

Riuso, interoperabilità, influenza: la cooperazione virtuosa tra i progetti SHARE e Wikidata.

di Tiziana Possemato (@Cult) e Claudio Forziati (Università degli studi di Napoli Federico II)

Abstract

Nell'epoca della frammentazione dell'informazione, l'integrazione e la cooperazione tra progetti diversi contribuisce ad accrescere la conoscenza. Il contributo presenta i processi di identificazione delle entità nei progetti di linked data della famiglia "SHARE" e il contributo prezioso offerto da progetti trasversali come quello di Wikidata. Indicatori di successo di un progetto, quali il riuso, l'interoperabilità e l'influenza, sono osservati per valutare questa cooperazione e proporre altre forme di arricchimento possibili.

La famiglia SHARE e la LOD Platform

I progetti inclusi nella famiglia SHARE sono promossi dalle biblioteche per stabilire procedure per l'identificazione e la riconciliazione di entità, la conversione di dati in Linked Data e la creazione di un ambiente di discovery virtuale basato sulla struttura a tre livelli del modello di dati BIBFRAME.

A questa famiglia appartengono importanti iniziative, quali SHARE Catalogue, il catalogo unico in linked open data, già presentato nel 2016 in occasione delle Stelline e che rientra in un più vasto progetto di cooperazione e di condivisione di servizi tra le biblioteche campane, lucane e salentine¹; SHARE-VDE (Virtual Discovery Environment), uno sforzo collaborativo basato sulle esigenze delle differenti biblioteche partecipanti, frutto della collaborazione tra Casalini Libri (agenzia bibliografica e fornitore di dati bibliografici e record di autorità come membro del Program for Cooperative Cataloging) e @Cult (un fornitore di Integrated Library System (ILS) e strumenti di discovery), con l'input iniziale di 16 biblioteche universitarie nordamericane. La governance dell'iniziativa è condivisa con la comunità delle biblioteche coinvolte.

Alla medesima famiglia appartengono progetti dedicati ad ambiti disciplinari specialistici in corso di realizzazione o di analisi, quali:

- SHARE-Art, prototipo per le biblioteche di storia dell'arte del Max Planck Institut quali il Zentralinstitut für Kunstgeschichte di Monaco di Baviera, il Kunsthistorisches Institut di Firenze, la Bibliotheca Hertziana a Roma e il Centre Allemand d'Histoire de L'Art di Parigi (gruppo Kubikat); una delle peculiarità di questo progetto è di prevedere l'integrazione della fototeca e di costruire un ponte per dare all'utente la possibilità di spostarsi dall'ambito bibliotecario a quello delle collezioni museali.
- SHARE-Music, sperimentazione nell'ambito della musica con la partecipazione della Bayerische Staatsbibliothek di Monaco, della Library of Congress e della Stanford University; questo progetto affronta aspetti molto specifici come, a titolo di esempio, la definizione e gestione ai vari livelli dell'Opera (Work) e l'applicazione dell'ontologia PMO (Performed Music Ontology).

Da un punto di vista tecnologico questi progetti sono per lo più basati sulla **LOD Platform**, un sistema tecnologico innovativo per la gestione dei dati bibliografici, archivistici e museali, e la loro trasformazione in LD.

Lo scopo del framework è di aprire le possibilità offerte dai LD a biblioteche, archivi e musei fornendo maggiore interoperabilità, visibilità e reperibilità per tutti i tipi di risorse.

¹ Homepage di SHARE Catalogue <http://catalogo.share-cat.unina.it/sharecat/clusters>.

