

Alessandra Boccone - Remo Rivelli*

*I metadati bibliografici in Wikidata:
Wikicite e il case study di «Bibliothecae.it»*

Discendum est enim ut vivamus,
non vivendum ut discamus.

Conrad Gesner

1. Citazioni aperte: sperimentazioni e progetti

Le citazioni bibliografiche mostrano l'orizzonte culturale in cui si è sviluppata un'opera creativa, tracciano la linea delle idee che hanno portato a determinate scoperte, collegano vari ambiti di attività e studi e, infine, formano una base solida per lo sviluppo di nuovi saperi e competenze. Negli ultimi anni esse hanno acquisito uno straordinario valore aggiunto, perché considerate una fonte primaria necessaria a valutare l'impatto e la qualità della ricerca scientifica e,

* L'articolo è stato ideato e redatto da entrambi gli autori. Nello specifico: Alessandra Boccone ha curato i paragrafi 2, 3, 5, mentre Remo Rivelli è responsabile dei paragrafi 1, 4, 5. Gli autori desiderano ringraziare la prof. Fiammetta Sabba e la dott.ssa Maria Rosaria Califano per l'incoraggiamento e il sostegno all'iniziativa, e la comunità dei wikibibliotecari per il costruttivo confronto sul tema. Tutte le ultime consultazioni dei siti sono state effettuate nel mese di aprile 2019.

di conseguenza, anche a stabilire il finanziamento e lo sviluppo dei migliori progetti.² In particolare, come sostiene Silvio Peroni,

reti di citazioni scientifiche possono essere caratterizzate topologicamente (definendo il grafo di connessione tra gli articoli citanti e citati nel tempo), in termini sociologici (per l'identificazione di cattivi processi di condotta della ricerca scientifica e di accesso elitario alla scienza), o sottostare a logiche prettamente quantitative (creando metriche basate su citazioni per valutare l'impatto scientifico di un'idea e/o di una persona) ed economiche (intese come moneta con la quale un ricercatore provvede alla sua sostentazione accademica).³

Tuttavia, gli indici citazionali più importanti⁴ sono consultabili solo a fronte di costi molto elevati, dunque rimangono al di fuori delle possibilità di milioni di possibili utenti.

È necessario, dunque, facilitare l'esplorazione delle innumerevoli connessioni esistenti tra diversi campi del sapere attraverso un accesso illimitato ai dati bibliografici e alle citazioni, in forme leggibili sia dagli umani che dalle macchine.

A tale scopo nasce nell'aprile del 2017 il progetto *Initiative for Open Citations* (I4OC)⁵ per iniziativa di *OpenCitations*⁶, Wikimedia Foundation, PLOS, eLife, DataCite, The Centre for Culture and Technology at Curtin University.⁷

L'obiettivo principale è dunque rendere liberi da copyright e disponibili a chiunque in maniera strutturata, separabile e aperta i

² Taraborelli 2018.

³ Peroni 2018.

⁴ Fra i più utilizzati, *Web of Science* di Clarivate Analytics e *Scopus* di Elsevier.

⁵ Cfr. *i4oc.org*, <<https://i4oc.org/>>.

⁶ *OpenCitations* è un progetto congiunto dell'Università di Oxford e del Dipartimento di Informatica - Scienza e Ingegneria dell'Università di Bologna per la pubblicazione dei dati bibliografici e delle citazioni aperte mediante l'uso di tecnologie Semantic Web (Linked Data). Fornisce il modello di dati *OpenCitations* e le ontologie SPAR (Semantic Publishing and Reference) per la codifica di dati bibliografici e citazionali accademici in RDF.

⁷ Taraborelli - Dugan 2017.

dati relativi alle citazioni di articoli scientifici e prodotti accademici pubblicati in tutto il mondo.⁸

I metadati che rappresentano ciascuna pubblicazione e ogni istanza di citazione sono espressi in formati comuni leggibili dalle macchine; sono aperti, affinché possano essere liberamente accessibili e riutilizzabili, e sono separabili, nel senso che le istanze di citazione possono essere consultate e analizzate senza la necessità di accedere ai prodotti bibliografici di origine.

Le citazioni sono immagazzinate in Crossref e rese disponibili attraverso le API; sono altresì conservate nell'*OpenCitations Corpus*, un database che raccoglie dati citazionali da Crossref e altre fonti. Nel momento in cui entrano a far parte dell'iniziativa, i dati devono essere considerati in pubblico dominio, e quindi usati con licenza CC0.⁹

Fino ad aprile del 2017 la lista dei riferimenti bibliografici degli articoli non era liberamente accessibile senza la esplicita concessione dall'editore: alla fine del 2016, soltanto l'1% degli articoli che presentavano liste di riferimenti bibliografici avevano effettivamente rilasciato questi dati in modo totalmente aperto.¹⁰ Ad oggi è stata superata la soglia dei venti milioni di articoli con elenchi di citazioni liberamente accessibili, con una percentuale di copertura vicina al 51% del totale. Restano ancora irrisolti allo stato attuale alcuni aspetti problematici: una parte degli editori¹¹ tra quelli che contano il più alto numero di dati citazionali depositati in Crossref non hanno accordato l'accesso a questi dati tramite la piattaforma.¹²

La citazione bibliografica è anche la base dei principali progetti Wikimedia, da un lato perché costituisce un elemento fondamentale per la verificabilità e la correttezza delle informazioni dichiarate, dall'altro perché i metadati che costituiscono la citazione, cioè il

⁸ Peroni 2018.

⁹ Heibi - Peroni - Shotton 2019, p. 3.

¹⁰ Peroni - Shotton - Vitali 2017, p. 185.

