documento’, ma di creare, a partire da ognuna di esse, delle schede di livello inferiore (*other level*). Queste permettono di descrivere nel dettaglio ciascun argomento emerso nel corso dell’intervista, corredandolo coi corrispettivi indici relativi ai nomi, ai luoghi e agli argomenti individuati, secondo un sistema di *segmenti* che corrisponde a scansioni temporali sia della traccia registrata che dell’evento documentato. La segmentazione, collegata al minutaggio, risulta di particolare importanza per l’individuazione, il reperimento e l’approfondimento delle informazioni, come sottolineato anche in altri progetti legati alle fonti orali (Valentini et al. 2023).

Le criticità incontrate durante la messa a punto dell’ambiente di lavoro non hanno impedito lo svolgimento dell’attività di descrizione. Tuttavia, per la pubblicazione del lavoro sul sito del progetto si dovrà attendere ancora qualche tempo, mentre risulta già consultabile presso la sede di booq dai terminali presenti nella sala di lettura.

L’archivio Pantera: le descrizioni

Il fondo della Pantera, la cui descrizione è stata interamente operata dagli studenti e dalle studentesse, ha seguito giocoforza gli stessi criteri in principio formulati per gli altri due fondi documentari di *Limone Lunare*.

La descrizione del patrimonio ha previsto una schedatura analitica di tutte le unità archivistiche che compongono il *corpus* documentario di cui sono stati rilevati: tipologia archivistica; denominazione dell’unità; responsabilità principali; nomi degli/delle intervistati/e; data cronica e topica; chiavi di accesso; durata della registrazione; specifiche tecniche dei file audio; abstract; lingua della documentazione; presenza di allegati. Sulla base delle chiavi di accesso, ciascuna unità è stata corredata da indici relativi ai nomi di persona, ai luoghi e agli argomenti individuati al suo interno costituendo delle parallele schede *Authority*, così da garantire un accesso alle informazioni tramite molteplici chiavi. L’intero lavoro, come si è già avuto modo di anticipare, è stato condotto facendo ricorso alla versione *in cloud* di xDams, per la cui attività regista.exe ha provveduto a creare un ulteriore ambiente *ad hoc* gemello di quelli creati per gli altri fondi di *Limone Lunare*, nel pieno rispetto di quanto previsto dagli standard internazionali per la descrizione archivistica (in particolare ISAD e ISAAR e i corrispettivi standard di struttura EAD e EAC-CPF).

Di seguito dunque vengono riportate le descrizioni in forma semplificata per la pubblicazione in questa sede. Sono state sostanzialmente mantenute le descrizioni operate dagli studenti. Per la descrizione completa anche di metadati si rinvia alla pubblicazione prossima sul sito del progetto.

1. Titolo: *Intervista a Barbara Amodeo*

Abstract: Barbara Amodeo, allora studentessa della facoltà di Lettere e Filosofia, descrive la Pantera come un movimento formato da organizzazioni itineranti nei suoi membri e funzioni, con una gestione assembleare. Sottolinea tra gli obiettivi del movimento la qualità della didattica e la democrazia della vita universitaria: temi e riflessioni varie erano trattati con un approccio trasversale.

Data: 2022 mag. 31

Intervistatore/rice: Chiara Bertozzino, Chiara Lindiner

Soggetti: occupazione; femminismo; logica della trasversalità; aziendalizzazione

Durata: 00:53:04

2. Titolo: *Intervista a Giuseppe Burgio*

Abstract: Giuseppe Burgio racconta come è nato il movimento della Pantera e in che modo si sia opposto alla trasformazione del sistema universitario proposta dalla riforma Ruberti. I principali temi affrontati sono: i momenti di vita quotidiana all'interno delle facoltà occupate; il rapporto con il mondo della politica e del giornalismo; la rappresentanza diretta e la socializzazione. Inoltre l'intervistato si sofferma sull'importanza dell'autoformazione come elemento più significativo dell'esperienza maturata all'interno del movimento; vengono infine evidenziati i fattori sociali e politici che portano al tramonto della Pantera.

