

# Metodi e strumenti di un progetto di valorizzazione delle riviste italiane di biblioteconomia in Wikidata

di Carlo Bianchini, Andrea Marchitelli, Alessandra Moi

## Introduzione

Se si vogliono indagare gli sviluppi più recenti della biblioteconomia e della bibliografia italiane, facendone emergere i temi, le linee di sviluppo, le tendenze, i protagonisti e valorizzando, in ultimo, la produzione scientifica nazionale, è indispensabile disporre di dati bibliografici completi e aggiornati.

Gli autori di questo studio hanno quindi deciso di procedere alla creazione di una banca dati bibliografica relativa alla produzione scientifica di ambito biblioteconomico in Italia, a partire dalle riviste di settore, che sono la sede più propria del dibattito e lo strumento più adatto a individuare le linee di sviluppo disciplinari e professionali.

Allo stato attuale, infatti, gli strumenti bibliografici disponibili per la letteratura biblioteconomica italiana sono sostanzialmente due: *BIB. Bibliografia italiana delle biblioteche, del libro e dell'informazione* di Alberto Petrucciani, Vittorio Ponzani e Giulia Visintin, con un CD-ROM allegato che copre l'arco temporale dal 1971 al 2004, e con una versione in linea per il periodo 1992-2001<sup>1</sup>, e la rubrica *Letteratura professionale italiana* pubblicata annualmente dal 1975 dapprima sul *Bollettino AIB* e poi su *AIB studi*, da Alberto Petrucciani e Vittorio Ponzani, e sulla quale si basa la banca dati citata.

Per un'elaborazione dei dati più facile e rapida sarebbe preferibile disporre di dati in formati trattabili in modo automatico, ma il formato con il quale sono disponibili i dati di questi preziosi e unici strumenti bibliografici rende complessa ogni possibile elaborazione di questo tipo. La banca dati *BIB* è infatti disponibile su un CD-ROM che richiede hardware e software sempre più difficilmente reperibili; la sua versione online, oltre a essere cronologicamente limitata, non è neppure più raggiungibile

CARLO BIANCHINI, Università di Pavia, Dipartimento di musicologia e beni culturali, Pavia, e-mail: carlo.bianchini@unipv.it.

ANDREA MARCHITELLI, EBSCO Information Services, Roma, e-mail: amarchitelli@ebSCO.com.

ALESSANDRA MOI, Università degli studi di Milano-Bicocca, Milano, e-mail: alessandra.moi@unimib.it.  
Ultima consultazione siti web: 19 luglio 2023.

1 Cfr. <<https://www.aib.it/aib/bib/bib.htm>>.

**AIB studi**, vol. 63 n. 2 (maggio/agosto 2023), p. 313-335. DOI 10.2426/aibstudi-13893

ISSN: 2280-9112, E-ISSN: 2239-6152 - Copyright © 2023 Carlo Bianchini,

Andrea Marchitelli, Alessandra Moi



in linea<sup>2</sup>; la rubrica annuale è disponibile al pubblico solo in versione cartacea e non ha una cumulazione pluriennale.

I dati da raccogliere dalle riviste di settore per svolgere accurate analisi bibliometriche sono diversi e raggruppabili in quattro tipologie:

1. i dati bibliografici relativi agli articoli (titolo, data, rivista, volume e fascicolo, pagine ecc.);
2. i dati relativi agli autori (i cui nomi si trovano associati agli articoli pubblicati sulle riviste, ma per i quali è necessario fornire informazioni più ricche tratte anche da altre fonti di natura bio-bibliografica);
3. i dati relativi agli argomenti trattati (per esempio, parole chiave)<sup>3</sup>;
4. i dati relativi alle citazioni, cioè sia alle fonti citate negli articoli pubblicati nelle riviste italiane oggetto di analisi (per esempio, opere citate più frequentemente, e quindi di probabile maggiore rilevanza nella produzione scientifica italiana) sia le citazioni ricevute dagli articoli stessi in qualsiasi altra sede (per esempio, in monografie italiane o straniere, o in riviste straniere).

Gli articoli pubblicati sulle riviste italiane di biblioteconomia sono davvero molti (la sola rivista *Biblioteche oggi*, per esempio, ne conta oltre 4.500 nell'arco di tempo dal 1993 al 2020) e, come si è visto, devono essere integrati con i dati relativi agli autori - persone ed enti - ai concetti e agli altri oggetti bibliografici collegati. Perciò per il progetto si è reso indispensabile individuare uno strumento molto potente per la registrazione, gestione e analisi dei dati, bibliografici e non.

A tale scopo è stato scelto come strumento di lavoro Wikidata<sup>4</sup>: si tratta di un database libero, collaborativo e multilingue che raccoglie in modo centralizzato dati strutturati per fornire supporto ai progetti di Wikimedia (per esempio Wikipedia, Wikimedia Commons, Wikisource ecc.). Wikidata, tuttavia, fornisce supporto anche a molti altri siti e servizi del web, grazie al fatto che il contenuto di Wikidata è disponibile sotto licenza libera, è esportabile usando formati standard e può essere facilmente interconnesso ad altri insiemi di dati aperti. Un altro punto di forza di Wikidata è quello di essere una base di conoscenza che può essere letta e modificata allo stesso modo da umani e macchine.

Oltre a questi interessanti e rilevanti aspetti, Wikidata si presenta anche come un *hub*, cioè uno strumento che, sulle entità che descrive, raccoglie informazioni di qualità e provenienti da fonti diverse; ciò avviene attraverso il raggruppamento o *cluster*<sup>5</sup> di identificatori (ID) provenienti dal web semantico che viene creato in una

<sup>2</sup> Secondo Internet Archive, la pagina di appoggio per la versione in linea di *BIB* (<[http://www.bibliotechevicentine.it/bib/Homebib\\_ita.htm](http://www.bibliotechevicentine.it/bib/Homebib_ita.htm)>) non è più raggiungibile dal 24 febbraio 2017.

<sup>3</sup> Sui problemi di indicizzazione per soggetto degli articoli pubblicati sulle riviste italiane di biblioteconomia, si veda nota 64.

<sup>4</sup> Cfr. <<https://www.wikidata.org/wiki/Wikidata:Introduction>>.

<sup>5</sup> Il termine 'cluster' viene utilizzato nell'ambito della *cluster analysis*, che è uno strumento di esplorazione dei dati che deve essere affiancato da tecniche per la visualizzazione dei dati; nella *cluster analysis* «gli algoritmi di clusterizzazione raggruppano oggetti, o item di dati, sulla base di indici di prossimità tra coppie di oggetti»; cfr. Anil K. Jain, Richard C. Dubes, *Algorithms for clustering data*. Englewood Cliffs (NJ): Prentice Hall, 1988, p. 8. Un cluster è quindi un insieme di oggetti (o entità) o di dati di un oggetto (o entità). Anche se la creazione dei cluster avviene di solito mediante algoritmi, e quindi in modo automatico, il termine cluster può indicare un insieme di dati relativi a un oggetto, come nel caso di Wikidata dove si ha un insieme di identificatori associati tutti alla medesima entità (item).

delle quattro parti della descrizione di un item previste da Wikidata. Infine, ma non meno importante, essa è uno strumento gratuito a disposizione di tutti.

I vantaggi dell'utilizzo di Wikidata come *dataset* per la creazione, elaborazione e condivisione dei dati sono numerosi. Wikidata, infatti, consente di migliorare o mettere in atto numerose funzioni di essenziale importanza: l'identificazione delle entità (*entity management*), l'arricchimento attraverso il riuso di dati provenienti da fonti diverse, l'integrazione di dati provenienti da tali fonti. Tutto ciò garantendo la qualità, l'indipendenza dei dati da hardware e software, la loro accessibilità come presupposto per la loro condivisione e interoperabilità, il decentramento del lavoro di creazione e gestione attraverso la modularità, la possibilità di riuso da parte di terzi, la visualizzazione di tutti i dati tramite strumenti grafici e la loro analisi. Un aspetto altrettanto importante per un progetto scientifico è che Wikidata consente di indicare la fonte di ogni singolo dato relativo a un'entità (per esempio indicando una fonte sulla quale si trova la data di nascita di una persona e un'altra che ne riporta la data di morte; oppure due fonti che riportano date di nascita o di morte divergenti, documentando l'origine di uno specifico dato).

Tuttavia, ai fini del progetto di creazione dei metadati relativi alle riviste italiane di biblioteconomia su Wikidata, come strumento per indagare gli sviluppi più recenti della biblioteconomia e della bibliografia italiana, alcune funzioni offerte da Wikidata come insieme di dati in formato LOD sono particolarmente interessanti: l'identificazione, la visualizzazione e l'analisi e, infine, il riuso.

Un processo fondamentale nel lavoro con i dati bibliografici è l'identificazione degli autori. Le grandi banche dati citazionali commerciali - come Scopus o Web of science - solo di recente, con l'aiuto di identificatori come ISNI e ORCID, hanno iniziato ad affrontare il problema dell'ambiguità nell'identificazione degli studiosi di cui avevano sempre sofferto<sup>6</sup>. Wikidata prevede invece che oltre ai dati descrittivi (proprietà) relativi a un item, si registrino nel modo più sistematico possibile anche gli identificatori di altri servizi. Perciò un item relativo a una persona si può collegare contemporaneamente alla rappresentazione di quella medesima entità descritta in molti altri servizi, come i cataloghi delle biblioteche nazionali, o i fornitori di identificatori per i ricercatori come ORCID, VIAF, ISNI, con gli IRIS e altri depositi istituzionali e i social come LinkedIn, Twitter, Facebook, ResearchGate ecc. Grazie al collegamento di molti identificatori a un item, quando i dati relativi agli autori di una certa rivista vengono collegati con Wikidata, diventa possibile identificare quell'autore a vantaggio di quella specifica rivista di partenza, ma anche per tutte le altre riviste in cui ha eventualmente pubblicato e tutti gli altri servizi del web in cui è identificato (per esempio, l'IRIS dell'università a cui afferisce).