¹¹ Fra cui *Elsevier*, *IEEE*, *Wolters Kluwer Health*, *IOP Publishing*, *ACS* detentori di una grossa fetta del rimanente 49% dei dati citazionali non pubblici in Crossref.

¹² Cfr. <<https://www.crossref.org/reports/members-with-closed-references/>>.

legame fra un articolo e la sua fonte, sono «dati fattuali» non coperti dal diritto d'autore, quindi vicini alla concezione 'open' che caratterizza tali progetti.

Fra le proposte formulate dalla comunità attiva nei progetti Wikimedia, la più importante in tal senso è Wikicite.

2. WikiCite: le informazioni bibliografiche in Wikidata

WikiCite è un'iniziativa che mira a costruire un vasto database di citazioni bibliografiche di qualità in *linked open data*, utilizzando l'infrastruttura di Wikidata.

2.1 La struttura di Wikidata.

Nato nel 2012 con lo scopo principale di raccogliere e strutturare i dati fondamentali delle voci e delle pagine degli oltre ottocento progetti Wikimedia, nell'arco di sette anni è diventato uno dei progetti più attivi.

Wikidata, dunque, si può definire una base di conoscenza secondaria, libera, collaborativa e multilingue, formata da dati interpretabili sia dall'uomo che dalle macchine; essa raccoglie dati contenuti in fonti primarie (secondo il criterio di verificabilità comune a tutti i progetti Wikimedia), pubblicati con licenza Creative Commons Zero (CC0), cioè in pubblico dominio, in modo da garantirne il più ampio riuso possibile, anche tramite l'esportazione in diversi formati, tra cui JavaScript Object Notation (JSON) e Resource Description Framework (RDF). Ciò significa che chiunque può modificare ed utilizzare i dati, anche per uso commerciale, con l'unico obbligo di citare la fonte da cui sono stati prelevati. I dati sono immessi e controllati in maniera collaborativa dalla comunità dei contributori, che determinano anche lo sviluppo del progetto, prendendo decisioni anche riguardo ai data models da adottare, modificare o creare. L'utilizzo, la modifica e il riuso dei dati è assicurato in trecento lingue,

oltre che in formati leggibili dalle macchine.¹³

Per ogni pagina di Wikipedia o di qualsiasi altro progetto Wikimedia viene creato un elemento su Wikidata per la gestione di tutti i dati strutturati e dei collegamenti correlati fra tutte le versioni linguistiche; questi elementi sono chiamati *item* e sono identificati dalla lettera Q seguita da un numero progressivo.

Per archiviare i dati, oltre alle etichette di testo e ai collegamenti linguistici, Wikidata utilizza un semplice modello, dove gli elementi sono descritti attraverso coppie proprietà-valore. Per ogni valore è possibile aggiungere dei qualificatori, che forniscono informazioni di contesto, e dei riferimenti, che elencano le fonti a supporto dei valori inseriti.

Le proprietà sono corredate da etichette, alias e descrizioni. A differenza degli articoli, tuttavia, queste pagine non sono collegate a voci di Wikipedia e vengono identificate dalla lettera P seguita da un numero progressivo.

In sintesi possiamo affermare che il *data model* di Wikidata è pienamente interpretabile in ottica LOD: gli *item* contengono dichiarazioni in forma di triple, dove il soggetto è l'*item* stesso, il predicato è rappresentato da una proprietà e l'oggetto da un valore che può essere una stringa, una data, un identificatore esterno o un altro *item*.¹⁴

La nascita del progetto Wikidata ha quindi semplificato la creazione e pubblicazione di LOD, che tradizionalmente ha sempre richiesto competenze tecniche raffinate per la trasformazione dei dati relazionali e la creazione dei vocabolari e delle ontologie.¹⁵ Nel documento di discussione presentato al congresso IFLA del 2016, Bartholomei e colleghi hanno evidenziato come la capacità di Wikidata di disegnare dati aperti collegati, integrare *authorities* di dati e diverse ontologie e tassonomie abbia un enorme potenziale di utilizzo per i ricercatori di

¹³ Martinelli 2016, p. 76.

¹⁴ Anderson 2018, p. 307.

¹⁵ Allison - Scott 2018.

tutto il mondo.¹⁶

Le informazioni bibliografiche raccolte in Wikidata, dunque, possono riguardare, fra l'altro, monografie, articoli, *preprints*, autori, organizzazioni, riviste ed editori. Questi *item* (corrispondenti al 'soggetto' nel linguaggio del Web semantico) possono essere interconnessi attraverso le proprietà 'bibliografiche' di Wikidata (corrispondenti al 'predicato'), come ad esempio 'autore' (P50), 'pubblicato in' (P1433), 'editore' (P123), 'serie' (P179), 'tema principale' (P921), 'istruito a' (P69), 'datore di lavoro' (P108), 'parte di' (P361), 'sponsor' (P859) e diverse altre.¹⁷

Esistono, inoltre, numerose proprietà per il collegamento diretto a risorse esterne, ad esempio attraverso DOI, PMID, PMCID, arXiv, ORCID, Google Scholar, VIAF, CrossrefID, ZooBank e Twitter.¹⁸ Attraverso queste, grazie a una serie di tool¹⁹, è possibile creare e popolare gli elementi in maniera automatica o semiautomatica; grazie a ciò, negli ultimi mesi, il numero di item relativi agli articoli scientifici si è sviluppato in maniera esponenziale: più di centoquaranta milioni di citazioni solo nell'ultimo anno.²⁰

Questa enorme mole di metadati bibliografici, circa il 40% sul totale degli item presenti in Wikidata, costituisce il nucleo centrale

¹⁶ Bartholmei et al. 2016, p. 5.