Intervistatore/ricce: Matteo Morici

Data: 2023 mag. 31

Soggetti: aziendalizzazione; non-violenza; autoformazione; rappresentanza diretta

Durata: 00:40:33

3. Titolo: *Intervista ad Annibale Cangemi*

Abstract: Annibale Cangemi ricorda l'esordio del movimento della Pantera a dicembre del 1989 nella facoltà di Lettere e Filosofia dell'Università di Palermo e il suo obiettivo principale, ovvero il blocco della riforma Ruberti nel suo aspetto di aziendalizzazione dell'università. Si sottolinea l'eco sociale molto ampia del movimento, che in pochissimo tempo si estese in molti atenei italiani. Ribadendo la natura politica del movimento, Cangemi sottolinea che la partecipazione degli studenti al movimento non era data solo dall'opposizione verso la riforma Ruberti, ma dall'avvertita esigenza di autonomia rispetto le autorità accademiche e di soluzione dei problemi interni alle facoltà.

Intervistatore/ricce: Giuliana Sulli

Data: 2022 mag. 31

Soggetti: pacifismo; collettivizzazione; logica della trasversalità; centri sociali

Durata: 00:34:43

4. Titolo: *Intervista ad Andrea Cozzo*

Abstract: Andrea Cozzo individua nella nuova logica della autonomia finanziaria degli atenei proposta dalla riforma Ruberti – nel quale vi leggeva il definanziamento pubblico dell'università secondo una logica imprenditoriale che mirava all'aziendalizzazione del sistema universitario – il punto centrale della contestazione. Più di altri Cozzo, che allora era dottorando, affronta il tema della *trasversalità*, presentando la Pantera come capace di tenere insieme tutte le figure presenti all'interno dell'università, dagli studenti ai docenti passando per il personale tecnico-amministrativo e attribuendo al movimento un taglio politico-culturale. Cozzo affronta anche il tema dei mass-media, coi suoi tentativi di strumentalizzazione, da un lato, e di supporto, in altri casi. Tra i limiti del movimento, Cozzo riconosce l'incapacità di far presa sulla società, cosa che determinò la risoluzione del movimento come solamente universitario.

Intervistatore/ricce: Alessia Ardizzone, Alessia Muratore

Data: 2022 mag. 31

Soggetti: vento del sud; privatizzazione; aziendalizzazione; logica della trasversalità

Durata: 00:52:27

5. Titolo: *Intervista a Davide Ficarra*

Abstract: L'intervista a Davide Ficarra prende le mosse proprio dall'avvio della Pantera, restituendo alle intervistatrici informazioni sull'organizzazione dell'occupazione, non tralasciando il coinvolgimento dei docenti universitari ma anche degli studenti medi, i quali pure presero parte al movimento. L'intervista si conclude con le riflessioni dell'intervistato sugli esiti della Pantera, la quale dal punto di vista sostanziale non si può dire che ottenne i risultati sperati, ma ebbe il merito di cambiare il clima all'interno delle università italiane.

Intervistatore/ricce: Antonina Bommarito, Maria Domenica Vitello

Data: 2022 giu. 01

Soggetti: movimento del '77; partecipazione scolastica; concerti; riforma Ruberti

Durata: 00:28:38

6. Titolo: *Intervista a Gilda Terranova*

Abstract: Gilda Terranova, rispondendo alla richiesta di ripercorrere le fasi iniziali del movimento, affronta i motivi della contestazione sottolineando il peso delle rivendicazioni locali della protesta, che hanno contribuito alla *trasversalità* del movimento. Fra gli altri temi emersi, sono degni di nota: il ruolo del "Libro Bianco", documento di denuncia della condizione dell'ateneo e della sua gestione; gli strumenti di comunicazione, tra fax e *tazebao*; i meccanismi di autofinanziamento; la *berlusconizzazione* della società; la vita quotidiana in facoltà tra autogestione degli spazi e rapporti coi docenti.