È questo che si intende quando si parla di Wikidata come un «hub del web semantico»<sup>7</sup>. La registrazione degli ID di molti *dataset* all'interno di un singolo item garantisce che, in fase di esplorazione e analisi dei dati, le interrogazioni al sistema abbiano due caratteristiche fondamentali per ogni sistema di recupero dell'informazione, ovvero precisione e richiamo.

In Figura 1 si può vedere un'applicazione concreta del processo di identificazione in Wikidata. Per ogni 'Silvia Pellegrini' finora individuata e inserita in Wikidata,

<sup>6</sup> Si veda tra gli altri il lavoro di Thomas Krämer, Fakhri Momeni, Philipp Mayr, *Coverage of author identifiers in Web of Science and Scopus*, «arXiv», 3 marzo 2017, DOI: 10.48550/arXiv.1703.01319.

<sup>7</sup> Linked Data for Production, *Wikidata as a hub for identifiers*. 11 giugno 2020, <<https://t.ly/677QP>>.

esiste un item dedicato e opportunamente distinto e descritto; a ciascuno di questi item vengono aggiunti gli identificatori di ciascuna diversa entità. Le proprietà descrittive di ciascun item in Wikidata consentono l'identificazione di Silvia Pellegrini all'interno di Wikidata, gli identificatori esterni collegati a quell'item consentono l'identificazione di quella entità negli altri *dataset* del web semantico<sup>8</sup>.

Search Input
Silvia Pellegrini
<b>Silvia Pellegrini</b> biomedica italiana
<b>Silvia Pellegrini</b> astrofisica italiana
<b>Silvia Pellegrini</b> patologa italiana (1965-)
<b>Silvia Pellegrini</b> teologa italiana (1965-)
<b>Silvia Pellegrini</b> archeologa italiana (1970-)
<b>Silvia Pellegrini</b> traduttrice italiana (1983-)
<b>Silvia Pellegrini</b> giurista italiana (1972-)

Figura 1 – Interrogazione di Wikidata (particolare)

Wikidata offre anche un secondo vantaggio molto rilevante per gli sviluppi del progetto, che consiste nella possibilità di analizzare, anche visivamente, i dati relativi alle pubblicazioni in modo più complesso di quando non si possa fare attraverso i normali strumenti bibliografici. Da quando è nato, nel 2012, Wikidata è sempre più utilizzato come strumento per la pubblicazione di dati bibliografici e fonte per la loro analisi.

Uno dei *tool* più noti che utilizzano i dati di Wikidata ai fini di analisi bibliometriche è Scholia<sup>9</sup>, creato da Finn Årup Nielsen. Lo scopo del sito è creare profili istantanei – e quindi sempre aggiornati grazie ai dati presenti su Wikidata – per diversi possibili oggetti di interesse: singoli ricercatori, organizzazioni (come un dipartimento universitario o un istituto di ricerca), riviste accademiche, editori, singole opere scientifiche e argomenti di ricerca<sup>10</sup>.

<sup>8</sup> Per esempio <<https://www.wikidata.org/wiki/Q102420507>>.

<sup>9</sup> Cfr. <<https://scholia.toolforge.org/>>; vedi anche <<https://www.wikidata.org/wiki/Wikidata:Scholia>>.

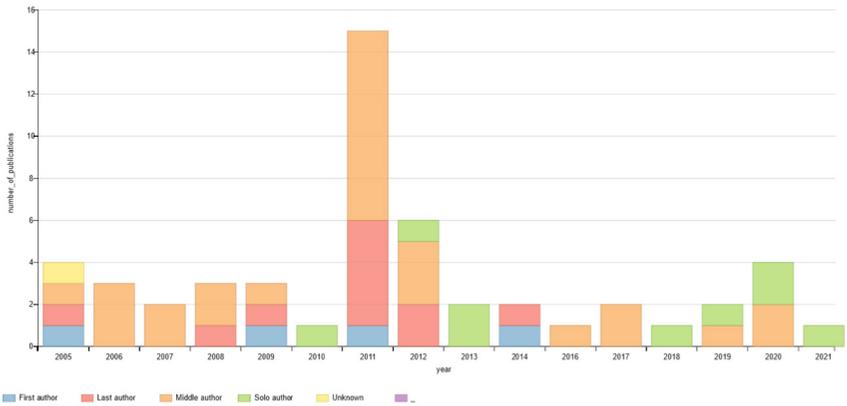
<sup>10</sup> Finn Årup Nielsen, Daniel Mietchen, Egon Willighagen, *Scholia, scientometrics and Wikidata*. In: *The Semantic Web: ESWC 2017 Satellite Events: ESWC 2017 Satellite Events, Portorož, Slovenia, May 28 – June 1, 2017, revised selected papers*, edited by Eva Blomqvist [et al.]. Cham: Springer International Publishing, p. 237-259, DOI: 10.1007/978-3-319-70407-4\_36; Mairelys Lemus-Rojas, Jere D. Odell, *Creating structured linked data to generate scholarly profiles: a pilot project using Wikidata and Scholia*, «Journal of librarianship and scholarly communication», 6 (2018), n. 1, art. eP2272, DOI: 10.7710/2162-3309.2272; Daniel Mietchen, Lane Rasberry, *Presenting Scholia: a scholarly profiling tool*. 2020,

Per esempio, Scholia permette di ottenere un elenco di tutte le pubblicazioni di un ricercatore (come fa un normale catalogo), ma offre anche molti altri punti di vista pronti per un'analisi approfondita (Figura 2): mostra un grafico con il numero delle pubblicazioni per anno (Figura 2.1) e un altro grafico con il numero di pagine pubblicate per anno (Figura 2.2; infatti una monografia è più impegnativa da produrre rispetto ad alcuni articoli, ma essendo rappresentata nel primo grafico da una sola pubblicazione, potrebbe mostrare una flessione nella continuità del lavoro che in realtà non c'è stata). Offre una visualizzazione tabellare, ma anche un grafico (Figura 2.3), con gli argomenti più trattati da quell'autore e una rappresentazione a grafo delle persone con cui collabora (Figura 2.4) ecc.

Showing 1 to 10 of 52 entries

Previous 1 2 3 4 5 6 Next

Number of publications per year

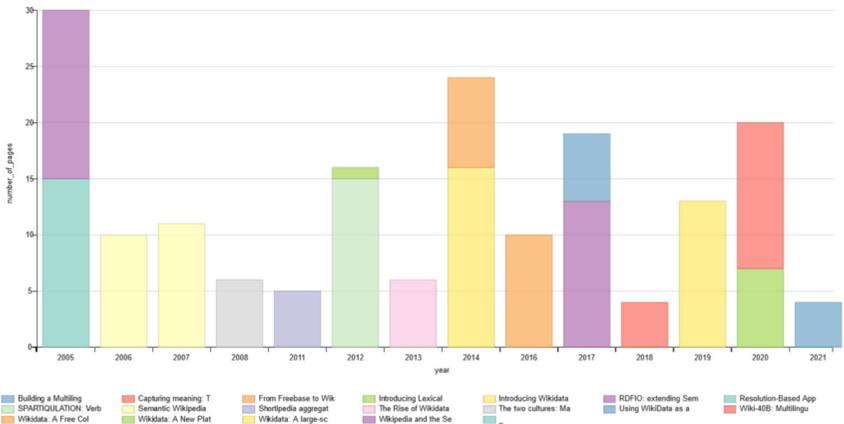


Wikidata Query Service

author: publications-per-year.sparql

Number of pages per year

(Only articles with number of pages set are displayed)



<[https://docs.google.com/presentation/d/1jYbYSnYSDh36-LxzSpedFyWUzusZAJuBbP-y46ji-ow/edit#slide=id.g35f391192\\_00](https://docs.google.com/presentation/d/1jYbYSnYSDh36-LxzSpedFyWUzusZAJuBbP-y46ji-ow/edit#slide=id.g35f391192_00)>. Una bibliografia su Wikidata completa e aggiornata automaticamente sulla base dei dati disponibili in Wikidata stessa è reperibile all'indirizzo <[https://www.wikidata.org/wiki/Wikidata:Bibliography\\_of\\_Wikidata](https://www.wikidata.org/wiki/Wikidata:Bibliography_of_Wikidata)>.



Home / Dettagli di: Semiotica e filosofia del linguaggio /

Visualizzazione normale Vista MARC Visualizzazione ISBD

**Semiotica e filosofia del linguaggio / Umberto Eco.**

Per: Eco, Umberto [autore]

Serie: Biblioteca studio ; 29.

Dettagli di pubblicazione: Torino : G. Einaudi, 1996

Descrizione: XVII, 318 p. : ill. ; 21 cm

ISBN: 8806140655

Classificazione Dewey: 401.41

Dewey: 401.41 SEMIOTICA

N.	Collocazione	Localizzazione	Copia	Barcode	In prestito fino al	Note di copia
1	401.41 / 2	Generale	Copy 001	160401	Disponibile	

**AuthorityBox** ↓

Eco, Umberto



05/01/1932 Alessandria -  
19/02/2016 Milano

Italiano; francese

Eco, Umberto Filosofo, scrittore, semiologo, professore emerito e presidente della Scuola Superiore di Studi Umanistici dell'Università di Bologna (2008). studioso di estetica.