¹⁷ Cfr. tutte le proprietà relative ai metadati bibliografici <https://www.wikidata.org/wiki/Template:Bibliographic_properties>.

¹⁸ «Linking is a small act of generosity that sends people away from your site to some other that you think shows the world in a way worth considering. [...] [Sources] that are not generous with linking [...] are a stopping point in the ecology of information. That's the operational definition of authority: The last place you visit when you're looking for an answer. If you are satisfied with the answer, you stop your pursuit of it. Take the links out and you think you look like more of an authority». D. Weinberger (2012), Linking is a public good, <<http://www.hyperorg.com/blogger/2012/02/26/2b2k-linking-is-a-public-good/>>.

¹⁹ I principali tool utili a tale scopo sono segnalati nel *WikiProject:Source Metadata* <https://www.wikidata.org/wiki/Wikidata:WikiProject_Source_Metadata/it>.

²⁰ Il dato è basato sulle statistiche di Wikidata <<http://wikicite.org/statistics.html>>.

dell'attività di WikiCite.

2.2 L'evoluzione di Wikicite

La prima proposta di costituire un progetto che accorpasse ogni riferimento, citazione o fonte utilizzata nelle pagine Wiki risale al 2006;²¹ negli anni successivi proliferano le singole iniziative in tale direzione: la creazione di Open Metadata Handbook, la messa a punto di Wikibase e del WikiProject Source MetaData, la ricerca e la riformulazione di *data model* per i metadati degli articoli scientifici e delle monografie nei progetti Wiki, gli *hackaton* di citazioni bibliografiche in alcune importanti biblioteche.

L'idea di un progetto unico in cui far confluire tali manifestazioni e proposte viene ripresa nel 2011²², ma non si riesce a trovare una soluzione valida per evitare la frammentazione delle risorse: solo nel 2016 la discussione si sposta su Wikidata, la cui infrastruttura si rivela perfettamente compatibile all'attuazione del progetto.²³

Durante la prima conferenza internazionale dedicata, tenutasi nel 2016 a Berlino, viene discusso ed elaborato in maniera organica l'ambizioso progetto;²⁴ un gruppo formato da cinquantacinque partecipanti provenienti da varie organizzazioni – incluse università, biblioteche, case editrici e *open data stakeholders* – si sono incontrati per analizzare e progettare i *data models* e la tecnologia utile a migliorare la copertura, la qualità, la conformità agli standard delle citazioni e dei metadati che hanno origine nei progetti Wikimedia. In particolare, l'obiettivo principale consisteva nel definire una *roadmap* tecnica per

²¹ Wikicite (2006 proposal) <[https://meta.wikimedia.org/wiki/Wikicite_\(2006_proposal\)](https://meta.wikimedia.org/wiki/Wikicite_(2006_proposal))>.

²² Wikicite (pre-Wikidata) <[https://meta.wikimedia.org/wiki/Wikicite_\(pre-Wiki-data\)](https://meta.wikimedia.org/wiki/Wikicite_(pre-Wiki-data))>.

²³ Cfr. la timeline di Wikicite <<https://meta.wikimedia.org/wiki/WikiCite/Timeline>>.

²⁴ Supportata da Wikimedia Foundation e Wikimedia Deutschland e finanziata da Alfred P. Sloan Foundation, Gordon e Betty Moore Foundation e Crossref.

la costituzione di un *repository* di fonti, utilizzando la spina dorsale semantica di Wikipedia, cioè Wikidata. Nelle *convention* WikiCite successive, diventate oramai un appuntamento annuale molto partecipato, sono stati discussi i progressi, l'impatto, i bisogni della comunità e le sfide a livello tecnico. In particolare, durante l'ultima conferenza tenutasi a Berkeley nel 2018, si sono delineati tre possibili scenari di sviluppo. Nella prima ipotesi, l'iniziativa si predispone a costituire il database completo di tutte le fonti citate nei progetti Wikimedia. Progressi significativi sono stati già conseguiti verso tale obiettivo: i dati citazionali sono oggi strutturati attraverso la creazione di *item* per tutte le opere citate nei progetti Wikimedia, nonché con la progettazione di una serie di modelli di dati bibliografici in Wikidata e con la creazione di *bot* e strumenti necessari a supportare un flusso di lavoro così imponente. Esistono ancora grandi lacune: c'è la necessità di importare su vasta scala fonti diverse da quelle accademico-scientifiche (articoli di giornale, pagine web, TV/radio, ect.), ma il progetto sta già registrando anche in questo campo notevoli sviluppi.

Nella seconda ipotesi elaborata a Berkeley, WikiCite si configura come una piattaforma per *corpora* bibliografici specialistici e aperti: *WikiFactMine* e *ScienceSource* ad esempio, utilizzano Wikidata come archivio bibliografico per dati estratti dalla letteratura accademica dei rispettivi ambiti; lo *Zika Corpus* è stato creato per costruire una base di conoscenza completa e ricca di riferimenti bibliografici che descriva lo stato della ricerca scientifica sul virus Zika; *Inventaire.io* permette di creare in maniera gratuita un corpus di metadati bibliografici con licenza CC0, utilizzando Wikidata come archivio dati; un enorme *set* di dati relativo ad articoli scientifici che utilizzano le proprietà di citazioni di Wikidata è stato inserito nel 2018, passando da 36634758 di citazioni a gennaio 2018 a 174259894 a dicembre 2018.²⁵

L'ultima possibilità, infine, immagina WikiCite come «*a bibliographic commons*»: una parte della comunità considera il progetto come la futura sede di un *corpus* aperto e completo di dati strutturati relativo