Intervistatore/ricce: Mariagrazia Ciaccio, Maria Costanzo

Data: 2022 mag. 31

Soggetti: Libro Bianco; tazebao; autogestione; Tienanmen.

Durata: 00:55:12

7. Titolo: *Intervista Avion Travel*

Abstract: L'intervista concessa da Peppe D'Argenzio e Mimì Ciaramella, componenti della band Avion Travel, è incentrata sul concerto tenuto dagli stessi all'Università di Palermo nel febbraio del 1990. Fervidi sostenitori dei movimenti giovanili, rivivono con entusiasmo il '68 e il '77, anni che li hanno coinvolti in prima persona. A questo proposito D'Argenzio riferisce come sia stato per loro sorprendente che venissero contattati prima ancora di altre band come i 99 Posse e gli Almagretta. Pur non definendosi militanti, condividono l'orizzonte politico e sociale del movimento abbracciandone pienamente le istanze.

Intervistatore/ricce: Ilaria Aliesia Pia Siragusa, Sarah Stabile

Data: 2023 apr. 17

Soggetti: ragazzi dell'89; rivoluzione punk; privatizzazione dell'università; musica

Durata: 00:22:03

8. Titolo: *Intervista a Maurizio Carta*

Abstract: L'intervista a Maurizio Carta ha messo in luce alcune questioni importanti del mondo universitario emerse in quel periodo storico. Carta riconosce nella scintilla accesa dalla protesta

un modo per avanzare istanze più generali e miranti a una revisione della didattica, individuando il momento zero nel desiderio di una riflessione più ampia da parte di coloro che vivevano l'università in tutte le sue sfaccettature, luci e ombre. Carta sottolinea il clima collaborativo che ha esperito in quei mesi tra i vari livelli universitari, i quali in vario modo condividevano alcune delle istanze espresse dal movimento. Il raggiungimento di alcuni traguardi e la necessità di portare a compimento, per molti, un percorso universitario che aveva già dato i suoi frutti, insieme all'esaurimento della spinta propulsiva che aveva animato sul nascere l'iniziativa, portò alla disoccupazione delle facoltà nel marzo del 1990.

Data: 2023 apr. 18

Intervistatore/ricce: Elisabetta Franzone, Francesca Salvia

Soggetti: riforma Ruberti; privatizzazione dell'università; politicizzazione; attività ludiche.

Durata: 00:43:01

9. Titolo: Intervista ad Antonio Caruso

Abstract: Antonio Caruso, tra i maggiori protagonisti della Pantera a Palermo, ripercorre le ragioni della protesta a partire dall'analisi della riforma Ruberti fino ad arrivare al racconto dei giorni delle prime occupazioni. Particolare attenzione viene data al tema della violenza/non-violenza, avendo egli stesso presieduto l'assemblea di Firenze del febbraio '90 in cui si discusse accesa-mente sull'argomento. Caruso, chiudendo l'intervista, si focalizza sull'aspetto etico e valoriale della protesta e sulla città di Palermo, teatro all'epoca di violenze, conflitti di mafia e corruzione, oggi profondamente mutata sul piano della consapevolezza e dell'evoluzione urbana.

Data: 2023 apr. 18; 2023 apr. 19

Intervistatore/ricce: Andrea Mazzei, Miriam Palazzolo, Riccardo Raccuglia

Soggetti: Intifada; Tienanmen; muro di Berlino; non-violenza

Durata: 00:32:56; 01:18:36

10. Titolo: Intervista ad Alessandro Crociata

Abstract: Alessandro Crociata, in quegli anni studente di Giurisprudenza, è stato sin dall'inizio tra i più attivi della Pantera, partecipando in prima persona alle occupazioni, alle manifestazioni, alle assemblee nazionali. Allora militante di Democrazia Proletaria, si sofferma sulle varie componenti politiche del movimento, dagli autonomi a coloro che invece erano legati alle sezioni giovanili dei partiti passando per gli studenti che, pur non identificandosi in uno gruppo politico ben preciso, hanno abbracciato con convinzione le istanze della Pantera, sottolineando la capacità di coinvolgimento del movimento. In chiusura dell'intervista, Crociata analizza il rapporto con la stampa e, nel bilancio finale a distanza di un trentennio, espone le sue idee sui deboli esiti della protesta.