Figura 3 – Esempio di riuso di dati Wikidata nell'AuthorityBox del Catalogo PUSC

Dal momento che l'esito dell'analisi della produzione di un ricercatore, di un dipartimento, di una rivista ecc. attraverso Wikidata o Scholia dipende strettamente dalla disponibilità dei dati di partenza su Wikidata stessa, per sviluppare il progetto di indagine sugli sviluppi più recenti della biblioteconomia e della bibliografia italiana era indispensabile in via strumentale procedere al caricamento di tutti i dati necessari su Wikidata.

Si tratta di un lavoro problematico e complesso per diverse ragioni. Prima di tutto la disponibilità difforme dei dati di partenza: i dati si possono trovare in formati digitali più o meno accessibili e riutilizzabili (per esempio, in formato PDF o CSV) o essere disponibili in solo formato cartaceo (come si è visto). In secondo luogo, i dati possono non essere omogenei, sia all'interno della stessa rivista (registrazione non uniforme dei nomi e delle parole chiave, eterogeneità nei formati disponibili per lo scarico) sia tra riviste diverse. In terzo luogo, il caricamento dei dati su Wikidata può avvenire con procedimenti diversi: semplificando in modo estremo, si può procedere in modo manuale (registrando un dato per volta), in modo semi-automatico (registrando un blocco di dati per volta e procedendo in seguito al loro miglioramento) o in modo automatico (e procedendo poi al loro controllo)<sup>12</sup>.

Il caricamento dei dati bibliografici su Wikidata ha quindi posto diverse questioni: considerata la complessità della procedura, esiste un metodo preferibile per procedere al caricamento dei dati? Esistono metodi, strumenti e tecniche per il controllo dei dati? Ne esistono per facilitare l'identificazione degli autori? E per l'arricchimento degli item (cioè un'entità descritta) in Wikidata? Quale modello cooperativo si può mettere in atto per fare in modo che il processo di aggiornamento dei dati bibliografici si sostenga nel tempo? Esistono procedure per migliorare i dati citazionali? I dati citazionali ottenibili sono comparabili con quelli di altri strumenti citazionali, almeno per quanto riguarda la situazione italiana?

### Materiali e metodi

<sup>12</sup> Alessandra Boccone, Remo Rivelli, *I metadati bibliografici in Wikidata: Wikicite e il case study di «Bibliothecae.it»*, «Bibliothecae.it», 8 (2019), n. 1, p. 227-248, DOI: 10.6092/issn.2283-9364/9503; Carlo Bianchini, *Wikidata for J LIS.it: a new step forward mapping Italian library and information science journals*, «J LIS.it», 12 (2021), n. 1, p. 29-38, DOI: 10.4403/jlis.it-12680.

Prima di approfondire la più complessa questione relativa alle modalità di caricamento dei dati, risulta utile chiarire alcuni passaggi preliminari che si sono rivelati fondamentali per la corretta impostazione del lavoro. In primo luogo, il censimento puntuale di tutte le riviste italiane di biblioteconomia, cessate e/o attive.

#### *Individuazione delle riviste di interesse*

Questa ‘rassegna’ ha costituito una base di partenza imprescindibile, basti pensare che ciascuna rivista italiana si basa su diverse modalità di conservazione e strutturazione dei dati e offre sistemi di interrogazione più o meno aperti verso l'esterno. A partire da un'apposita query di Wikidata<sup>13</sup>, sono state individuate 25 riviste di biblioteconomia afferenti al contesto italiano: per ciascuna rivista si è poi deciso di stabilire una priorità per il caricamento sulla base dell'autorevolezza, ma anche del periodo di attività, della rivista<sup>14</sup>. I criteri di scelta della priorità sono stati anche determinati dalla relativa maggiore fattibilità del caricamento dei dati e non rispecchiano assolutamente una valutazione di ordine gerarchico o qualitativo delle riviste stesse.

#### *Definizione del modello dei dati*

Il secondo passaggio è stato definire il modello logico dei dati, cioè un'ontologia di riferimento per il progetto. Si è trattato cioè di: individuare chiaramente quali entità, ovvero item, sarebbero state oggetto delle attività di caricamento e successivo arricchimento; stabilire quali dati, espressi sotto forma di proprietà, si sarebbero dovuti associare a ciascuna entità; stabilire il livello minimo di completezza potenzialmente raggiungibile. Trattandosi di un progetto finalizzato alla valorizzazione disciplinare della biblioteconomia italiana attraverso le riviste e le pubblicazioni di settore, gli item di principale interesse sono stati logicamente individuati in: articoli di ciascuna rivista, autori a essi associati, riviste italiane individuate, parole chiave degli articoli.

In tutti i casi, si tratta di entità molto ben definite e documentate in Wikidata, per le quali sono già state rese disponibili un'ampia serie di proprietà: in particolar modo per gli articoli e per gli autori è stata fatta una precisa selezione, volta a individuare tutte quelle proprietà giudicate imprescindibili ai fini di una corretta identificazione degli item. Tali liste sono mantenute e messe a disposizione della comunità<sup>15</sup>.

Per gli articoli, le principali proprietà individuate come tali – oltre all'etichetta (*label*), che è il nome in linguaggio umano, la descrizione e le eventuali forme varianti del nome<sup>16</sup> – sono state:

- istanza di (P31);
- titolo (P1476);
- autore (P50);
- lingua dell'opera o del nome (P407);
- data di pubblicazione (P577);

**13** <<https://w.wiki/5hcV>>.

**14** Nella pagina del progetto è disponibile una griglia relativa allo stato di avanzamento dei lavori, in cui sono elencate tutte le riviste di interesse unitamente a un punteggio numerico che indica il livello di priorità: <[https://www.wikidata.org/wiki/Wikidata:Gruppo\\_Wikidata\\_per\\_Musei,\\_Archivi\\_e\\_Biblioteche/Riviste\\_di\\_biblioteconomia](https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Riviste_di_biblioteconomia)>.

**15** Per gli articoli si veda nello specifico: <[https://www.wikidata.org/wiki/Template:Bibliographic\\_properties](https://www.wikidata.org/wiki/Template:Bibliographic_properties)>.

**16** I metadati *Etichetta* e *Descrizione* contengono i dati visibili in Figura 1.

- pubblicato in (P1433);
- volume di un'opera (P478);
- numero della pubblicazione (P433);
- pagina (P304);
- numero di pagine (P1104);
- URL (P2699);
- identificativo DOI (P356).

The catalogue (and the cataloguer) : a defenseless entity? (Q61696527)

journal article from 'JLIS.it' published in 2012

Other properties	
number of pages	11
published in	JLIS.it Italian Journal of Library, Archives, and Information Science
title	Il catalogo (e il catalogatore): vaso di coccio tra vasi di ferro? [it]
URL	<a href="https://jlis.fupress.net/index.php/jlis/article/view/288">https://jlis.fupress.net/index.php/jlis/article/view/288</a>
page(s)	1-11
instance of	scholarly article article in an academic publication, usually peer reviewed
issue	1
volume	3
author	Carlo Revelli Italian librarian (1926-2026)
publication date	2012
main subject	cataloging process of creating meta-data for information resources to include in a catalog database
	librarian person who works professionally in a library, and is usually trained in librarianship
	library science application of practices, methods, technologies, tools of management, education, training, perspectives, and related areas, to the organized collection and the systematic classification of information in libraries

Free images Google search

External sources	
DOI	10.4403/JLIS.IT-5621
Scholia	Scholia

Wikimedia projects

Concept cloud

Figura 4 – Esempio di articolo con relative proprietà, visualizzato tramite Reasonator<sup>17</sup>

L'analisi del lavoro di caricamento dei dati relativi agli autori ha mostrato la differenza tra un approccio orientato alle entità e quello tradizionalmente orientato alla registrazione degli oggetti di interesse catalografico. Si tratta di una differenza che si può rilevare già nella prima edizione di RDA (Resource description and access), lo standard di metadateazione internazionale derivato dalla trasformazione delle AACR<sup>218</sup>. Anche se i dati che si registrano sono sostanzialmente i medesimi, muta l'approccio concettuale. L'autorialità, infatti, è solo uno dei possibili aspetti con i quali è possibile descrivere una persona e, assumendo un diverso punto di vista focalizzato sull'entità, ci si colloca in una prospettiva nella quale una persona non è necessariamente un

<sup>17</sup> Reasonator è uno strumento gratuito per la visualizzazione delle informazioni del database di Wikidata in una forma leggibile dalle persone e sviluppato da Magnus Manske, uno dei più attivi e importanti volontari di Wikimedia. È accessibile, con alcuni esempi pronti, al seguente indirizzo: <<https://reasonator.toolforge.org/?>>.

<sup>18</sup> «Unlike codes of the past, RDA ties the description and access to the user tasks and the specific entities that we are describing [...]. Yet RDA has ties with the codes of the past in the objectives to collocate all the works of a creator, all the expressions of a work, all the manifestations of an expression, all the items held. RDA enables us to indicate other related resources and information, to identify the persons, families, corporate bodies, and places, so we can collocate things by or about those entities», Barbara B. Tillett, *RDA and the semantic web, linked data environment*, «JLIS.it», 4 (2013), n. 1, p. 139-145: p. 142, DOI: 10.4403/jlis.it-6303.

autore. Perciò un autore è, prima di tutto, una persona o un ente (per esempio, una famiglia o una istituzione) in grado di agire. In questo senso, è un agente, cioè «un'entità capace di azioni intenzionali, di godere di diritti e di essere ritenuta responsabile delle proprie azioni»<sup>19</sup>. Un agente può svolgere attività anche molto differenti tra loro che lo qualificano per esempio, di volta in volta, come autore (in molti possibili ruoli come autore principale o secondario, curatore, traduttore, prefatore ecc.), editore, proprietario, pittore, scultore, restauratore, fornitore ecc. Le proprietà di norma attribuite a una persona nelle anagrafi a esse destinate sono relative alla sua essenza (nome e cognome, data e luogo di nascita, sesso, lingua e nazionalità e così via) e, in misura proporzionata, alle sue attività (di autore ecc.). È evidente che le proprietà attribuibili all'essenza della persona saranno quelle di maggiore interesse per la generalità dei servizi del web semantico (potranno interessare biblioteche, ma anche archivi, musei e altre istituzioni culturali e di diversa natura) mentre i dati relativi all'attività di autore, supponiamo, saranno di interesse primario per le biblioteche<sup>20</sup>.

