

Wikidata for JLIS.it. A new step forward mapping Italian library and information science journals

Carlo Bianchini^(a)

a) Università degli Studi di Pavia, <http://orcid.org/0000-0002-6635-6371>

Contact: Carlo Bianchini, carlo.bianchini@unipv.it

Received: 3 September 2020; Accepted: 14 September 2020; First Published: 15 January 2021

ABSTRACT

The paper describes the process of publication of the metadata of the articles published on *JLIS.it* in the period 2010-2019 as Linked Open Data on Wikidata. The aim of the publication was to improve the number of Italian Library and Information Science papers on Wikidata, after the upload of the metadata of the papers published on the Italian journal *Bibliothecae.it* and to test new automatic tools for the publication process. OpenRefine was used to implement the process, as it allowed to upload metadata of different entities (journal, authors, names and family names of the authors, etc.) in sequential steps, and, in the last step, to upload the papers and all their relevant relationships. Metadata of two Italian LIS journals makes possible a few quantitative analyses on Italian scientific literature, by means of Scholia, a free Wikidata tool created for this purpose.

KEYWORDS

Jlis.it; Italian Library and Information Science Literature; Wikidata; Scholia; OpenRefine.

CITATION

Bianchini, C. "Wikidata for JLIS.it. A new step forward mapping Italian library and information science journals." *JLIS.it* 12, 1 (January 2021): 29–38. DOI: [10.4403/jlis.it-12680](https://doi.org/10.4403/jlis.it-12680).

Introduzione

Il 29 novembre 2018, Remo Rivelli annunciava alla lista dei bibliotecari wikimediani italiani che il gruppo Wiki dell'Università di Salerno aveva completato le prime due fasi del lavoro d'inserimento degli articoli della rivista *Bibliothecae.it* su Wikidata, ovvero la creazione degli item Wikidata per tutti gli articoli, gli autori e gli argomenti degli articoli. Il lavoro non era stato svolto in modo solo manuale, ma avvalendosi di “una serie di tool che hanno permesso di gestire gli inserimenti in maniera automatica, mentre le operazioni di rifinitura sono state necessariamente manuali (anche se penso che Open Refine possa essere d'aiuto)”.¹ Come scriveva con giusto orgoglio ed entusiasmo Rivelli, si trattava della “prima rivista italiana (e tra le primissime al mondo) di ambito bibliografico e biblioteconomico a essere completamente caricata in Wikidata!”.² Il risultato raggiunto dal gruppo Wiki dell'Università di Salerno – descritto in dettaglio in un articolo pubblicato su *Bibliothecae.it*³ – era un grande stimolo non solo per lo sviluppo della terza fase prevista (l'arricchimento degli item degli articoli con tutte le citazioni bibliografiche contenute all'interno degli stessi, per contribuire al progetto Wikicite), ma anche per l'avvio di lavori analoghi su altre riviste italiane del settore. Da quella esperienza positiva ed entusiasmante è nata l'idea di pubblicare su Wikidata i dati relativi alla rivista *JLIS.it. Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, sperimentando al tempo stesso una modalità tecnica diversa da quella utilizzata nel primo caso di studio per creare nuovo know-how a disposizione di tutti. L'obiettivo del nuovo studio non era quindi solo il caricamento dei dati di una seconda rivista italiana di biblioteconomia, ma verificare se e quali metodi alternativi fossero disponibili per quello scopo e studiare quali analisi diventassero possibili grazie ai dati a disposizione in questo campo d'interesse.

Materiali e metodi

JLIS.it si presentava come particolarmente adatta a costituire il secondo caso di studio per diversi motivi: prima di tutto perché essa è dello stesso ambito disciplinare di *Bibliothecae.it* e ciò rendeva più direttamente confrontabili esperienze e difficoltà; in secondo luogo l'eventuale disponibilità su Wikidata dei dati di una seconda rivista di settore avrebbe consentito di tentare alcune prime elaborazioni complessive e comparative dei dati; in terzo luogo era garantita la disponibilità dei dati perché – nell'ambito della discussione sulla lista dei wikimediani – da parte della redazione di *JLIS.it* era stata data immediata disponibilità a fornire tutti i dati accessibili; infine, la presenza su Wikidata

¹ Remo Rivelli, *Inserimento articoli rivista “Bibliothecae.it” in Wikidata*, mailing list bibliotecari@wikimedia.it, 29 novembre 2018.