²⁵ Statistiche di Wikicite <<http://wikicite.org/statistics.html>>

a ogni possibile fonte, non limitato a ciò che è attualmente citato nei progetti Wikimedia o a *corpora* bibliografici su argomenti specifici. Un bene comune bibliografico, creato in modo collaborativo e che può fungere da *hub* che collega una varietà di basi di conoscenza bibliografica, definito spesso ‘*moonshot*’. Questa ‘*bibliotheca universalis*’ potrebbe competere con i grandi *database* proprietari o semi-proprietari esistenti, fornendo dati strutturati aperti e liberi, riutilizzabili da chiunque perché privi di restrizioni sul *copyright*. I requisiti tecnici, programmatici e finanziari per raggiungere questo obiettivo sono imponenti e vanno ben oltre le possibilità attuali dell’iniziativa, ma è auspicabile un allargamento degli orizzonti in tal senso.

Per identificare i punti di forza e di debolezza delle varie soluzioni ai problemi insiti nella repentina e costante crescita del numero di pubblicazioni inseriti in Wikidata, si sono discussi quattro possibili scenari tecnici, coprendo una vasta gamma di opzioni: si va da una soluzione centralizzata, che corrisponde allo *status quo*, in cui tutti i dati bibliografici strutturati si inseriscono in Wikidata, a un’organizzazione completamente decentralizzata, che va a sfruttare la capacità di federazione di Wikibase, in assenza di un *data store* o un *hub* centrale.²⁶

Fra i maggiori sostenitori del progetto, numerose biblioteche sono state promotrici di iniziative di largo respiro all’interno delle attività di WikiCite.

3. Le biblioteche per WikiCite: una prospettiva internazionale

Migliaia di biblioteche negli ultimi anni stanno transitando alla tecnologia LOD per la gestione dei dati bibliografici, avendo

²⁶ Per un approfondimento sugli scenari e gli eventuali sviluppi di Wikicite, con un’analisi comparata rischi/benefici, si rimanda alla Wikicite Roadmap <<https://m.wikidata.org/wiki/Wikidata:WikiCite/Roadmap>>.

compreso che la connessione dei propri metadati con gli altri dati strutturati presenti nel Web 3.0 permette di aprire nuove strade per il reperimento delle informazioni.

Tali progressi facilitano la partecipazione alle iniziative di WikiCite, per cui negli ultimi anni sono andati moltiplicandosi i progetti delle biblioteche in questa realtà.

Wikimedia Svezia sta collaborando con la *National Library of Sweden* per importare metadati relativi a fonti e dati di autorità su Wikidata. La transizione della biblioteca a Bibframe 2.0 facilita la sinergia con Wikidata, fornendo un primo passo verso l'importazione su vasta scala di informazioni bibliografiche relative alle monografie. Raccogliendo e analizzando i dati sulle fonti più popolari utilizzate nelle voci di Wikipedia in svedese, si è compreso quali metadati bibliografici potessero essere di maggior utilità per la comunità wikipediana, mentre al personale della *National Library of Sweden* che partecipa al progetto sono state mostrate le opportunità offerte dall'utilizzo dei *linked open data*: l'obiettivo più importante è sviluppare una strategia di lavoro che possa giovare contemporaneamente alla comunità Wikimediana e alla biblioteca stessa sul lungo termine.²⁷

Anche la *National Library of Wales*, dopo aver curato un progetto di immissione in Wikidata di dati molto dettagliati relativi al commercio librario nella loro regione, ha predisposto un piano per condividere i metadati di ogni libro pubblicato in Galles, o di interesse galles. Il progetto '*Sum of all Welsh Literature*', dunque, ha lo scopo di raccogliere, preservare e fornire l'accesso a tutte le forme di conoscenza registrata relative al Galles e alla cultura delle popolazioni celtiche, a beneficio di un pubblico molto più ampio di quello delle singole biblioteche gallesi; tutto ciò si traduce nella creazione di circa mezzo milione di nuovi *item* di Wikidata.²⁸

²⁷ Strategisk inkludering av biblioteksdata på Wikidata 2018 <https://se.wikimedia.org/wiki/Projekt:Strategisk_inkludering_av_biblioteksdata_p%C3%A5_Wikidata_2018/Ans%C3%B6kan>.

²⁸ The sum of all Welsh literature, Jason Evans, <<https://docs.google.com/pre>

Con l'elaborazione del *Linked Data Wikibase Prototype*²⁹, OCLC, in collaborazione con sedici biblioteche *partner*, ha portato avanti un progetto pilota estremamente innovativo relativo alla creazione e gestione di descrizioni bibliografiche in *linked data* attraverso la piattaforma Wikibase. Lo scopo è la creazione di strumenti utili a fornire un servizio di riconciliazione (o importazione) delle descrizioni catalografiche delle biblioteche partecipanti e quelle in LOD di Worldcat, attraverso gli identificatori del *Controlled Vocabulary* di WC presenti in Wikidata e un servizio di *editor* utile a visualizzare, creare e modificare le descrizioni e le relazioni fra tali dati.

*Wikibib*³⁰, la sperimentazione in corso alla Biblioteca Nazionale Centrale di Firenze, propone di esplicitare e convertire, a livello sintattico, l'ontologia del MARC in strutture Linked Data/RDF mediante l'uso del Wikibase *data model*. La gestione dei metadati bibliografici su Wikibase consente di salvaguardare la ricchezza dei dati già prodotti e dei cataloghi *legacy*, rendendoli al contempo più facilmente reperibili tramite i motori di ricerca e immediatamente disponibili per il riutilizzo.³¹

4. Case study: il periodico *Bibliothecae.it*

Il progetto di inserimento in Wikidata dei metadati bibliografici degli articoli della rivista *Bibliothecae.it* nasce dalla collaborazione tra il gruppo di lavoro GLAM del Centro Bibliotecario di Ateneo dell'Università degli Studi di Salerno e la Direzione del periodico. È il

sentation/d/1OTLE52im-luTx1l6YhmbxSuJSLYXyRPptF7GKVZTQkw/edit#slide=id.p>.