Intervistatore/ricce: Giulia Bruno, Vanessa Lamberti

Data: 2023 apr. 17

Soggetti: sciopero; Giornale di Sicilia, Democrazia Proletaria; Federazione Giovanile Comunista Italiana

Durata: 00:27:37

11. Titolo: *Intervista a V.C.*

Abstract: Intervista a V.C, studente della facoltà di Lettere e Filosofia che ha partecipato da laureando al movimento della Pantera. L'intervistato descrive il clima che si viveva all'interno dell'ateneo palermitano e analizza, oltre certamente all'opposizione alla riforma Ruberti, le istanze locali che il movimento portava avanti. Viene raccontata l'espansione a macchia d'olio della protesta da Palermo verso tutti gli atenei italiani, tenendo in considerazione il quadro geopolitico, la situazione nazionale e i fatti locali che hanno influenzato l'esperienza della Pantera.

Data: 2023 apr. 18

Intervistatore/ricce: Francesco Paolo Li Castri, Giulia Maganuco

Soggetti: occupazione; riforma universitaria; vento del sud; diritto allo studio

Durata: 00:54:15

12. Titolo: *Intervista a Simone Lucido*

Abstract: L'intervista a Simone Lucido, studente della facoltà di Lettere e Filosofia, prende le mosse dalle questioni alla base del movimento, affrontando il tema dei rapporti con gli atenei italiani e descrivendo le attività che scandivano i momenti all'interno delle facoltà. Si è parlato del rapporto tra studenti, stampa e forze dell'ordine e della logica della *trasversalità*.

Intervistatore/ricce: Valeria D'Acquisto, Simona Romano

Data: 2023 apr. 17

Soggetti: commissioni stampa; strumentalizzazione; pacifismo; cortei

Durata: 00:23:57

13. Titolo: *Intervista a Ottavio Navarra*

Abstract: Ottavio Navarra, allora studente fuorisede di Giurisprudenza, racconta il movimento della Pantera con riferimenti importanti al contesto storico del tempo, descrivendolo come pacifico, di ribellione e solidarietà e legato alla necessità del diritto allo studio. Un movimento plurale, quello dei ragazzi dell'89 – come venne definitivo –, spesso associati ai giovani del '68 e del '77, che si ritrovano a gestire le loro vite dentro l'università, organizzando le assemblee e rispondendo alla stampa che mostra un atteggiamento troppo spesso paternalistico nei loro confronti. Come tutti i movimenti, anche quello della Pantera è destinato a finire, "come palline su uno zampillo d'acqua" (Ralf Gustav Dahrendorf). Tuttavia, dal punto di vista editoriale, ne rimane traccia nel romanzo *Inseguendo la Pantera* di Peppe Bettoliere e in *Cento giorni. Cronache del movimento studentesco della Pantera '90* di Massimiliano Denaro.

Intervistatore/ricce: Daria Castellese, Rosolino Gallina

Data: 2023 apr. 19

Soggetti: editoria; brandizzazione; meridionalismo; fax

Durata: 00:49:34

14. Titolo: *Intervista a Vincenzo Puleo*

Abstract: Vincenzo Puleo è stato l'organizzatore del grande concerto tenutosi il 17 gennaio 1990 durante l'occupazione all'Università di Palermo. Durante l'intervista si raccontano le modalità con

cui è stato organizzato l'evento, le band che hanno risposto all'appello e gli ideali che stavano alla base del concepimento di questo progetto.