² Ibid.

³ Alessandra Boccone e Remo Rivelli, “I metadati bibliografici in Wikidata: Wikicite e il case study di *Bibliothecae.it*,” *Bibliothecae.it* 8, n. 1 (12 giugno 2019): 227–48, <https://doi.org/10.6092/ISSN.2283-9364/9503>.

di articoli di *JLIS.it* già caricati in modo non sistematico⁴ consentiva la valutazione e l'analisi anche degli eventuali problemi connessi alla duplicazione dei dati.

È utile una breve rassegna di alcuni dei diversi possibili metodi di caricamento dei dati su Wikidata (oltre a quello manuale): un primo approccio, poi abbandonato, era stato tentato nel primo caso di studio. Si basa sulla disponibilità dei DOI (Digital Object Identifier) degli articoli e dell'uso di *SourceMD* (o Source Metadata Tool),⁵ un prototipo che sulla base dell'identificatore permanente di una pubblicazione scientifica ne elabora i dati e importa metadati bibliografici negli item di Wikidata: al momento, gli identificatori utilizzati sono ISBN-13, DOI, ORCID, PubMed id e PMC id. Tuttavia, la soluzione di SourceMD non era stata adottata per un problema legato alla disponibilità dei dati originati dai DOI: "La maggiore criticità evidenziata nell'utilizzo del tool è stata il mancato funzionamento con i DOI della rivista [*Bibliothecae.it*], in quanto non registrati in Crossref o in progetti similari".⁶ L'esperienza con i dati provenienti dai DOI degli articoli di *JLIS.it* in effetti è stata analoga: dopo avere ottenuto dalla redazione i DOI di tutti gli articoli pubblicati su *JLIS.it* grazie all'aiuto di Andrea Marchitelli, si è purtroppo constatato che i dati associati su CrossRef sono piuttosto lacunosi.

Per verificare facilmente la qualità e la riusabilità dei dati che si ottengono dal DOI di un articolo si possono utilizzare due strumenti; una prima soluzione è il sito di mEDRA,⁷ alla quale si può lanciare una richiesta con un link.⁸ Un altro strumento molto utile per comprendere la qualità dei dati disponibili presso le agenzie DOI è *doi2Bib*,⁹ perché mostra come i dati vengono suddivisi nei campi e quindi mostra più chiaramente la disponibilità di dati semantici o in forma di stringhe.¹⁰

Il deposito ufficiale per i dati legati al DOI di *JLIS.it* è mEDRA;¹¹ nei dati DOI registrati su mEDRA per gli articoli di *JLIS.it* quelli relativi al volume e al fascicolo sono presenti in modo irregolare e imprevedibile, mentre la pagina iniziale e la pagina finale dell'articolo non sono mai registrate (come nell'esempio di Figura 1).¹² In entrambi i casi di studio, quindi, i dati di partenza ottenibili dai DOI erano lacunosi.

Il secondo metodo, utilizzato da Boccone e Rivelli per la rivista *Bibliothecae.it*, prevede l'utilizzo di Zotero, associato a QuickStatements e a un affinamento manuale dei dati. Zotero¹³ è un programma

⁴ Con una query si potevano vedere su Wikidata alcuni articoli del primo volume, caricati da Claudio Forziati con alcune procedure batch e "anche un paio di articoli sparsi, qualcuno creato da me per sperimentare con Scholia, altri creati come item da altri utenti dopo la pubblicazione dell'ultimo fascicolo di *JLIS.it*". Claudio Forziati, comunicazione e-mail all'autore, 22 dicembre 2018.

⁵ <https://www.wikidata.org/wiki/Wikidata:SourceMD>; da agosto 2018 il tool è stato incluso in ORCIDator.

⁶ Boccone e Rivelli, 238.

⁷ mEdra (multilingual European Registration Agency of DOI) è l'agenzia europea di registrazione dei DOI, l'identificativo standard persistente di qualsiasi forma di proprietà intellettuale in una rete digitale (<https://www.medra.org/en/index.htm>).