Perciò, gli autori rientrano in Wikidata nella più vasta categoria degli 'umani' (il termine utilizzato in Wikidata per identificare l'entità persona), ai quali, come si potrà immaginare, è possibile attribuire svariate decine di proprietà. A fronte di questo, le informazioni di norma fornite dalle riviste sugli autori sono estremamente ridotte: solo di recente si è consolidata la prassi di registrare affiliazioni e altri dati rilevanti come l'ORCID.

Per questo motivo, si è scelto di optare inizialmente per un numero minimo di attributi - da registrare se disponibili sulle fonti utilizzate - con l'idea passare a una successiva attività di arricchimento manuale in una seconda fase del progetto:

- istanza di (P31);
- sesso o genere (P21);
- prenome (P735);
- cognome (P734);
- paese di cittadinanza (P27);
- data di nascita (P569);
- data di morte (P570);
- luogo di nascita (P19);
- luogo di morte (P20);
- occupazione (P106);

**19** International Federation of Library Associations and Institutions, *IFLA Library reference model: a conceptual model for bibliographic information*, edited by Pat Riva, Patrick Le Boeuf, Maja Zumer. Den Haag: IFLA, 2017, p. 28.

**20** Ciò non significa che, automaticamente, sia compito delle biblioteche registrare tutti quei dati. Si possono tuttavia fare due considerazioni in proposito. La prima è che già da tempo è esperienza comune nelle biblioteche che quei dati diventano indispensabili alle biblioteche stesse quando occorre distinguere tra omonimi: in questi casi, la data e il luogo di nascita, la professione, le notizie biografiche ecc. diventano di interesse primario e una necessità per le biblioteche. In secondo luogo, non tutte le biblioteche sono strutturate – in termini finanziari e di personale, se non di competenze – per svolgere questo lavoro; nonostante ciò, il lavoro qualificato di alcune biblioteche ha una ricaduta positiva sull'intero sistema. È questa la prospettiva con la quale guardare ai dati che si associano alle persone: sono un bene comune, di cui altri si potranno avvantaggiare. Senza contare che, in un sistema aperto come il web semantico, è una certezza che anche le biblioteche si possono avvantaggiare del bene comune prodotto da istituzioni diverse e condiviso nelle modalità opportune.

- lingue parlate o scritte (P1412);
- identificativo VIAF (P214);
- identificativo ISNI (P213);
- identificativo SBN (P396);
- identificativo LCNAF (P244);
- identificativo BNF (P268);
- identificativo GND (P227).

Di queste proprietà, solamente le prime quattro sono state considerate obbligatorie.

Non è stato quasi mai necessario creare gli item delle riviste, già esistenti su Wikidata. Le più importanti proprietà associate alle riviste sono riassunte nell'apposito *template* dedicato alle *bibliographic property*<sup>21</sup>, tra cui quelle essenziali possono essere individuate in:

- istanza di (P31);
- titolo (P1476);
- data di fondazione o creazione (P571);
- fondatore (P112);
- editore (P123);
- paese di origine (P495);
- sito web ufficiale (P856);
- frequenza di pubblicazione (P2896).

Ben diversa e più impegnativa, invece, è stata la gestione delle parole chiave. La principale differenza rispetto alla procedura applicata per gli item di articoli, autori e riviste, è che per l'inserimento delle parole chiave si è scelto di non procedere necessariamente alla creazione dei relativi item in Wikidata.

In Wikidata non esistono item la cui proprietà "istanza di" abbia come valore "parola chiave", "termine di soggetto", "stringa di soggetto" o simili. Che un concetto sia soggetto di un'opera, infatti, è una questione 'relazionale' e non 'ontologica': i soggetti non esistono in forma a sé stante, ma qualunque oggetto o concetto rappresentato da un item può essere soggetto di una pubblicazione, rappresentata da un altro item, stabilendo tra i due una relazione "è soggetto di". Questo approccio deriva dal modello logico che Wikidata adotta per esprimere il soggetto: si tratta di una relazione tra due item, nella quale la qualifica di soggetto è espressa dalla proprietà P921 ("main subject"). Tale modello è in perfetta sintonia con quello di IFLA LRM, nel quale si prevede una relazione "has as subject" (LRM-R12) per collegare un'opera a una res. Un esempio concreto, preso dall'attività che abbiamo condotto all'interno del progetto, è quello dell'item "bibliotecario" (Q182436), spesso utilizzato come parola chiave negli articoli biblioteconomici; l'item in questione è una istanza di "professione" e non una parola chiave o un soggetto in sé. È solo la relazione con gli articoli - espressa dalla proprietà "argomento principale" (P921) - che ne certifica in quello specifico caso la funzione di parola chiave.

All'interno di questo progetto, la decisione di non creare gli item utilizzabili come parole chiave è stata dunque motivata dalla complessità di individuare di volta in volta la tipologia di item corretta e le relative proprietà; questa attività avrebbe certamente portato a un aggravio del lavoro e a un allungamento complessivo dei tempi. Il legame tra parole chiave e articolo è stato quindi valorizzato solo quando esisteva già un item per quel concetto, senza doverlo creare.

<sup>21</sup> Si veda *Riviste scientifiche* alla pagina <[https://www.wikidata.org/wiki/Template:Bibliographic\\_properties](https://www.wikidata.org/wiki/Template:Bibliographic_properties)>.

### *Il processo di caricamento dei dati in Wikidata*

Dopo aver definito gli item da creare e le relative proprietà, è iniziata la fase di ricerca di una metodologia di lavoro che fosse la più rapida e semplice possibile, e la più facilmente replicabile. Questo obiettivo non era banale, perché in conflitto con la grande eterogeneità di sistemi, anche tecnologici, ai quali ciascuna rivista si affida per la pubblicazione e l'accessibilità dei propri dati.

Il punto di partenza è stato la messa alla prova di metodi già sperimentati in progetti analoghi. Il primo esempio risale al 2019, quando Alessandra Boccone e Remo Rivelli pubblicarono i risultati del primo caricamento in Wikidata di dati bibliografici provenienti da una rivista online di area italiana, ovvero *Bibliothecae.it*<sup>22</sup>. Il processo di caricamento, che è tuttora utilizzato dal gruppo redazionale di *Bibliothecae.it*, prevede un sistema semi-automatizzato, in cui alla creazione manuale degli autori segue il caricamento degli articoli da Zotero a Wikidata tramite il tool Zotkat<sup>23</sup>. Sono tuttavia necessari interventi manuali per l'inserimento delle citazioni bibliografiche originali in Zotero, per la creazione delle relazioni (per esempio, tra l'item dell'articolo con dati provenienti da Zotero e l'autore precedentemente creato manualmente) e per la necessaria revisione dei dati caricati<sup>24</sup>.

Considerata la significativa componente di lavoro manuale, il flusso di lavoro utilizzato da *Bibliothecae.it* risultava poco adatto alla grande mole di dati che si prevedeva sarebbe stato necessario caricare su Wikidata.

Si è perciò proceduto alla ricerca di strumenti tramite i quali gestire grandi quantità di dati e che offrissero anche moduli di integrazione con Wikidata. L'unico strumento che al momento consentiva azioni di import/export massivo verso grandi *dataset* di autorità era OpenRefine<sup>25</sup>.

Il problema successivo era avere a disposizione i dati degli articoli in un formato adatto a essere utilizzato con OpenRefine (OR).

Dopo alcuni tentativi di *web scraping*<sup>26</sup> – un approccio che dava risposte molto eterogenee dovute alle differenti modalità di creazione dei siti web delle riviste di interesse – la soluzione vincente è stata, in questo caso, quella di sfruttare le potenzialità

**22** Cfr. A. Boccone, R. Rivelli, *I metadati bibliografici in Wikidata cit.*

**23** L'interazione tra Wikidata e il *reference management software* Zotero (<<https://www.zotero.org/>>) viene garantita dall'apposita estensione Zotkat (<<https://github.com/UB-Mannheim/zotkat>>). A questa si è aggiunto più di recente il tool Cita (<<https://www.wikidata.org/wiki/Wikidata:Zotero/Cita>>) che, seppur ancora in fase di sviluppo, permette un aggancio diretto tra le citazioni bibliografiche di Zotero e i corrispondenti item in Wikidata, laddove esistenti.

**24** Questo aspetto dipende in buona misura dal fatto che il colloquio tra Zotero e Wikidata si basa sul tool QuickStatements (<<https://quickstatements.toolforge.org/#/batch>>). QuickStatements consente di effettuare azioni di modifica, creazione ed eliminazione di item in modalità *batch*, attraverso l'uso di una precisa serie di comandi e di una strutturazione tabellare dei dati. Perciò le proprietà di Wikidata espresse come relazioni tra entità non possano essere caricate tramite QuickStatements (a meno che le entità verso le quali punta una relazione non siano state tutte precedentemente identificate).