Data: 2023 apr. 18

Intervistatore/ric: Gabriele Riela, Ekaterina Shirkunova

Soggetti: concerto; Avion Travel; Litfiba; autofinanziamento

Durata: 00:31:50

15. Titolo: Intervista ad Antonio Rampolla

Abstract: L'intervista ad Antonio Rampolla inizia con il racconto del passato politico dell'intervistato e del sostrato socio-politico di Palermo in quegli anni, portando la sua esperienza, comune e condivisa con altri/e, come esempio per motivare l'esplosione del movimento della Pantera proprio a Palermo. Centrale si rivela la descrizione della città flagellata dalle guerre di mafia ma che trova il modo di rialzarsi attraverso gli studenti.

Intervistatore/ric: Ginevra Cardile, Calogera Maria Andrea Zappardo

Data: 2023 apr. 18

Soggetti: guerre di mafia; anarchia; lotte anti-carcerarie; centri sociali

Durata: 00:42:29

16. Titolo: Intervista ad Antonio Bellia

Abstract: Antonio Bellia inizia la testimonianza con una riflessione sulla ricerca dell'identità politica da parte del movimento, leggendo la presa di posizione sulla non-violenza, assunta non di buon grado da tutti, sotto quella lente. Analizzando le cause scatenanti del movimento, Bellia ritiene che la riforma Ruberti fu la miccia che accese una situazione che a Palermo, città che stava vivendo dei grossi cambiamenti (fra cui la nota Primavera), sobbolliva da tempo. In ultimo, l'intervistato ragiona sui possibili esiti del movimento qualora da universitario si fosse trasformato in qualcosa di più ampio, nella cui direzione si era tentati di andare stringendo rapporti con gli operai della Keller allora in fermento.

Intervistatore/ric: Antonio Bellia (?)

Data: 2000

Soggetti: costruzione dell'identità; primavera di Palermo; non-violenza; operai Keller

Durata: 00:01:41; 00:03:45; 00:01:18; 00:01:59

17. Titolo: Intervista ad Annibale Cangemi (2000)

Abstract: Annibale Cangemi affronta il carattere – ritenuto offensivo – della riforma Ruberti la quale, in quanto legge dello Stato, andava ostacolata da tutti su scala nazionale. Per questa ragione, Cangemi è fra quelli che spingono verso il coinvolgimento delle altre università italiane nel movimento di protesta, in città dove peraltro la tradizione politica studentesca aveva creato terreno fertile per lo sviluppo delle idee della Pantera. Come molti altri, Cangemi sottolinea il ruolo invasivo della stampa nella definizione delle azioni del movimento, individuando in essa una delle debolezze della Pantera, accanto alla mancanza di ragionamenti sulle prospettive future della protesta.

Intervistatore/ric: Antonio Bellia (?)

Data: 2000

Soggetti: riforma Ruberti; cattolici popolari; politicizzazione; rapporti con la stampa

Durata: 00:03:11; 00:02:28; 00:03:20; 00:02:39

18. Titolo: Intervista a Francesco Costanzo

Abstract: Francesco Costanzo, studente di Giurisprudenza, descrive la Pantera come un movimento politico non violento; racconta, dunque, che la scelta della non-violenza fu un metodo per non evitare la strumentalizzazione e per potersi esprimere al meglio. Difatti, le assemblee svolte durante il periodo dell'occupazione erano uno strumento utilizzato da tutti gli studenti per esprimere la propria opinione, in contrasto con una rappresentanza vuota e, quindi, inutile. Infine, l'intervistato si sofferma sull'involuzione del movimento, su alcuni aspetti contraddittori e poco coerenti.

Intervistatore/ricce: Antonio Bellia (?)

Data: 2000

Soggetti: movimento antimafia; autonomia universitaria; non-violenza; rappresentanza

Durata: 00:02:57; 00:03:32; 00:03:24

19. Titolo: Intervista a Davide Ficarra (2000)

Abstract: Davide Ficarra ripercorre i momenti delle prime assemblee a Lettere e Filosofia in cui si decide per l'occupazione, ricordando il ruolo di alcuni gruppi universitari, fra cui quello dei Cattolici popolari che si opponeva tanto fermamente quanto inefficacemente alle occupazioni. Per Ficarra, Palermo è stata la culla del movimento a causa o grazie alle forti contraddizioni che la caratterizzavano. Infine, individua nel rapporto coi mezzi di informazione uno dei limiti del movimento.