⁸ Per esempio, il link <https://data.medra.org/10.4403/jlis.it-11>. Naturalmente, è possibile utilizzare il link come modello e sostituire nella stringa di esempio la parte che inizia con "10.4403" ecc. con il DOI di qualsiasi altro articolo di *JLIS.it*.

⁹ <https://www.doi2bib.org>.

¹⁰ Per esempio, <https://www.doi2bib.org/bib/10.4403/jlis.it-11>.

¹¹ Per conoscere quale agenzia registra i dati DOI di una particolare rivista, è possibile consultare il sito di CrossRef; un esempio d'interrogazione per la rivista *AIB studi* è <https://api.crossref.org/works/10.2426/aibstudi-11820/agency>; la documentazione relativa si può trovare a: <https://github.com/CrossRef/rest-api-doc>.

¹² È registrato invece talvolta il numero complessivo delle pagine dell'articolo; questo dato è poi stato effettivamente inserito in Wikidata, per migliorare l'analisi dei dati.

¹³ <https://www.zotero.org/>.

di reference manager software particolarmente integrato con Wikidata tramite ZotKat,¹⁴ un'estensione di Zotero per catalogare in senso ampio e che consente anche di tradurre i riferimenti bibliografici contenuti in Zotero in un formato adatto al caricamento su Wikidata tramite QuickStatements (QS). QuickStatements a sua volta è uno strumento prodotto dalla comunità Wikidata per creare e modificare gli item di Wikidata, grazie a un piccolo insieme di comandi testuali.¹⁵ Grazie a questi comandi è possibile aggiungere o modificare le *dichiarazioni* (triple del web semantico), etichette, descrizioni e alias e anche aggiungere dichiarazioni con qualificatori e fonti opzionali; per esempio, rispetto alla dichiarazione che una collezione contiene un certo tipo di risorse – libri, dischi, immagini, ecc. – è possibile ‘qualificarla’ sia aggiungendo il numero degli oggetti che indicando la fonte dei dati che si stanno inserendo (per esempio, i qualificatori *data d'inizio*, *numero* e *URL di riferimento* in alcune dichiarazioni dell'item Q64347227,¹⁶ che rappresenta il Fondo Elio Borsetto della Fondazione Levi di Venezia). Tramite ZotCat si producono comandi testuali che possono essere utilizzati da QuickStatements per modificare (o creare) i dati su Wikidata.

L'uso di Zotero e dell'estensione ZotKat per il caricamento degli articoli di *JLIS.it* presentava tuttavia un altro aspetto critico: gli item di Zotero sono stati inseriti in maniera semiautomatica, ma poi è stato necessario procedere “a un lavoro manuale di ‘ripulitura’ degli stessi: sono stati associati gli *item* autore creati nella prima fase del lavoro, sostituendo ad esempio la P2093 (proprietà Wikidata utilizzata quando l'autore della pubblicazione non ha un elemento Wikidata) inserita di *default* da Zotero, con la P50, che è la proprietà specifica per il creatore principale di un'opera. Sono stati poi aggiunti ulteriori metadati riguardanti, ad esempio, il numero di pagine (P1104)”.¹⁷ L'operazione di sostituzione della proprietà P2093 *stringa del nome dell'autore* con la proprietà P50 *autore* è assolutamente fondamentale e condivisibile, nell'ottica dei linked data e nella prospettiva Wikicite. Infatti, la proprietà P2093 serve per registrare come valore una *stringa* (non semantica), mentre la proprietà P50 *autore* consente di esprimere una *relazione qualificata* tra due *entità*. L'item articolo e l'item persona sono collegati da una relazione di tipo *autore* e ciò permette una rappresentazione dei dati molto più significativa. Tuttavia, questo procedimento implica l'inserimento manuale di ogni autore non già presente e il collegamento manuale tra item articolo e autore o autori relativi.