**25** <<https://openrefine.org/>>.

**26** Il *web scraping* consiste nell'estrarre dati strutturati nell'HTML delle pagine web e raccogliarli in database o tabelle locali per analizzarli o riutilizzarli (Andrea Fumagalli, *Web scraping: cos'è, perché si usa e come difendersi da "intrusioni" indesiderate*, «AgendaDigitale.eu», 23 aprile 2021, <<https://www.agendadigitale.eu/sicurezza/web-scraping-cose-perche-si-usa-e-come-difendersi-da-intrusioni-indesiderate/>>).

del software gestionale in uso presso un buon numero di riviste, ovvero OJS (Open Journal System)<sup>27</sup>, che prevede l'esposizione dei metadati pubblicati dalla rivista tramite il protocollo OAI-PMH<sup>28</sup>. L'adesione alla Open Archives Initiative delle riviste ad accesso aperto, inoltre, permette che tali metadati siano messi a disposizione attraverso licenze copyleft a diversi gradi di apertura, tali da consentirne il riuso.

Il protocollo OAI-PMH, interrogabile direttamente tramite OR, ha permesso di ottenere in maniera semplice e veloce i metadati completi degli articoli per le riviste *JLIS.it*, *AIB studi*, *Biblioteche oggi* e *Biblioteche oggi trends*, secondo una procedura che, in seguito a vari cicli di elaborazione e riflessione, può essere riassunta nei seguenti punti fondamentali<sup>29</sup>:

1. Raccolta dei soli ID degli articoli in OR tramite la specifica baseURL OAI-PMH (verb: ListIdentifiers);
2. Creazione di una seconda baseURL – basata sugli ID – per l'*harvesting* dei metadati sempre tramite OAI-PMH (verb: GetRecord);
3. *Parsing* dei metadati in OR tramite l'uso di comandi formulati il linguaggio GREL<sup>30</sup>;
4. Riconciliazione del *dataset* con Wikidata tramite i servizi API offerti da OR.

### Importazione dei dati su Wikidata

Non sempre tali passaggi si sono svolti in maniera lineare; in alcuni casi, a seconda della rivista trattata, si è reso necessario effettuare più cicli di importazione tramite il protocollo OAI-PMH, il quale è in grado di gestire più formati per l'esposizione dei metadati. Solitamente, quando presente, si è utilizzato il formato NLM (National Library of Medicine), perché forniva i dati più completi, ricchi e altamente strutturati, oppure in alternativa il formato OAI DC (in genere più carente poiché basato su Dublin Core simple).

Per ottenere i dati delle riviste che non prevedono l'utilizzo di OJS, essendo spesso periodici dalla grande tradizione e importanza scientifica e indispensabili per raggiungere le finalità del progetto (per esempio, *Accademie & biblioteche d'Italia* o *AIDAinformazioni*) si è deciso di contattare direttamente gli editori. La richiesta dei dati è stata quasi sempre accolta e ha permesso di raggiungere notevoli risultati, in termini di copertura e completezza delle informazioni. In particolare, grazie alla disponibilità di Casalini Libri<sup>31</sup>, è stato possibile ottenere i dati bibliografici di oltre 1.500 articoli pubblicati su *Accademie & biblioteche d'Italia*<sup>32</sup>, *AIDAinformazioni*<sup>33</sup> e *Il bibliotecario: rivista di studi bibliografici*<sup>34</sup>, che vengono indicizzate, grazie a specifici accordi con gli editori, sulla piattaforma Torrossa. Purtroppo, non è stato possibile ottenere i dati disaggregati degli articoli de *La Bibliofilia* e dei *Nuovi annali della scuola*

27 <<https://pkp.sfu.ca/software/ojs/>>.

28 <<https://www.openarchives.org/pmh/>>.

29 L'intero flusso di lavoro è comunque accuratamente descritto in <<https://www.protocols.io/workspaces/wikidata>>.

30 Il General Refine Expression Language (GREL) è il linguaggio utilizzato per lavorare e manipolare dati, celle e colonne in OpenRefine.

31 Gli autori ringraziano Michele Casalini per la pronta disponibilità e Luisa Gaggini per l'aiuto tecnico.

32 <<https://www.torrossa.com/en/resources/an/4541575>>.

33 Fino al 2010: <<https://www.torrossa.com/it/resources/an/2193001>>.

34 <<https://www.torrossa.com/en/resources/an/2467946>>.

*speciale per archivisti e bibliotecari*, perché, per «una scelta dell'editore e delle Direzioni delle riviste, non vengono resi disponibili su Torrossa i metadati a livello articolo» (e-mail di Luisa Gaggini, 20 gennaio 2021).

Inoltre, l'archivio dei numeri di *Biblioteche oggi* precedenti al 2008, i cui dati non sono disponibili tramite OAI-PMH, è stato fornito in formato XLSX da parte della redazione della rivista e ha permesso di ottenere i dati di circa 3.200 articoli, colmando una lacuna che sarebbe stata assolutamente significativa<sup>35</sup>.

*La qualità dei dati: correttezza, completezza e arricchimento*

La fase successiva al caricamento massivo dei dati è consistita nel controllo di qualità dei dati. I potenziali problemi noti erano due: gli errori dovuti a un'errata identificazione degli item di Wikidata in fase di riconciliazione dei dati con OR (per esempio, l'assegnazione di un articolo a un item che rappresenta un omonimo dell'autore reale) e l'esistenza di item descritti in modo molto scarso (per esempio, autori con le sole proprietà essenziali).

Per monitorare le riviste di interesse del progetto è stata assegnata alle riviste italiane di scienze dell'informazione e delle biblioteche una specifica proprietà che le metteva sulla *focus list* del GWMA (P5008)<sup>36</sup>. Attraverso questa marcatura è stato possibile avviare controlli mirati sulle principali entità oggetto del caricamento: gli articoli e gli autori.

Secondo lo stile di lavoro in Wikidata, i problemi relativi ai dati sono stati descritti in un'apposita pagina del progetto<sup>37</sup>, sulla quale venivano segnalati i punti critici, si offriva una query per individuare ed estrarre gli item relativi a ciascun caso problematico e si aggiornava lo status di risoluzione. In Wikidata, la pagina di un progetto, oltre a fornire un quadro dei problemi e il progresso del lavoro, ha il vantaggio di permettere a chiunque di intervenire e correggere gli errori, a partire proprio dalla descrizione del problema e dalla query predisposta. Il problema non è necessariamente limitato alle persone che hanno avviato il progetto, ma è proposto all'intera comunità di Wikidata e può essere gestito e risolto da qualsiasi volontario interessato.

Per gli articoli, la bonifica dei dati ha riguardato gli aspetti più disparati, dalla sistemazione di valori erroneamente doppi per alcune proprietà (data di pubblicazione, pagine) fino a casi di duplicazione identificati grazie alla presenza del medesimo DOI su item diversi.

Anche per gli autori sono stati previsti sistemi di monitoraggio, tramite l'elaborazione di specifiche query<sup>38</sup> per l'individuazione di autori descritti in maniera troppo povera (per esempio, privi del cognome e privi di qualsiasi identificatore esterno, come VIAF, SBN, ISNI, ORCID ecc.).

Il lavoro di revisione degli autori non è concluso ed è da considerarsi continuamente 'in corso' per diverse ragioni. In primo luogo, la natura automatizzata della procedura di creazione degli autori si è basata necessariamente su un numero davvero limitato di proprietà, unitamente a una descrizione molto stringata e generica («autore di ...»). In secondo luogo, non essendo disponibile un'opera di consultazione<sup>39</sup>

**35** Gli autori ringraziano Sara Speciani, Massimo Belotti e Piero Cavalieri per la disponibilità.

**36** La query è <<https://w.wiki/4Lgf>>.

**37** <[https://www.wikidata.org/wiki/Wikidata\\_talk:Gruppo\\_Wikidata\\_per\\_Musei,\\_Archivi\\_e\\_Biblioteche/Riviste\\_di\\_biblioteconomia](https://www.wikidata.org/wiki/Wikidata_talk:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Riviste_di_biblioteconomia)>.

**38** Una query di esempio è <<https://w.wiki/4abn>>.

**39** Non può essere SBN perché non effettua uno spoglio organico, sistematico e completo delle riviste.

contenente l'anagrafe onnicomprensiva degli autori, per gli item creati *ex novo* e privi di qualsiasi identificatore, le strategie di miglioramento della qualità devono necessariamente passare da azioni manuali sui singoli item. Infine, la grande mole di dati caricati rende oneroso il lavoro e molto lungo il processo.

### Risultati

Dall'avvio del progetto, nell'ottobre 2020, il gruppo di lavoro ha realizzato il caricamento su Wikidata di tutti gli articoli pubblicati nelle seguenti riviste nell'arco cronologico indicato tra parentesi:

1. *Accademie e biblioteche d'Italia* (2010-2018);
2. *AIB studi* (2012- )<sup>40</sup>;
3. *AIDAinformazioni* (1995-2007);
4. *Biblioteche oggi* (2013-2020);
5. *Biblioteche oggi trends* (2015-2020);
6. *Bollettino AIB* (1992-2011);
7. *Bollettino d'informazioni. Associazione italiana biblioteche* (1961-1991, con lacune di intere annate);
8. *DigItalia* (2005-2022);
9. *Il bibliotecario: rivista di biblioteconomia, bibliografia e scienza dell'informazione* (1984-2011);
10. *JLIS.it* (2010- )<sup>41</sup>;
11. *La Bibliofilia: rivista di storia del libro e di bibliografia* (1950-1979 [lac. 1951-52 e 1980-83] e 1984-2018 [lacunos]).

Ciascun articolo risulta corredato dei metadati essenziali (titolo, data di pubblicazione, numero di volume e fascicolo, DOI ecc.) ed è collegato alla rivista sulla quale è stato pubblicato e ai propri autori. Solo per il 18% degli articoli inseriti sono state collegate anche le parole chiave, perché molte riviste non le attribuiscono e perché si è scelto di collegare la parola chiave solo se era già rappresentata da un item in Wikidata, come già detto<sup>42</sup>.