²⁹ Linked Data Wikibase Prototype <<https://www.oclc.org/research/themes/data-science/linkedata/linked-data-prototype.html>>.

³⁰ Wikibib <https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteca_Nazionale_Centrale_di_Firenze>.

³¹ Per la descrizione completa del progetto si rimanda a Bergamin - Bacchi 2018.

primo esempio italiano e fra i primissimi al mondo di *corpus* relativo ad un periodico scientifico di ambito bibliografico-biblioteconomico interamente caricato in Wikidata.

La realizzazione del progetto si è articolata in tre fasi distinte. Il primo *step* ha comportato la creazione degli *item* per tutti gli autori degli articoli; l'ultima versione del *tool Resolve authors* si è rivelata molto utile sia per creare nuovi *item* che per conciliare eventuali duplicati, nonché per individuare il nome dell'autore sul sito di OrcId e importarne tutti i metadati presenti nella pagina personale.

In questa fase è stato sperimentato anche *Source Meta Data*, un programma che, attraverso l'identificatore persistente di un articolo accademico, di una monografia o di un autore elabora e importa i relativi metadati in Wikidata. Nell'agosto 2018 è stato fuso con *Orcidator* (che raccoglie e importa vari tipi di informazioni da profili OrcId degli autori) in un unico strumento che consente l'elaborazione in *batch* completamente automatizzata. La maggiore criticità evidenziata nell'utilizzo del *tool* è stata il mancato funzionamento con i DOI della rivista, in quanto non registrati in Crossref o in progetti simili.

La seconda fase del progetto ha previsto l'inserimento dei metadati descrittivi di ogni singolo articolo tramite l'utilizzo di Zotero³² e dell'*addon* Zotkat, sviluppato nel corso del 2017. Questa estensione rende possibile l'esportazione dei metadati in un formato comprensibile da QuickStatements,³³ consentendo di creare in maniera automatizzata

³² *Zotero* è un software per la gestione di riferimenti bibliografici e dei materiali ad essi correlati (ad esempio file in formato PDF), libero e open source. Tra le sue principali caratteristiche vi sono l'integrazione all'interno dei più famosi web browser e editor di testo, la sincronizzazione on-line delle bibliografie, la generazione automatica di citazioni, note e bibliografie. <<https://www.zotero.org>>.

³³ *QuickStatements*, oggi alla sua seconda versione, è uno strumento che può modificare gli elementi di Wikidata, basandosi su un semplice set di comandi di testo. Lo strumento può aggiungere, unire e rimuovere statements, etichette, descrizioni e alias, nonché aggiungere statements con qualificatori e fonti. La sequenza di comandi può essere digitata nella finestra di importazione o creata in un foglio di

gli elementi Wikidata relativi alle opere salvate nella libreria di Zotero. Poiché il software, tramite specifici traduttori o metadati generici come CoinS,³⁴ in grado di leggere i metadati su decine di altri siti Web, tra cui Wikipedia, o file strutturati come BibTeX, può essere agevolmente utilizzato come strumento per importare tali dati.

Una volta inseriti gli elementi degli articoli in maniera semiautomatica con QuickStatements, si è proceduto ad 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) e l'argomento principale del contributo (P921).

Per arricchire le etichette nelle varie lingue, in particolare in italiano e inglese, si sono utilizzati *AltLabels*, *Label Collector* e *Name as labels*; per il popolamento degli *item* autore, grazie ad *Orcidator* sono stati importati tutti i dati riferiti nel profilo OrcId; con *DuplicateReferences* si è velocizzata l'aggiunta dei riferimenti per le varie dichiarazioni, mentre *Propbrowse* ha consentito di sfogliare, filtrare e visualizzare rapidamente tutte le proprietà su Wikidata.

La principale problematica riscontrata in questo stadio del lavoro è stata la difficoltà nell'indicare il soggetto degli articoli, dovuta sia alla scarsa copertura di argomenti bibliografici e biblioteconomici in Wikipedia, e di conseguenza in Wikidata, sia al mancato utilizzo da

calcolo o attraverso un editor testuale e successivamente incollata nella finestra. Può anche essere creata da un codice esterno. I dati presenti in OpenRefine possono anche essere esportati nel formato CSV per QuickStatements. <https://tools.wmflabs.org/wikidata-todo/quick_statements.php>.

³⁴ *ContextObjects in Span*, comunemente abbreviato COINS, è un metodo per incorporare i metadati bibliografici nel codice HTML delle pagine Web. Ciò consente al software bibliografico di pubblicare elementi bibliografici leggibili dai software per la gestione di riferimenti bibliografici.

parte degli editori delle riviste scientifiche di un *thesaurus* nella fase di indicizzazione dei documenti.³⁵

La terza fase del progetto, più complessa e tuttora in corso, consiste nell'arricchimento degli *item* degli articoli con tutte i relativi riferimenti bibliografici attraverso la proprietà P2860 (cita).