Rispetto all'obiettivo di pubblicazione su Wikidata di tutti i dati di *JLIS.it*, si trattava di un limite rilevante, perché questo progetto non poteva contare sulla collaborazione di un gruppo di lavoro (come peraltro capita spesso in biblioteca). Inoltre, la povertà dei dati iniziali ottenibili attraverso i DOI implicava già un importante sforzo manuale per il loro completamento. Perciò un prerequisito fondamentale era che le eventuali procedure manuali successive alla pubblicazione fossero minime, pena l'insuccesso del progetto stesso. Per queste ragioni, per pubblicare su Wikidata i dati di *JLIS.it*, si è deciso di utilizzare *OpenRefine*,¹⁸ cioè si è stabilito di sperimentare una soluzione tecnica alternativa a quella utilizzata per *Bibliothecae.it*, che partendo dai dati ricavabili dai DOI in Zotero, li

¹⁴ <https://github.com/UB-Mannheim/zotkat>.

¹⁵ <https://www.wikidata.org/wiki/Help:QuickStatements>.

¹⁶ <https://www.wikidata.org/wiki/Q64347227>.

¹⁷ Boccone e Rivelli, 239. Nella sperimentazione parziale di questo approccio tentata da Claudio Forziati per alcuni articoli di *JLIS.it* (vedi oltre), era stato necessario anche provvedere al collegamento dei singoli articoli con la rivista.

¹⁸ <https://openrefine.org>.

caricava su Wikidata tramite QuickStatements e si concludeva con un completamento e una revisione manuale dei dati.

Creazione e gestione dei dati con OpenRefine

OpenRefine (OR) è uno strumento gratuito e molto potente per la pulizia, la trasformazione e la conversione tra formati diversi dei dati originariamente in forma poco o non strutturata e la loro riconciliazione con i dati del web semantico. OR consente di visualizzare i dati in forma tabellare (come in un foglio di calcolo) ma offre strumenti per analizzare, scremare e modificare i dati in base alle necessità con funzioni già pronte o anche grazie a uno specifico linguaggio (GREL, General Refine Expression Language).¹⁹

Il GREL si è rivelato subito indispensabile per passare da un elenco tabellare di DOI a un foglio di dati contenenti, nelle diverse colonne, i metadati necessari per il caricamento degli articoli (titolo, autore o autori, parole chiave, volume, fascicolo, ecc.). Una funzione di OR (*fetch*) consente di lanciare richieste di dati (Link-based Content Type Requests) alle API dei siti web d'interesse e di scaricare i dati ottenuti nel proprio foglio dati (progetto) in OR. Per esempio, sulla base di una lista di nomi e cognomi di autori – e se possibili altri dati come date di nascita e di morte, per evitare omonimie – è possibile lanciare una richiesta di dati di certe entità-autore al VIAF e ottenere in risposta tutti o alcuni dati (per esempio, le opere, o le forme alternative del nome ecc.) associati a ciascun autore della lista. Per conoscere la 'sintassi' di una richiesta, ovvero come deve essere costruito il link che si usa per ottenere i dati, è necessario guardare la documentazione dello specifico sito d'interesse (nell'esempio, il VIAF).²⁰

Nel caso specifico, per ottenere una tabella di dati suddivisa in colonne contenenti i valori degli elementi del riferimento bibliografico a partire dal DOI degli articoli, si è consultata la documentazione di mEDRA e si è seguita la seguente procedura:

1. a partire dalla colonna contenente i DOI si è creata una seconda colonna che conteneva il link completo per la richiesta dati (cioè da 10.4403/jlis.it-11 a <https://data.medra.org/10.4403/jlis.it-11>);²¹
2. a partire dalla colonna ottenuta con gli URL di richiesta di dati a mEDRA, si sono scaricati i dati con la funzione di *fetch*;²²
3. i dati ottenuti, registrati tutti insieme nella nuova colonna, sono stati separati in colonne diverse (*parsing*);²³

¹⁹ <https://github.com/OpenRefine/OpenRefine/wiki/General-Refine-Expression-Language>.

²⁰ Per esempio, per istruzioni su come effettuare una richiesta di dati a DataCite, si deve consultare la pagina <https://support.datacite.org/docs/datacite-content-resolver>; per una richiesta di dati a CrossRef si deve consultare la pagina <https://github.com/CrossRef/rest-api-doc>.

²¹ Comando OR: Add column; comando GREL: "https://data.medra.org/" + value.