Le riviste totalmente o parzialmente indicizzate sono quindi 17, su un insieme di 25 testate inizialmente individuate<sup>43</sup>. A queste si aggiungono *Bibelot*, che viene indicizzata autonomamente da un gruppo di bibliotecari toscani e *Bibliothecae.it*, da sempre in carico al gruppo redazionale della rivista<sup>44</sup>.

Nel complesso, come risultato del progetto e degli inserimenti occasionali e non sistematici operati da altri utenti di Wikidata, sono stati indicizzati in Wikidata 9.749<sup>45</sup> articoli, dei quali 8.382 sono articoli in senso stretto e 1.367 sono recensioni<sup>46</sup>, distribuiti per rivista come risulta da Figura 5.

<sup>40</sup> In aggiornamento continuo da parte della redazione.

<sup>41</sup> In aggiornamento continuo da parte della redazione.

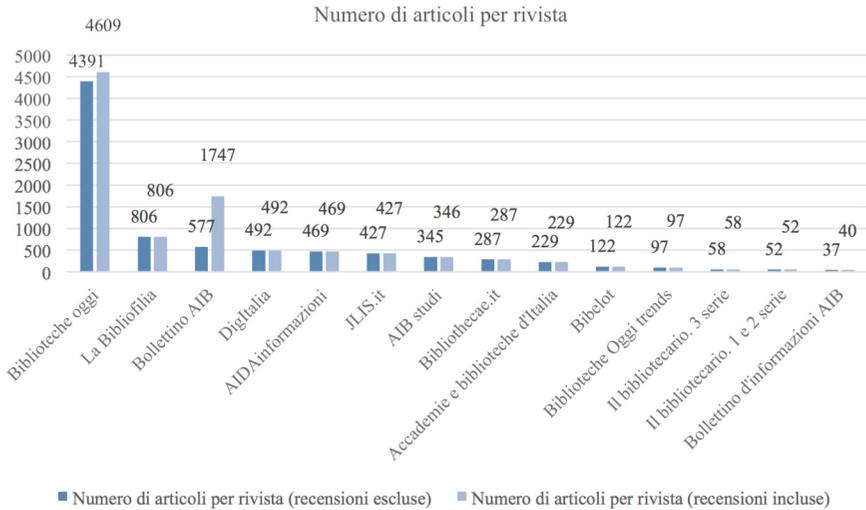
<sup>42</sup> La query è <<https://w.wiki/6iUH>>.

<sup>43</sup> La query è <<https://w.wiki/6jBg>>.

<sup>44</sup> Risultano prive di articoli inseriti in Wikidata le seguenti riviste: *Archivi & computer*, *Bollettino dell'Istituto centrale per la patologia del libro* Alfonso Gallo, *Bollettino di informazione ABEI*, *Culture del testo*, *Giornale della libreria*, *La rivisteria*, *Rara volumina* e *SBN notizie*.

<sup>45</sup> La query è <<https://w.wiki/6jyP>>.

<sup>46</sup> La query è <<https://w.wiki/6jBa>>.



**Figura 5** - Numero di articoli per rivista incluse<sup>47</sup> o escluse<sup>48</sup> le recensioni

Dalla Figura 5 emerge chiaramente l'importanza della rivista *Biblioteche oggi*, che è seguita per numero di articoli e recensioni dalla rivista *Bollettino AIB* e poi da *La Bibliofilia*, e a seguire da *Digitalia*, *AIDAinformazioni*, *JLIS.it*, *AIB studi*, *Bibliothecae.it*, *Accademie e biblioteche d'Italia* e le altre. Se si considerano il *Bollettino AIB* e *AIB studi* come un'unica rivista (dal momento che la seconda è prosecuzione della prima), si conferma l'importanza della rivista pubblicata dall'AIB. Il dato relativo a *La Bibliofilia*, che è stata fondata nel 1899, risulta decisamente più basso del reale dal momento che – come si è visto – il caricamento dei dati in Wikidata parte dal 1950 ma è anche lacunoso. È ragionevole pensare che sia proprio quest'ultima la rivista italiana di biblioteconomia che ha pubblicato il maggior numero di articoli.

Il numero degli autori per rivista rispecchia sostanzialmente la situazione degli articoli, con 5.123 autori per *Biblioteche oggi*, seguiti a molta distanza dai 1.861 autori del *Bollettino AIB*, gli 837 de *La Bibliofilia* (dato che presenta il limite che si è detto) e dalle altre a seguire<sup>49</sup>. Diversi sono i dati se si escludono dal conteggio gli autori di sole recensioni<sup>50</sup>, come si vede dal grafico di confronto (Figura 6).

<sup>47</sup> La query è <<https://w.wiki/6oaH>>.

<sup>48</sup> La query è <<https://w.wiki/6oaG>>.

<sup>49</sup> La query è <<https://w.wiki/6jBf>>.

<sup>50</sup> La query è <<https://w.wiki/6kzT>>.

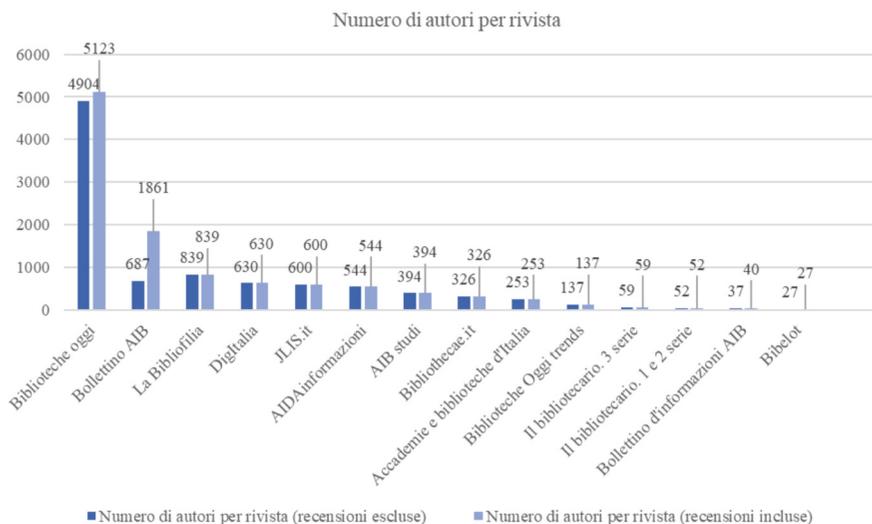


Figura 6 - Numero di autori per rivista

Un altro dato quantitativo che si può trarre dall’elaborazione consiste nell’individuazione degli autori con il numero più alto di articoli pubblicati sulle riviste italiane di biblioteconomia e inseriti in Wikidata (Figura 7) includendo<sup>51</sup> o escludendo<sup>52</sup> le recensioni. Si osserva, analogamente ai grafici precedenti, una lieve variazione dei dati in base al conteggio dei soli articoli o degli articoli e delle recensioni.

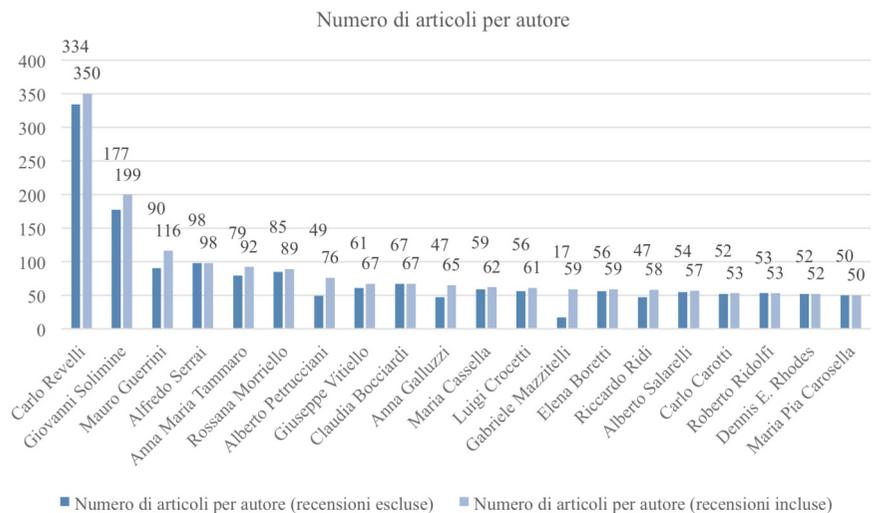


Figura 7 - Autori con almeno 50 articoli censiti in Wikidata (aprile 2023)

51 La query è <<https://w.wiki/6XrQ>>.

52 La query è <<https://w.wiki/6oaP>>.

Anche in questo caso, i dati attualmente presenti risentono del caricamento parziale degli articoli della letteratura, oppure di progetti particolari dedicati all'inserimento della produzione totale di un autore specifico che possono in qualche modo falsare la panoramica complessiva delle pubblicazioni.

Questo è sicuramente il caso di Carlo Revelli, l'autore che risulta in assoluto il più prolifico, la cui produzione risulta più alta della media perché è stata oggetto di un *editathon* promosso dal GWMA in sua memoria il 12 aprile 2021<sup>53</sup> e di uno specifico progetto di caricamento dei dati. L'alto numero degli articoli di Giovanni Solimine si spiega, almeno in parte, con la pluridecennale direzione e collaborazione con *Biblioteche oggi*, rivista sulla quale risultano pubblicati il 65% (132 su 203) degli articoli censiti<sup>54</sup> e con la direzione decennale del *Bollettino AIB*.