Questa parte del lavoro, allo stato attuale, è svolta principalmente in maniera manuale perché, anche in riviste presenti in DataCite (come nel nostro caso), è predominante l'assenza dei metadati citazionali; nel caso specifico di OJS,³⁶ usato dalla maggioranza degli editori, le relazioni con altri articoli, ovvero le citazioni, non sono presenti. Si potrebbe pensare ad una modifica del software, ma si avrebbe comunque un unico campo di testo, non controllato e non strutturato, quindi tendenzialmente inutilizzabile in maniera automatica e che andrebbe sottoposto ad un'operazione di *parsing*.³⁷ Allo stato attuale la comunità di WikiCite sta riflettendo sulle possibili soluzioni; una proposta utile sarebbe quella di separare e organizzare i metadati relativi alla bibliografia direttamente all'origine, cioè in fase di pubblicazione del manoscritto, per inserire nei software i dati citazionali già parcellizzati e strutturati, ad esempio attraverso l'utilizzo di BibTex o programmi simili.

Oltre ai *tool* utilizzati per l'inserimento dei metadati bibliografici di *Bibliothecae.it* in Wikidata, ve ne sono molteplici anche per la visualizzazione, analisi e riuso di tali dati. Il più utile e completo è senza dubbio *Scholia*, che risponde a diverse esigenze: oltre a generare i profili degli autori, si presenta come *information discovery tool*, strumento di gestione dei riferimenti bibliografici e *tool* scientometrico.³⁸ Dalla sua nascita, nel 2016, lo strumento è stato continuamente aggiornato:

³⁵ Sarebbe utile che gli editori, in fase di submission del manoscritto, chiedano agli autori di inserire le parole chiave utilizzando vocabolari controllati; nel caso italiano lo strumento adeguato è senza dubbio il thesaurus del Nuovo Soggettario.

³⁶ Open Journal Systems (OJS) è un software open source per la gestione di riviste scientifiche elettroniche. <<https://pkp.sfu.ca/ojs/>>.

³⁷ Cfr. <<https://it.wikipedia.org/wiki/Parsing>>.

³⁸ Scholia <<https://tools.wmflabs.org/scholia/>>.

gli ultimi sviluppi prevedono anche la possibilità di mostrare gli *item* correlati sulla base di cosiddetti *knowledge graph embeddings*, dalle citazioni bibliografiche fino all'indicazione di tempo e luogo degli eventi descritti. Sono state realizzate anche diverse estensioni di Scholia, ad esempio per i *feed RSS* o per la visualizzazione dei dati delle sostanze chimiche. Scholia ha beneficiato della continua espansione delle informazioni bibliografiche in Wikidata grazie all'iniziativa Wikicite, ed è concepito come una componente di *knowledge-as-a-service*.³⁹

Altri strumenti, anche se non destinati prettamente ai metadati bibliografici, possono essere utilizzati in questo senso.

Reasonator è il *tool* che permette di visualizzare gli elementi Wikidata in una versione accattivante e completa, giacché mostra e consente di esplorare tutte le correlazioni di un elemento. Le sue funzionalità sono apprezzabili in maniera particolare per *item* riconducibili agli esseri umani: in un'unica pagina si trovano, ad esempio, tutte le versioni linguistiche del nome, i media caricati in Commons, le relazioni familiari e personali, gli identificativi e le fonti esterne, i *link* alla voce nelle varie versioni linguistiche di Wikipedia, il *concept cloud*, la linea del tempo, il codice QRpedia e tutte le altre possibili proprietà ed elementi, il tutto navigabile attraverso gli *interlink*. Per ciò che riguarda i dati bibliografici, sono presentati in maniera chiara e facilmente leggibile, sotto forma di lista complessa, ricca di tutte le relazioni e i relativi *link*.

³⁹ Per ulteriori approfondimenti e per seguire le evoluzioni del programma si rimanda a: Finn Årup Nielsen - Daniel Mietchen - Egon Willighagen (13 March 2017), «Scholia and scientometrics with Wikidata», *Joint Proceedings of the 1st International Workshop on Scientometrics and 1st International Workshop on Enabling Decentralised Scholarly Communication*, arXiv:1703.04222 , doi:10.5281/ZENODO.1036595; Finn Årup Nielsen - Daniel Mietchen - Egon Willighagen (October 2017), «Scholia, Scientometrics and Wikidata», *The Semantic Web: ESWC 2017 Satellite Events*, doi:10.1007/978-3-319-70407-4_36; Jere D. Odell; Mairelys Lemus-Rojas (16 February 2018), *From Wikidata to Scholia: creating structured linked data to generate scholarly profiles*, doi:10.17605/OSF.IO/GXQ8D.

Ispirato a *Reasonator*, ma con un focus diverso, è *SQID*, giacché vengono mostrate le informazioni sulle classi e sulle proprietà dell'elemento Wikidata, comprese le informazioni e le statistiche derivate.

Wikidata Graph Builder permette di visualizzare i risultati di una Wikidata *query* sotto forma di grafo, attraverso il *Data-Driven Documents*, detto anche *d3.js.*; in tal modo si mostrano in maniera immediata le connessioni fra gli *item* basati sulle proprietà in comune.

Il *Wikidata Query Service* (WDQS) è un pacchetto *software* e un servizio pubblico progettato per fornire un *endpoint* SPARQL che consente di eseguire *query* sul set di dati di Wikidata. Prevede un'interfaccia semplice e intuitiva: grazie all'editor visuale dell' 'Assistente' è possibile lanciare le *query* anche senza conoscere il linguaggio SPARQL, oppure si possono facilmente modificare schemi di interrogazione già preimpostati, a seconda delle informazioni che si vogliono ottenere.⁴⁰ Molto varie sono le possibilità di visualizzazione dei risultati sotto forma di linee del tempo, mappe ad albero, grafici lineari, a dispersione, ad area, diagrammi a bolle, etc., così come la possibilità di scaricare i risultati in formati CSV, TSV, JSON, tabella HTML o immagine SVG.⁴¹

5. Conclusioni

In conclusione, la citazione bibliografica ha ritrovato una nuova centralità nella comunità scientifica e nel mondo dell'informazione. La necessità di rendere liberi, aperti e riutilizzabili i metadati delle pubblicazioni accademiche fa convergere verso uno sforzo unico in tale direzione le varie comunità, che si trovano a collaborare nella progettazione di linee guida e piani di lavoro comuni.