²² Comando OR: Add column by fetching URLs, specificando nel campo Accept "application/x-bibtex".

²³ Dal momento che i campi sarebbero stati separati dal carattere "," proveniente dal formato .csv (comma separated value), e che quel carattere poteva tuttavia essere presente anche nei titoli degli articoli, si è utilizzato come delimitatore dei diversi campi il carattere "=". Comando OR: Split.

4. i dati di ciascuna colonna ottenuta dal parsing sono stati ripuliti (per esempio, nella colonna titolo compariva il valore “title={Improvement of editorial quality of journals indexed in DOAJ ...}” anziché semplicemente “Improvement of editorial quality of journals indexed in DOAJ”).²⁴

Il risultato finale di questo primo processo di rielaborazione è stato un insieme di dati in forma tabellare, con righe che rappresentavano tutti i dati associati ai singoli articoli. La rappresentazione dei dati era ‘bidimensionale’, cioè non consentiva di esprimere le relazioni, per esempio tra un articolo e la rivista su cui è pubblicata (dato necessario per la proprietà P1433) o tra l’articolo e il suo autore (o i suoi autori in caso fossero più di uno, per la proprietà P50).

Riconciliazione dei dati con OpenRefine

La creazione di una relazione, in generale, e in Wikidata in particolare, necessita che le entità da collegare esistano già. Se un’entità non esiste, il dato viene registrato non in forma di relazione, ma in forma di attributo di tipo stringa: questo, come si è visto, è ciò che succede se si caricano i dati con Zotero, che crea una proprietà P2093 (stringa del nome dell’autore) invece che una relazione basata sulla proprietà P50. Di conseguenza, per costruire relazioni ben formate, al momento di caricare i dati degli articoli in Wikidata dovevano essere stati preventivamente creati, rielaborati e riconciliati in OpenRefine i dati relativi alla rivista *JLIS.it*, a tutti gli autori e a tutte le parole chiave (proprietà P921). La riconciliazione è una funzione inclusa in OR che prevede la comparazione tra uno o più dati presenti nella tabella locale con i dati presenti in Wikidata per trovare delle corrispondenze ed eventualmente stabilire un’identità (riconciliazione).

Per esempio, volendo riconciliare una serie di persone, si partirà da una colonna di OR che contiene il nome e il cognome delle persone e si indicherà nella funzione di tenere conto anche dei valori eventualmente presenti nella colonna contenente data di nascita e data di morte. I dati della tabella locale vengono poi confrontati con i dati presenti in Wikidata – per esempio, l’etichetta per il nome e cognome e le proprietà P569 (data di nascita) e P570 (data di morte) – e si ottiene un’indicazione della corrispondenza (match) tra i dati in locale e quelli in Wikidata.

Dopo avere dato il comando per avviare il processo, nella colonna relativa di OR si può trovare, sulla base della corrispondenza alta, media o bassa tra i dati locali e quelli ritrovati in Wikidata:

- un match accettato (con l’opzione “Scegli un nuovo match” se quello scelto non fosse corretto);
- più match proposti come probabili (in ordine di probabilità espressa in percentuale);
- nessun match e l’opzione “crea un nuovo item”; i match accettati automaticamente hanno un grado di affidabilità proporzionale alla quantità di dati utilizzati per la riconciliazione: più dati trovano corrispondenza, più è probabile la corrispondenza esatta (l’identità) tra i due item.

A questo punto, è necessario provvedere ad analizzare i risultati, accettando i match automatici validi, scegliendo l’opzione corretta tra quelli proposti come probabili, o confermando di volere creare in seguito gli item che non hanno avuto match. Questa fase di validazione del processo di riconciliazione

²⁴ Per correggere le stringhe si sono usati il comando OR: Edit cell, Transform e il comando GREL: `value.replace("{", "")`, oppure `value.replace("@article{", "")` ecc.

è la parte più chiaramente intellettuale del lavoro di authority control dell'intero processo; solo un bibliotecario, o un esperto di una certa disciplina, sa stabilire se il match proposto in modo automatico può essere confermato o meno. Si tratta d'instaurare una relazione d'identità, che ha profondamente a che fare con il significato – la semantica – dei dati che si stanno comparando.