In merito agli autori, sarebbe interessante compiere valutazioni quantitative anche di altra natura, quali l'occupazione, la distribuzione geografica, le fasce di età più ricorrenti o il loro sesso, le affiliazioni istituzionali. Per fare due esempi, si può osservare che, tra le 264 occupazioni degli autori che sono state censite, la prima e di gran lunga più diffusa è quella di bibliotecario (5.378), seguita da docente universitario (1.387), biblioteconomista (1.175), archivista (648), bibliografo (467), storico del libro (440) e paleografo (405)<sup>55</sup>; una mappa basata su 836 autori<sup>56</sup> – cioè quelli per i quali è disponibile il dato del luogo di nascita – mostra che la distribuzione geografica è su base mondiale per gli autori nati prima del 1960, e che si riduce progressivamente all'Europa e all'Italia con la crescita dell'anno di nascita.

Analisi simili a queste consentirebbero di delineare un quadro preciso sugli autori delle riviste italiane di biblioteconomia, per esempio, sulle posizioni professionali e il contributo dell'accademia e della professione allo sviluppo della biblioteconomia in Italia, e anche di gettare luce diversa e nuova sulle riviste italiane e il loro posizionamento in Italia e in ambito internazionale.

Tuttavia, queste analisi allo stato attuale non sono ancora sostenute da una sufficiente quantità di dati (per esempio, il 48% degli autori non ha il dato dell'occupazione e l'84% non ha il dato del luogo di nascita). Possono essere indicative, ma non essere considerate affidabili. Inoltre, la quantità dei dati che caratterizza gli item degli autori è spesso molto diversa: si passa infatti da item di autori associati a identificativi internazionali e con informazioni di qualità a item provvisti del solo nome e cognome<sup>57</sup>.

**53** <[https://www.wikidata.org/wiki/Wikidata:Gruppo\\_Wikidata\\_per\\_Musei,\\_Archivi\\_e\\_Biblioteche/Bibliografia\\_Carlo\\_Revelli](https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Bibliografia_Carlo_Revelli)>. Da notare che gli articoli su rivista risultano 350, ma la bibliografia di Carlo Revelli conta 437 riferimenti, perché include tutte le altre forme di pubblicazione (per esempio monografie, capitoli o parti di libro ecc.).

**54** La query è <<https://w.wiki/6iTe>>; una visualizzazione analitica e più chiara di questi dati si può ottenere all'indirizzo <<https://scholia.toolforge.org/author/Q3768160>>. Si può confrontare inoltre questa visualizzazione analitica con l'analisi bibliometrica offerta in Ernest Abadal, Gerard Vidal, *Las publicaciones de Giovanni Solimine: su presencia y su impacto en internet*, «AIB studi», 61 (2022), n. 3, p. 533-545, DOI: 10.2426/aibstudi-13335.

**55** Per un elenco delle prime occupazioni associate ad almeno 50 autori delle riviste italiane di biblioteconomia censite e inserite si veda <<https://w.wiki/6obB>>.

**56** Per una mappa dei luoghi di nascita degli autori di riviste italiane di biblioteconomia colorate secondo la fascia d'età si veda <<https://w.wiki/6oaR>>.

**57** Per esempio, gli autori totalmente privi di qualsiasi identificatore associato sono 1.309 (21 giugno 2023), <<https://w.wiki/4gYk>>.

Un livello troppo difforme, dunque, per poter proporre delle valutazioni che diano conto delle ‘macro-caratteristiche’ degli autori biblioteconomici italiani.

### Discussione

#### *Automazione e caricamento parziale dei dati*

Rispetto ai quesiti iniziali, la realizzazione del progetto ha mostrato che esistono metodi di caricamento degli articoli delle riviste più efficienti rispetto a quelli precedentemente utilizzati con *Bibliothecae.it*<sup>58</sup> e con *JLIS.it*<sup>59</sup>, se l’obiettivo è quello di gestire grandi quantità di dati.

Le tecniche di controllo della qualità dei dati, sotto forma di query preimpostate e lanciate periodicamente per l’individuazione di item con descrizioni ‘critiche’ (per esempio, articoli privi di legame con l’autore, autori privi di identificatori largamente diffusi come VIAF, ISNI o SBN ecc.) si sono dimostrate efficaci per l’estrazione dei dati che richiedevano un miglioramento. Le query utilizzate per il controllo hanno dimostrato anche che la qualità dev’essere spesso raggiunta con interventi di gestione delle entità, che richiedono tipicamente un intervento umano qualificato e specializzato: questa parte del processo non può essere demandata alle macchine e dev’essere svolta da figure professionali come catalogatori o creatori di metadati.

A causa delle dimensioni del progetto, non è ancora stato completato il controllo di qualità sugli item degli autori; in prospettiva, comunque, non sarà possibile mai portare allo stesso livello qualitativo tutti gli item, per la natura stessa dei dati esterni a Wikidata. Per esempio, molte riviste sono spogliate in SBN, mentre alcune, come *JLIS.it* e *Bibliothecae.it*<sup>60</sup>, non lo sono; questo implica che non sono state create registrazioni di autorità, neppure minime, per gli autori di queste ultime due riviste. Quindi, se quegli autori avessero pubblicato in Italia solo in quelle riviste, non risulterebbero censiti in SBN e perciò non avrebbero un identificatore SBN da associare in Wikidata e i loro item rimarrebbero necessariamente di qualità semantica inferiore rispetto agli altri.

A causa dell’incompletezza della fase di descrizione degli articoli (soprattutto in relazione all’inserimento delle parole chiave) e degli autori, il progetto non ha potuto includere per il momento un’analisi dei dati citazionali con dettaglio e ampiezza maggiori di quella offerta nel paragrafo precedente (numero di articoli per rivista, numero di autori per rivista ecc.).

Infine, il progetto ha dimostrato che le analisi quantitative sugli articoli e sugli autori sono possibili e saranno significative e utili per gettare nuova luce sugli studi biblioteconomici italiani quando i dati caricati in Wikidata raggiugeranno una soglia adeguata.

#### *La sostenibilità del progetto*

Il progetto di caricamento dei dati si proponeva di creare e rendere disponibili i dati degli articoli pubblicati su ciascuna rivista da quando erano disponibili fino all’epoca del progetto stesso. Idealmente, da parte del gruppo di lavoro, sarebbero stati caricati i dati relativi agli articoli pubblicati fino al 31 dicembre 2021.

**58** Cfr. A. Boccone, R. Rivelli, *I metadati bibliografici in Wikidata* cit.

**59** Cfr. C. Bianchini, *Wikidata for JLIS.it* cit.

**60** Si confronti, per esempio, la presenza della sezione ‘Comprende’ nel record di *AIB studi* (<<http://id.sbn.it/bid/CF0794721>>) o di *Nuovi annali della Scuola speciale per archivisti e bibliotecari* (<<http://id.sbn.it/bid/RAV0107645>>) e la sua assenza in quelli di *JLIS.it* (<<http://id.sbn.it/bid/UF0577978>>) e di *Bibliothecae.it* (<<http://id.sbn.it/bid/BVE0846142>>). Tale controllo è stato effettuato il 14 luglio 2023.

Questo obiettivo apriva il problema di stabilire se e come il progetto di valorizzazione si sarebbe potuto mantenere oltre quella data, in relazione a due nuclei principali di lavoro: il lavoro sugli articoli e quello sugli autori.

L'unica soluzione che sembrava percorribile, pur se con esiti incerti, era quella del coinvolgimento delle redazioni delle riviste di biblioteconomia; ognuna di esse avrebbe dovuto farsi carico dell'aggiornamento dei dati dal 2021 in poi: l'idea sembrava concretizzabile prima di tutto perché proprio in quel contesto era nata la prima esperienza di quel genere<sup>61</sup>.

Sono state quindi contattate le direzioni e le redazioni di *AIB studi*, *Biblioteche oggi trends*, *Biblioteche oggi*, *Bibliothecae.it*, *JLIS.it* e *Digitalia*, per presentare una proposta di collaborazione in questo senso: il nostro gruppo di lavoro avrebbe presentato il lavoro retrospettivo svolto e avrebbe dato la disponibilità a formare i collaboratori delle redazioni; queste poi si sarebbero assunte il compito di provvedere al caricamento sistematico delle pubblicazioni uscite a partire dal gennaio 2022 e all'arricchimento retrospettivo dei dati relativi alle pubblicazioni e ai loro autori.

Il corso, che aveva come contenuto l'estrazione dei set di metadati degli articoli tramite OJS, il trattamento con OpenRefine e il caricamento su Wikidata, si è svolto in tre giornate tra febbraio e marzo 2022, ed è stato suddiviso in un primo incontro introduttivo su Wikidata (dedicato a chi non aveva conoscenze su Wikidata), un secondo dedicato a illustrare le caratteristiche e il funzionamento di OAI-PMH e OpenRefine, con esempi di applicazione del GREL e un terzo incontro dedicato alla riconciliazione, alla creazione dello schema per Wikidata e al caricamento dei dati.

Il corso ha dato l'esito sperato per due redazioni: quella di *AIB studi* e quella di *JLIS.it*, che ora caricano autonomamente i dati di ogni nuovo fascicolo in uscita.

Riguardo al tema dell'arricchimento degli item degli autori, sono stati messi in atto due tentativi: il primo è stato coinvolgere la comunità dei bibliotecari italiani, contando sul fatto che molti di essi erano anche autori di pubblicazioni comparse sulle riviste censite. Questo approccio richiedeva di contattare gli autori e di metterli in condizione di fornire i dati anche nel caso – più probabile che no – in cui non fossero in grado di utilizzare Wikidata in modo autonomo. Si è pensato perciò di diffondere una richiesta di collaborazione tramite AIB-CUR, di predisporre una query che fornisse l'elenco degli autori sui quali c'erano informazioni scarse e un modulo Google attraverso il quale si potevano inviare le informazioni possedute. La mail è stata inviata il 15 gennaio 2022, ma l'esito è stato totalmente fallimentare, perché un solo bibliotecario ha fornito i dati per un solo item.