Grazie anche a grandi progetti come I4OC e WikiCite, negli ultimi

⁴⁰ Bielefeldt - Gonsior - Kröttsch 2018.

⁴¹ Wikidata SPARQL query service <<https://query.wikidata.org/>>.

tempi c'è stata una crescente evoluzione verso la possibilità di accesso libero e riutilizzo di tali dati, consentendo la nascita e l'estensione di molti servizi, software, eventi e progetti anche da parte di istituzioni e aziende legate a una visione strettamente tradizionale del mondo dell'editoria e della ricerca.⁴²

La scelta di partecipare all'iniziativa Wikicite includendo in Wikidata l'intero *corpus* dei metadati descrittivi relativi agli articoli di *Bibliothecae.it* rientra a pieno titolo in questa nuova sensibilità ed esigenza: l'inserimento permette l'utilizzo di tali dati in qualsiasi progetto o programma che impieghi la tecnologia dei *linked open data*, nell'ottica di agevolare l'interscambio fra risorse culturali e soddisfare i bisogni informativi dell'utente finale.

Tale operazione si pone anche come buona pratica nella produzione e diffusione di metadati di qualità, giacché un utilizzo non conforme agli standard o la scarsa qualità delle informazioni descrittive può frenare la circolazione della risorsa.⁴³ Disseminare una metadattazione di qualità produce molteplici vantaggi che investono vari campi: la prima ricaduta positiva si riscontra nell'arricchimento dei principali *repositories* e *database* citazionali (WebCite, Mendeley, LibraryThing, etc.), ma grandi benefici sono apportati, di conseguenza, anche alla comunità bibliotecaria, che può utilizzare tali dati per l'*information retrieval*. L'utilizzo di metadati strutturati in *linked open data* diventa fondamentale anche nella creazione e nel mantenimento degli *authority file* locali, ad esempio nella gestione delle forme varianti dei nomi.

In ambito accademico e scientifico, la disponibilità di dati citazionali aperti favorisce la misurazione e la valutazione dell'impatto dei saggi pubblicati sulla rivista *Bibliothecae.it*, sia qualitativamente, mettendo in luce le relazioni esistenti tra vari filoni di studio, sia quantitativamente, con la creazione di opportune metriche di analisi dei dati.

Per i contributori dei progetti Wikimedia il vantaggio si evidenzia

⁴² Peroni 2018.

⁴³ Weston - Sardo 2017, p. 20.

nella possibilità di utilizzare riferimenti bibliografici di alta qualità, nell'opportunità di unificare e centralizzare le citazioni e nella facilitazione di trasclusione dei metadati via API.

Grazie all'iniziativa, i fruitori dei progetti Wikimedia possono valutare la qualità delle citazioni relative al periodico analizzandole nel loro contesto, vedere tutte le risorse in cui viene utilizzata la citazione, trovare e confrontare fonti simili.

Oltre ai vantaggi descritti, le pratiche collaborative proprie dei progetti Wikimedia appaiono adeguate anche all'assolvimento della cosiddetta 'terza missione' delle Università, cioè all'impegno delle comunità accademiche nel favorire l'applicazione diretta, la valorizzazione e l'impiego della conoscenza per contribuire allo sviluppo sociale, culturale ed economico della società.

Si auspica che la comunità bibliotecaria condivida anche questa nuova sfida, contribuendo, ognuno a seconda dei propri mezzi, conoscenze e abilità tecniche, alle varie iniziative a favore della conoscenza libera e condivisa, creando basi solide per una ulteriore evoluzione delle scienze biblioteconomiche e dell'informazione.

Bibliografia

- Allison - Scott 2018 = Stacy Allison-Cassin - Dan Scott, *Wikidata: a platform for your library's linked open data*, «Code4Lib Journal», 40 (2018), <<https://journal.code4lib.org/articles/13424>>.
- Anderson 2018 = Clifford B. Anderson, *Introduction to Wikidata, The Wikipedia of Facts*, In-Conference Workshop ATLA 2018 Proceedings, p. 307-309 <<https://serials.atla.com/proceedings/article/view/117/620>>.
- Ayers - Matthews - Yates 2008 = Phoebe Ayers - Charles Matthews - Ben Yates, *How Wikipedia Works: And How You Can Be a Part of It*, San Francisco CA, No Starch Press, 2008.
- Bartholmei et al. 2016 = Stephan Bartholmei - Rachel Franks - James Heilman - Mylee Joseph - Vicki McDonald - Anna Raunik - Mia Ridge - Mark Robertson, *Opportunities for Academic and Research Libraries and Wikipedia*, Columbus, Ohio, 2016. <<https://www.ifla.org/files/assets/hq/topics/info-society/iflawikipediaopportunitiesforacademicandresearchlibraries.pdf>>.
- Bergamin - Bacchi 2018 = Giovanni Bergamin - Cristian Bacchi, *New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data*. «JLIS.it», 9 (2018), n. 3, p. 35-74, <<https://www.jlis.it/article/view/12458/11344>>.
- Bielefeldt - Gonsior - Kröttsch 2018 = Adrian Bielefeldt - Julius Gonsior - Markus Kröttsch, LDOW'2018, April 2018, Lyon, France, *Practical Linked Data Access via SPARQL: The Case of Wikidata* <<http://ceur-ws.org/Vol-2073/article-03.pdf>>.
- Boccone - Rivelli 2018 = Alessandra Boccone - Remo Rivelli, *Biblioteche e progetti Wikimedia: un percorso comune*, «Biblioteche oggi», 36 (2018), n. 4, p. 37-48. DOI: 10.3302/0392-8586-201803-037-1.
- Heibi - Peroni - Shotton 2019 = Ivan Heibi, Silvio Peroni, David Shotton, *COCI, the Open Citations Index of Crossref open DOI-to-DOI citations*, arXiv:1904.06052v1 [cs.DL] 12 apr 2019