La corrispondenza (match) che si ottiene con il processo di riconciliazione è una condizione essenziale per potere intervenire a qualsiasi titolo (creazione, modificazione, integrazione, ecc.) sui dati di Wikidata.

Se non si procede alla riconciliazione e alla validazione dei match, o non si procede correttamente, c'è il forte rischio di creare duplicati o di fare modifiche errate.²⁵ Perciò il processo di riconciliazione ha riguardato nell'ordine tutti gli autori (con i rispettivi nomi e cognomi, che sono entità autonome in Wikidata), tutte le parole chiave per la descrizione degli argomenti, la rivista e i singoli articoli (perché alcuni erano già presenti in Wikidata, anche se con dati parziali).

Pubblicazione dei dati con OpenRefine

Il passo successivo e finale consiste nel caricamento dei dati delle tabelle: per gli item degli articoli di JLIS.it da creare, viene creato l'item in Wikidata e viene corredato di tutti i dati e delle relazioni (rivista, autore/i, parole chiave) presenti nella tabella locale.

Per gli articoli già presenti e che hanno trovato corrispondenza durante la riconciliazione, vengono caricati solo i dati e le relazioni della tabella locale eventualmente mancanti in Wikidata. Infatti, con la procedura di riconciliazione – effettuata sul titolo dell'articolo – è stato possibile individuare quali articoli fossero già stati inseriti in Wikidata e quali no, e quindi 'completare' i dati mancanti, secondo il principio di modularità che contraddistingue il web semantico.

Per pubblicare i dati su Wikidata è necessario creare uno "schema" in OR: si tratta di una descrizione che serve per indicare a quale proprietà devono essere associati i dati di una determinata colonna della tabella di OR. Per esempio, tra i dati locali da pubblicare è registrato il volume e il fascicolo di ogni articolo; lo schema permette di associare le colonne "Volume" e "Fascicolo" predisposte in OR alle rispettive proprietà P478 (volume di un libro) e P433 (numero della pubblicazione). Lo schema contiene quindi tutti dati da caricare nella procedura che si intende avviare e i dati dipendono dalle proprietà degli item che si stanno registrando.

Visualizzare i dati

Se si pubblicano i dati relativi agli articoli di una rivista in forma Linked open data (LOD) diventa molto più facile fare analisi sulla rivista stessa, sui suoi autori, sul numero di articoli pubblicati per ciascun anno, sul numero di pagine medie di un articolo, oppure degli articoli di un certo anno, oppure degli articoli con una certa parola chiave e così via.

²⁵ Durante la sperimentazione, il rischio è diventato reale e il caricamento di dati non riconciliati su Wikidata ha portato a modifiche indesiderate di item già presenti. Fortunatamente, la comunità Wikidata è presente e attiva: Alessandra Boccone ha tempestivamente segnalato il problema e si è potuto provvedere al ripristino dei dati precedenti, tutto nella stessa giornata (22 febbraio 2019).

Se è vero che per queste analisi è in genere necessaria una minima conoscenza del linguaggio SPARQL,²⁶ è altrettanto vero che l'enorme interesse che Wikidata suscita nel mondo accademico e le immense possibilità che offre nell'ambito delle analisi citazionali unite al contesto cooperativo e collaborativo che caratterizza tutti i progetti wiki consente di avere a disposizione strumenti di alto livello, gratuiti e pronti all'uso. Un esempio tra i migliori è *Scholia*,²⁷ creato da Finn Årup Nielsen. Scholia è uno strumento per fare analisi bibliometriche quantitative in ambito scientifico utilizzando Wikidata come database. Lo scopo del sito è creare profili istantanei per ricercatori, organizzazioni, riviste, editori, singole opere scientifiche e argomenti di ricerca, sfruttando i dati presenti su Wikidata.²⁸