Intervistatore/ricce: Antonio Bellia (?)

Data: 2000

Soggetti: assemblee; cattolici popolari; movimento antimafia; mezzi di informazione

Durata: 00:02:30; 00:03:46; 00:03:22

20. Titolo: Intervista a Fabio Montagnino

Abstract: Fabio Montagnino, della facoltà di Scienze, nell'analizzare l'eruzione del malcontento universitario a Palermo, racconta l'adesione della propria facoltà all'occupazione. Nel corso dell'intervista viene affrontato il tema dell'utilizzo dei mezzi di informazione i quali, se da un lato hanno potenziato la comunicazione degli eventi, dall'altro hanno appiattito lo spessore del confronto politico all'interno del movimento stesso. In ultimo avanza delle ipotesi sull'eco del movimento nelle scelte dei suoi partecipanti di rimanere a vivere e lavorare a Palermo.

Intervistatore/ricce: Antonio Bellia (?)

Data: 2000

Soggetti: occupazione; eco della Pantera; processo Andreotti; mafia

Durata: 00:01:52; 00:02:56; 00:02:40; 00:01:32

21. *Titolo: Intervista a Elisabeth Parisi*

Abstract: In questa intervista si tratta del ruolo attivo delle donne all'interno del movimento tramite la partecipazione alle varie commissioni istituite nel corso dell'occupazione. Altra tematica è la stesura del menzionato Libro Bianco. L'intervistata racconta degli studenti della facoltà di Fisica, che ritardarono la fine della disoccupazione in quanto il direttore dell'istituto si rifiutava di incontrare i ragazzi per la consegna delle chiavi. La non-violenza fu un elemento caratterizzante che accomunava tutti i partecipanti al movimento, il quale inoltre rifiutava che le strutture di partito prendessero il sopravvento. Infine si riflette sull'impatto che la Pantera ha avuto sulla città al termine della propaganda mediatica, ma soprattutto sui partecipati che tutt'oggi formano un gruppo capace di condividere opinioni differenti.

Intervistatore/ricce: Antonio Bellia (?)

Data: 2000

Soggetti: Libro Bianco; commissioni; non-violenza; partecipazione femminile

Durata: 00:01:08; 00:01:49; 00:01:34; 00:01:51; 00:00:51; 00:01:00

22. *Titolo: Intervista a Orazio Rosalia*

Abstract: Orazio Rosalia ripercorre il clima politico, economico e sociale nazionale e locale in cui si sviluppa la Pantera, facendo riferimento al 'craxismo', alla realtà palermitana e naturalmente alla riforma Ruberti. Raccontando la grande partecipazione iniziale, Rosalia si sofferma sulle differenze in seno al movimento che hanno poi condotto, dopo la grande manifestazione di Napoli, alla perdita di vigore della Pantera.

Intervistatore/ricce: Antonio Bellia (?)