Inoltre, sempre riguardo all'arricchimento dei dati relativi agli autori, si è pensato che anche le redazioni delle riviste avrebbero potuto svolgere un ruolo di importanza fondamentale nel raccogliere dati autorevoli e di qualità direttamente dagli autori che sottomettono gli articoli da pubblicare. Nell'ambito del corso, si è cercato di sensibilizzare le persone coinvolte, ma questo approccio richiede la riorganizzazione del flusso di lavoro delle redazioni e l'inserimento di un passaggio di raccolta (e magari pubblicazione) di brevi profili biografici degli autori.

#### *Lo sviluppo futuro dell'analisi citazionale e bibliometrica*

Tra gli obiettivi iniziali del progetto era prevista una terza fase relativa all'analisi citazionale e bibliometrica sulla letteratura biblioteconomica nazionale, grazie al caricamento in Wikidata delle citazioni presenti in ogni articolo. Questi dati avrebbero

<sup>61</sup> Cfr. A. Boccone, R. Rivelli, *I metadati bibliografici in Wikidata cit.*

infatti consentito un'analisi approfondita delle relazioni tra le riviste e gli autori ma soprattutto di portare alla luce quella rete di connessioni di temi, di problemi, di approcci e il loro variare diacronico, tutti elementi che caratterizzano una disciplina e la sua professione.

Purtroppo, non è stato possibile procedere nemmeno all'analisi di un singolo caso per l'impossibilità di trattare le citazioni bibliografiche utilizzando strumenti e procedure automatici o semi-automatici. Le citazioni bibliografiche sono infatti gestite nella pratica consuetudinaria delle riviste come blocchi unitari di testo, inserite nel file PDF o HTML dell'articolo: quasi mai l'autore è vincolato a presentare la sua bibliografia in un formato strutturato (RIS, EndNote, RefWorks ecc.) e gli unici cenni in tal senso sono formulati dalle riviste come un invito<sup>62</sup>.

Alla luce di queste considerazioni, l'unica modalità di inserimento di tali citazioni bibliografiche in Wikidata rimane a oggi quella manuale, che ovviamente risulta inapplicabile a una mole di dati così elevata come quella gestita all'interno del progetto, con un dispendio di tempo ed energie considerevole e in maniera del tutto insostenibile per il futuro.

Una strada preferibile è certamente quella di sensibilizzare gli autori verso l'uso di *reference manager software* come Zotero o simili, e dunque vincolare l'elaborazione delle bibliografie a formati ben strutturati e potenzialmente elaborabili con processi automatici; tale strategia, d'altra parte, implicherebbe una presa di posizione netta, nonché un orientamento comune e condiviso da parte delle riviste del settore.

Altra soluzione, più impegnativa per il carico di lavoro che ne deriverebbe, potrebbe consistere nella creazione di tali bibliografie strutturate da parte delle stesse redazioni su ogni singolo articolo, prevedendo per esempio delle modalità condivise tra gruppi come quella offerta da Zotero<sup>63</sup>.

## Conclusioni

Seppur dunque in maniera incompleta rispetto alle intenzioni iniziali, il progetto ha portato a risultati sufficienti a impostare una prima valutazione, quantitativa e in parte qualitativa, della produzione scientifica biblioteconomica italiana, raggiungendo inoltre l'importante traguardo di sensibilizzare a questo proposito le redazioni di alcune tra le maggiori riviste del settore.

Tuttavia, per il completamento del progetto sarà necessario proseguire con diverse azioni:

1. completare il caricamento delle riviste ancora *in progress* (per esempio, *La Bibliofilia*) e l'inserimento di riviste finora non coinvolte nel progetto, con la possibilità di esplorare nuove modalità di reperimento dei dati laddove non siano liberamente accessibili sul web;
2. coinvolgere e responsabilizzare maggiormente le riviste, così che il lavoro di caricamento rimanga costante nel tempo;
3. impegnarsi in un costante incremento qualitativo degli item degli autori attraverso, per esempio, l'organizzazione di *editathon* dedicati;
4. proporre concretamente alle riviste del settore strategie comuni per la definizione

<sup>62</sup> Per esempio: «JLIS.it recommends its authors to use a reference manager software (e.g. Zotero and the like) connecting it to a Word processor. In this case, the file in .bib or .ris format containing the bibliographic references used in the article should be uploaded during submission», cfr. <<https://www.jlis.it/index.php/jlis/about/submissions>>.

<sup>63</sup> Cfr. <<https://www.zotero.org/support/groups>>.

di processi condivisi su specifiche tematiche (uso di *reference manager software* per la gestione dei dati bibliografici citazionali, definizione di un vocabolario condiviso per le parole chiave<sup>64</sup> ecc.);

5. sviluppare un progetto per il caricamento efficace ed efficiente dei dati bibliografici citazionali (per esempio, con l'aiuto di strumenti come Cita e Zotero);

6. coinvolgere nel progetto un maggior numero di studiosi e bibliotecari, ovvero professionisti di un settore in cui la sensibilità verso questi temi dovrebbe essere massima.

A questi obiettivi primari si aggiunge infine il più ambizioso proposito di aver in qualche maniera ispirato iniziative analoghe che, travalicando i limiti del contesto bibliotecario, nascano e si diffondano anche in ambiti culturali affini quali musei e archivi, enti e progetti culturali di più ampio respiro.

Articolo proposto il 22 giugno 2023 e accettato il 19 luglio 2023.

---

## ABSTRACT

AIB studi, 63 n. 2 (maggio/agosto 2023), p. 313-335. DOI 10.2426/aibstudi-13893  
ISSN: 2280-9112, E-ISSN: 2239-6152 - Copyright © 2023 Carlo Bianchini,  
Andrea Marchitelli, Alessandra Moi

---

CARLO BIANCHINI, Università di Pavia, Dipartimento di musicologia e beni culturali, Pavia, e-mail: carlo.bianchini@unipv.it.

ANDREA MARCHITELLI, EBSCO Information Services, Roma, e-mail: amarchitelli@ebSCO.com.

ALESSANDRA MOI, Università degli studi di Milano-Bicocca, Milano, e-mail: alessandra.moi@unimib.it.

### **Metodi e strumenti di un progetto di valorizzazione delle riviste italiane di biblioteconomia in Wikidata**

Gli autori di questo studio intendono condurre in futuro un'analisi bibliometrica incentrata sui recenti progressi della biblioteconomia e degli studi bibliografici italiani. Per raggiungere questo obiettivo, Wikidata è stato identificato come un potente strumento per la generazione, l'elaborazione e lo scrutinio dei dati, ma manca di informazioni riguardanti le riviste bibliotecarie italiane degli ultimi decenni. In questo caso, l'ingente volume di dati da caricare su Wikidata richiede notevoli risorse umane. Di conseguenza, questo articolo studia i metodi per snellire e gestire in modo efficiente il processo di caricamento dei dati. Inoltre, l'articolo analizza i metodi, gli strumenti e le tecniche esistenti per un'accurata identificazione degli autori all'interno del set di dati. In aggiunta, esso, esamina il potenziale per stabilire sinergie che promuovano la continuità del progetto e facilitino i progressi futuri. Oltre a ciò, la ricerca valuta le procedure volte a migliorare la qualità dei dati citazionali. Lo studio analizza anche la fattibilità dell'analisi dei dati bibliografici su un campione caricato su Wikidata rispetto ad altri strumenti bibliometrici, con particolare attenzione alla produzione in lingua italiana.

Facendo luce su questi aspetti fondamentali, l'articolo fornisce indicazioni preziose per la conduzione di ricerche bibliometriche nel campo della biblioteconomia e degli studi bibliografici italiani. I risultati non solo contribuiscono all'avanzamento delle conoscenze in questo campo, ma aprono anche la strada a una ricerca bibliometrica più efficiente ed efficace.

**64** In merito al processo di assegnazione delle parole chiave per l'indicizzazione semantica della letteratura scientifica e ai problemi di recupero dell'informazione che ne seguono si permetta di rinviare all'intervento *Author tagging: potenzialità e limiti delle parole chiave nella produzione scientifica professionale* di Alessandra Moi, Andrea Marchitelli e Carlo Bianchini al convegno internazionale "Look beyond. Indicizzazione per soggetto delle risorse non librarie" (Roma, 6 febbraio 2023) organizzato dal Gruppo di studio AIB-CILW e da ISKO Italia.t15

**Methods and tools of a project for the enhancement of Italian LIS journals in Wikidata**

The authors of this study would aim at conducting a bibliometric analysis focusing on recent advancements in Italian library science and bibliographic studies. To achieve this objective, Wikidata was identified as a powerful tool for generating, processing, and scrutinizing data, but it lacks information concerning Italian library journals from the latest decades. In this case, the substantial volume of data to be uploaded onto Wikidata would necessitate considerable human resources. Consequently, this article looks into methods to streamline and manage the data uploading process efficiently. Moreover, the article investigates existing methods, tools, and techniques for accurate author identification within the dataset. Furthermore, it examines the potential for establishing synergies that promote project continuity and facilitate future advancements. Additionally, the research assesses procedures aimed at enhancing citation data quality. The study investigates also the feasibility of conducting bibliographic data analysis on a sample uploaded to Wikidata compared to other bibliometric tools, with a particular focus on Italian-language production.

By shedding light on these vital aspects, this article provides valuable insights into conducting bibliometric research within the field of Italian library science and bibliographic studies. The findings not only contribute to the advancement of knowledge in this domain, but they pave the way for more efficient and effective bibliometric research.