Le possibilità relative alle misurazioni quantitative dei risultati delle ricerche che si aprono con questo approccio sono enormi: a condizione che i dati siano presenti su Wikidata, per un determinato ricercatore è possibile avere, per esempio, elenco delle pubblicazioni, numero di pubblicazioni per anno (unico autore, primo autore, ultimo autore, ecc.), numero di pagine annue, grafico e dati del contesto di pubblicazione (riviste, convegni, ecc.), rappresentazione grafica dei coautori, grafico e dati relativi agli argomenti più trattati.²⁹ Oltre alle analisi e alle visualizzazioni proposte in forma standard, il vero valore aggiunto è la possibilità illimitata di estendere l'analisi dei dati tramite apposite query direttamente su Wikidata, per esempio per calcolare l'H-index di un autore.³⁰ Inoltre, l'analisi dei dati è aggiornabile in tempo reale, tramite il caricamento dei dati su Wikidata; per esempio, se un dato in una query è lacunoso ma disponibile, si può integrare inserendolo in Wikidata e 'vederlo' immediatamente dopo nella query aggiornata.

Analogamente, anche la produzione di un intero dipartimento universitario può essere visualizzata, sempre in termini quantitativi, attraverso l'elenco del personale (in ordine di quantità di produzione), un grafico dei coautori, argomenti più trattati dal personale interno e dai collaboratori, pubblicazioni più recenti, quantità di pagine prodotte per autore, citazioni recenti ottenute, articoli più citati del personale del dipartimento, premi e riconoscimenti, ecc.³¹

Scholia prevede la possibilità di analizzare i dati presenti su Wikidata per una molteplicità di profili: oltre che per gli autori, esistono profili pronti anche per le opere, le collane, i cataloghi, gli editori, gli stampatori; per le organizzazioni in generale e gli sponsor; per i luoghi e le nazioni; per eventi singoli o periodici; per progetti di ricerca; per argomento; e, infine, per rivista scientifica.

²⁶ <https://www.w3.org/TR/rdf-sparql-query/>.

²⁷ <https://scholia.toolforge.org/>; vedi anche: <https://www.wikidata.org/wiki/Wikidata:Scholia>.

²⁸ Finn Årup Nielsen, Daniel Mietchen, Egon Willighagen, "Scholia, Scientometrics and Wikidata," in *The Semantic Web: ESWC 2017 Satellite Events*, ed. Eva Blomqvist et al., v. 10577, Lecture Notes in Computer Science (Cham: Springer International Publishing, 2017), 237–59, https://doi.org/10.1007/978-3-319-70407-4_36. Una presentazione chiara e completa di Scholia è stata fatta da nell'incontro dell'11 agosto 2020 al meeting online del LD4 Wikidata Affinity Group: Daniel Mietchen e Lane Rasberry, "Presenting Scholia: A Scholarly Profiling Tool", https://docs.google.com/presentation/d/1jJbYSnYSDh36-LxzSpedFyWUzusZAjuBbP-y46ji-0w/edit#slide=id.g35f391192_00.

²⁹ Per esempio, <https://scholia.toolforge.org/author/Q21562060> è la pagina di Scholia sulla produzione di Dario Taraborelli.

³⁰ Nielsen, Mietchen, Willighagen, 242.

³¹ Per esempio, <https://scholia.toolforge.org/topic/Q131626> è la pagina di Scholia sulla produzione della Smithsonian Institution.

Visualizzare i dati di una rivista

Se l'interesse dello studioso si sposta sull'analisi della produzione di una rivista, Scholia mette a disposizione specifici grafici e tabelle creati al volo su:

1. il numero e l'elenco degli articoli pubblicati, in ordine cronologico inverso;
2. l'elenco degli argomenti più trattati (in base alle parole chiave fornite) e il link a tutti gli articoli che ne parlano;
3. l'elenco degli autori che hanno contribuito maggiormente alla rivista, con il relativo numero di articoli (in forma di link) e, se disponibile su Wikicommons, una loro immagine;
4. un grafico delle collaborazioni tra gli autori;
5. la possibilità di vedere gli articoli più citati e gli autori più citati tra gli autori;
6. le fonti più citate negli articoli comparsi sulla rivista;
7. in quali fonti vengono citati gli articoli della rivista presa in esame;
8. i dati relativi al genere, distinti nel conteggio degli autori uomini e donne e nel numero di articoli firmati da uomini e donne;
9. l'eventuale presenza di autori che hanno avuto riconoscimenti e premi scientifici.³²

È necessario sottolineare ancora una volta che la visualizzazione di un particolare insieme di dati dipende dal loro previo caricamento su Wikidata; per esempio, per le due riviste di biblioteconomia italiane sono stati caricati i dati degli articoli e dei relativi autori, ma non quelli delle eventuali citazioni dei loro articoli pubblicate in altre sedi. In questo caso, la risposta di Scholia è povera perché i dati in Wikidata sono ancora poveri. Tuttavia, la sola presenza degli articoli delle due riviste è un ottimo presupposto affinché questi dati possano essere creati da terzi (per esempio, da altri contributori che vogliono pubblicare dati di altre riviste di settore, italiane o estere).