Data: 2000

Soggetti: craxismo; strumentalizzazione; politicizzazione; Democrazia Proletaria

Durata: 00:05:07; 00:03:33; 00:01:58

Riferimenti bibliografici

- Agostini, Giovanni, Andrea Giorgi, e Leonardo Mineo, a cura di. 2014. *La memoria dell'Università. Le fonti orali per la storia dell'Università degli studi di Trento (1962-1972)*. Bologna: Il Mulino.
- Bonatesta, Antonio. 2022. "L'Italia della Pantera e delle posse. Conflitto sociale e capitale subculturale nell'hip hop degli anni Ottanta-Novanta". *Ricerche di storia politica* 3: 257-277.
- Brunetti, Dimitri, Diego Robotti, e Elisa Salvalaggio, a cura di. 2021. *Documenti sonori. Voce, suono, musica in archivi e raccolte*. Torino: Centro Studi Piemontesi.
- Caruso, Loris, Alberta Giorgi, Alice Mattoni, e Gianni Piazza, a cura di. 2010. *Alla ricerca dell'Onda. I nuovi conflitti nell'istruzione superiore*. Milano: FrancoAngeli.
- Casellato, Alessandro, a cura di. 2021. *Buone pratiche per la storia orale. Guida all'uso*. Firenze: Editpress.
- Coordinamento per le fonti orali. 2021. *Vademecum per il trattamento delle fonti orali*, <https://icar.cultura.gov.it/attivita-e-progetti/progetti-in-collaborazione/vademecum-per-il-trattamento-delle-fonti-orali> (consultato il 17 agosto 2023).
- D'Alessandro Leonardo Pompeo. 2019. "Storia delle istituzioni e fonti orali: il "laboratorio" di studi e ricerche su Università e Sessantotto". *Le carte e la storia* 2: 153-163.
- Geraci, Marcella. 2019. *Letizia era Letizia. La memoria di Letizia Colajanni tra lotte e conquiste, speranze e futuro*. Caltanissetta: Lussografica.
- Giorgi, Andrea, e Leonardo Mineo. 2012. "«Grazie ad un lavoro costante e capillare». Fonti documentarie per lo studio del Sessantotto (1966-1970)". In *Le istituzioni universitarie e il Sessantotto*, a cura di Alessandro Breccia, 231-266. Bologna: Clueb.
- Giorgi, Andrea, e Leonardo Mineo. 2014. "Dal «gran vento di maggio» a «una miriade di gruppetti». Fonti documentarie per lo studio dei movimenti di contestazione (1966-1970)". *Ventunesimo Secolo* 34: 19-48.
- Giuva, Linda. 2014. "Gli archivi storici in Italia: la mappa della conservazione". In *Archivistica. Teorie, metodi, pratiche*, a cura di Linda Giuva, e Maria Guercio, 99-135. Roma: Carocci.
- Grifo, Marco. 2021. *Le reti di Danilo Dolci: Sviluppo di comunità e nonviolenza in Sicilia occidentale*, Milano: FrancoAngeli.
- Grispigni, Marco, e Leonardo Musci, a cura di. 2003. *Guida alle fonti per la storia dei movimenti in Italia (1966-1978)*, Roma: Ministero per i beni e le attività culturali.
- Legrottaglie, Luana. 2019. "L'autonomia universitaria e i suoi critici in parlamento e nelle piazze (1989-1990)". *Diacronie. Studi di storia contemporanea*, 37(1): 1-26.
- Procaccia, Michela. 2009. "Fonti scritte e fonti orali nella politica di tutela degli archivi sindacali". In *Lavoratrici e lavoratori: le fonti archivistiche*, a cura di Ariella Verrocchio, 31-36. Trieste: EUT Edizioni Università di Trieste.
- Stamuli, Maria Francesca. 2019. "Fonti orali, documenti e archivi: riflessioni e proposte per la nascita di un 'archivio vivo'". In *Gli archivi sonori al crocevia tra scienze fonetiche, informatica uma-*

nistica e patrimonio digitale, a cura di Duccio Piccardi, Fabio Ardolino, e Silvia Calamai, 95-109. Milano: Officinaventuno.

Valacchi, Federico. 2023. “Se l’archivio è artificiale. Verso uno ius archivi partecipativo?”. *AIDAinformazioni* 41 (1-2): 153-170.

Valacchi, Federico. 2010. “Bonaini, Top’ivio e il “gato Archivaldo”: possono gli archivi essere (anche) divertenti?”. *Il Capitale culturale. Studies on the Value of Cultural Heritage* 1: 57-81.

Valentini, Cecilia, Duccio Piccardi, Silvia Calamai, e Maria Francesca Stamuli. 2023. “Da cornice a soggetto. Il documento sonoro nell’infrastruttura Archivio Vi.Vo.”. *Archivi* XVIII(1): 87-125.

Viola, Paolo. 2006. *Oligarchie. Una storia orale dell’Università di Palermo*, Donzelli: Roma.