- Martinelli 2016 = Luca Martinelli, *Wikidata: la soluzione wikimediana ai linked open data*, «AIB Studi», 56 (2016), n. 1, p. 75–85. DOI: 10.2426/aibstudi-11434.
- Peroni - Shotton - Vitali 2017 = Silvio Peroni - David Shotton - Fabio Vitali, *One Year of the OpenCitations Corpus*, in *The Semantic Web – ISWC 2017*, a cura di Claudia d'Amato et al., Cham, Springer, 2017. DOI 10.1007/978-3-319-68204-4_19.
- Peroni 2018 = Silvio Peroni, *I4OC o la fine dell'oligarchia delle citazioni*, 2018 <<https://www.roars.it/online/?p=64905>>.
- Schiermeier 2017 = Quirin Schiermeier, *Initiative aims to break science's citation paywall*, Nature News (6 April 2017) DOI: 10.1038/nature.2017.21800.
- Taraborelli - Dugan 2017 = Dario Taraborelli - Jonathan Dugan, *How we know what we know: The Initiative for Open Citations (I4OC) helps unlock millions of connections between scholarly research*, Wikimedia Blog, <<https://blog.wikimedia.org/2017/04/06/initiative-for-open-citations/>>.
- Taraborelli 2018 = Dario Taraborelli, *The citation graph is one of humankind's most important intellectual achievements*, Boing Boing, <<https://boingboing.net/2018/04/14/open-graphs.html>>.
- Vrandečić - Krötzsch 2014 = Denny Vrandečić - Markus Krötzsch, *Wikidata: A Free Collaborative Knowledgebase*, «Communications of the ACM», 57 (2014), n. 10, p. 78-85 DOI 10.1145/2629489.
- Weinberger 2012 = David Weinberger, *Linking is a public good*, 2012, <<http://www.hyperorg.com/blogger/2012/02/26/2b2k-linking-is-a-public-good/>>
- Weston - Sardo 2017 = Paul Gabriele Weston, Lucia Sardo, *Metadati*, Roma, Associazione Italiana Biblioteche, 2017.

Abstract

Negli ultimi anni l'attenzione del mondo bibliotecario volto ai progressi della tecnologia Linked Open Data e ai progetti Wikimedia si è concentrata su Wikidata, la cui infrastruttura si è rivelata perfettamente compatibile con Wikicite, iniziativa che mira a costruire un vasto database di citazioni bibliografiche di qualità in linked open data. Nell'ambito di tale iniziativa, molti sono i progetti nati dalle collaborazioni fra la Wikimedia Foundation e le biblioteche: anche il Centro Bibliotecario di Ateneo dell'Università degli Studi di Salerno ha partecipato attivamente all'iniziativa, inserendo i metadati bibliografici della rivista Bibliothecae.it. È il primo esempio italiano e fra i primissimi al mondo di corpus relativo ad un periodico scientifico di ambito bibliografico-biblioteconomico interamente caricato in Wikidata. L'inserimento permette l'utilizzo di tali dati in qualsiasi progetto o programma che impieghi la tecnologia LOD, nell'ottica di agevolare l'interscambio fra risorse culturali e soddisfare i bisogni informativi dell'utente finale. In ambito bibliografico e biblioteconomico, tale azione rappresenta un nuovo modo di creare e condividere i dati bibliografici, di fare information retrieval e di condurre le analisi bibliometriche. Le ricadute positive del progetto riguardano sia la comunità wiki, che ha bisogno di fonti affidabili e verificate su cui basare i propri contenuti, sia la comunità scientifica: pubblicando in maniera aperta i set di citazioni si rende open access un enorme grafo di conoscenza scientifica, favorendo la divulgazione, la pubblicazione e la valutazione aperta dei risultati della ricerca.

LOD; Wikidata; Bibliothecae.it

In recent years, the attention of the librarian world aimed at the progress of Linked Open Data technology and Wikimedia projects has focused on

Wikidata, whose infrastructure has proved to be perfectly compatible with Wikicite, an initiative that aims to build a vast database of bibliographic references in linked open data. As part of this initiative, many projects originate from collaborations between the Wikimedia Foundation and libraries: the Library of the University of Salerno also actively participated in the initiative, including the bibliographic metadata of the journal Bibliothecae.it. It is the first Italian example of corpus related to a scientific periodical of bibliographic-biblioteconomical scope entirely loaded in Wikidata. The project allows the use of data in any project or program that uses the LOD technology, in order to facilitate the interchange between cultural resources and satisfy the information needs of the user. In the area of bibliography and librarianship, this action represents a new way of creating and sharing bibliographic data, of doing information retrieval and of conducting bibliometric analysis. The positive effects of the project concern the wiki community, which needs reliable and verified sources to base its contents, and the scientific community: by openly publishing the citations sets, an enormous graph of scientific knowledge is opened, encouraging the dissemination, publication and open evaluation of research results.

LOD; Wikidata; Bibliothecae.it