Un'ultima considerazione sulla visualizzazione dei dati delle riviste è relativa alla possibilità prevista da Scholia di effettuare comparazioni tra item di Wikidata, per esempio, tra autori, luoghi, editori ecc. Le riviste non fanno eccezione; è possibile impostare la query per *Bibliothecae.it* e *JLIS.it* e confrontare, a titolo puramente esemplificativo, i dati ottenuti.³³

Nella query preimpostata il confronto è limitato ad alcuni aspetti quantitativi: numero di articoli pubblicati in assoluto e grafico con il numero degli articoli pubblicati per anno; citazioni per anno e citazioni in rapporto alle opere. Mancano quindi alcuni altri aspetti direttamente comparabili, come il numero degli autori, divisi per genere, o il numero di articoli per genere; dal punto di vista disciplinare sarebbe stata utile anche una comparazione relativa agli argomenti trattati.

Conclusioni

La sperimentazione relativa al processo di pubblicazione dei dati relativi agli articoli di *JLIS.it* per mezzo di OpenRefine ha mostrato che questo approccio è pratico ed efficace e consente di ridurre al minimo l'intervento manuale sui dati di Wikidata. Le difficoltà riscontrate in partenza con i metadati

³² Un esempio per la rivista *Antiquity* si può vedere con il link <https://scholia.toolforge.org/venue/Q4775205>. Per la visualizzazione dei dati di *JLIS.it* si può usare il link <https://scholia.toolforge.org/venue/Q3805294>; per visualizzare i dati relativi a *Bibliothecae.it* si può usare il link <https://scholia.toolforge.org/venue/Q50811189>.

³³ Per il confronto si può utilizzare il link: <https://scholia.toolforge.org/venues/Q3805294,Q50811189>.

ricavati dai DOI hanno evidenziato che “i linked data sono utili quanto i metadati da cui dipendono, e metadati di scarsa qualità finiscono per ricreare i problemi che molti bibliotecari sperano di risolvere tramite i linked data”.³⁴

La pubblicazione dei dati degli articoli di *JLIS.it* ha consentito di analizzare la rivista nel suo complesso, tramite l’uso di Scholia. L’analisi degli articoli e degli argomenti individuati tramite le parole chiave, consente di trarre indicazioni utili relative all’orientamento della rivista, agli argomenti più trattati, alla quantità dei contributi pubblicati e – in prospettiva, quando miglioreranno i metadati – relativi alle citazioni e all’impatto della rivista. Il confronto diretto tra le due riviste è, al momento, poco significativo sia rispetto alle singole riviste che rispetto al quadro generale della disciplina in Italia. In questo senso, i dati sono ancora troppo pochi, ma è verosimile che con ulteriori caricamenti diventerà più chiaro il collocamento disciplinare e di pubblico delle singole riviste, anche su base comparativa e d’integrazione delle specificità tra le riviste stesse.

È infine evidente che l’analisi quantitativa non è mai sufficiente per trarre conclusioni definitive; tuttavia ci sono almeno due ragioni per continuare a svilupparla: da un lato risulta molto utile per supportare l’analisi qualitativa; dall’altro, i metadati – di una singola rivista e ancor più di tutte le riviste di un settore disciplinare – stanno diventando così tanti che è ormai sempre più difficile realizzare un’analisi qualitativa che non sia per campione e quindi diventa indispensabile supportare questo lavoro con strumenti di analisi di grandi quantità di dati.

³⁴ Kyle Banerjee, “The Linked Data Myth,” *Library Journal*, consultato 24 agosto 2020, <https://bit.ly/37ECnWD>.