Come funziona la LOD Platform

Componenti e strumenti inclusi nella LOD Platform sono pensati per gestire progetti di conversione dei dati in Linked Open Data, creando un ambiente integrato che consente di selezionare le fonti con le quali arricchire i propri dati, definire le ontologie da utilizzare, pubblicare i risultati in maniera utile non solo alle macchine ma anche a fruitori ‘umani’. Il framework è costituito da tool tecnologici che provvedono al trattamento dei dati dall’origine fino alla pubblicazione nel web:

- **Authify**, un modulo RESTful che fornisce servizi di ricerca e full-text di dataset esterni (scaricati, memorizzati ed indicizzati nel sistema), relativi soprattutto ad Authority file (VIAF, Library of Congress Name Authority file, etc...) ma estendibile anche ad altre tipologie di dataset. È composto da due parti principali: un’infrastruttura SOLR per l’indicizzazione dei dataset e relativi servizi di ricerca, ed un livello logico che orchestra tali servizi per trovare una corrispondenza all’interno dei cluster delle Entità definite
- **Cluster Knowledge Base**, su database PostgreSQL, è il risultato del processo di elaborazione e arricchimento dei dati con fonti esterne per ogni entità definita; tipicamente: cluster di nomi (forme, autorizzate e varianti, dei nomi delle Persone) e cluster di titoli (punti di accesso autorizzato e forme varianti per i titoli delle Opere)
- **Lodify**, modulo RESTful che automatizza l’intero processo di conversione e pubblicazione dei dati in RDF secondo l’ontologia BIBFRAME 2.0 in modo lineare e scalabile. È flessibile e adattabile a molteplici situazioni: permette, quindi, di gestire le classi e le proprietà non solo di BIBFRAME ma anche di altre ontologie, a seconda delle esigenze
- **Blazegraph/Stardog**, triplestore per lo storage dei file RDF. La LOD Platform integra due possibili triplestore, uno open source (Blazegraph²), più adatto a progetti di piccole o medie dimensioni (fino a circa 2.000.000 di record bibliografici), e uno commerciale (Stardog³) più adatto a situazioni di importanti dimensioni.
- **Portale di presentazione dei dati**, per una modalità user-friendly di pubblicazione e fruizione dei dati.

Il ciclo di lavorazione dei dati in un progetto SHARE

Nello schema riportato in figura 1 si illustra un processo ‘tipo’ di trattamento dei dati, dalla ricezione alla pubblicazione sul Portale SHARE.

Il flusso è solo indicativo, ma esprime in modo esaustivo gli step inclusi in un processo di elaborazione dei dati.

Partendo dalla sinistra del grafico, i dati sono ricevuti dagli Enti partecipanti (biblioteche, archivi, musei) in differenti formati (MARC, xml etc.). I dati possono essere bibliografici e di autorità.

I dati ricevuti sono elaborati secondo processi di *Text analysis* e *String matching* (rappresentati nel box “*Similarity’s score*”), per identificare le Entità incluse nei testi ‘piatti’ (i record), e preparare la creazione dei cluster.

La funzione di identificazione delle entità è potenziata ed estesa attraverso analoghi processi di *Text analysis* e *String matching* lanciati su fonti esterne (VIAF, ISNI, LCNAF, GND, Nuovo soggetto etc.), attraverso il framework *Authify*: questi processi generano l’arricchimento del dato con altre forme varianti provenienti dalle fonti esterne e con gli URI attraverso i quali la medesima entità è identificata su queste fonti (*riconciliazione*): il cluster di origine si arricchisce e consente, nel processo di conversione in Linked data, di attivare la funzione di *interlinking*, essenziale per la condivisione e il riuso dei dati nel web.

² <https://www.blazegraph.com/>.

³ <https://www.stardog.com/>.

Il risultato di questi processi è triplice:

- Identificazione delle entità
- Arricchimento dei dati
- Creazione dei cluster attraverso processi di riconciliazione

I dati così ottenuti sono pronti per essere nuovamente processati, attraverso differenti canali:

- arricchimento manuale e quality check (nel caso in cui la biblioteca richiedesse uno specifico servizio ad agenzie esterne o gestisse internamente i dati arricchiti ricevuti);
- estrazione delle relazioni ‘nascoste’ per la generazione e alimentazione di un *Database delle relazioni* (che verrà riutilizzato in possibili successivi step di arricchimento del dato e nelle fasi di pubblicazione, per estendere i collegamenti tra dati);
- generazione della *Cluster Knowledge Base* di progetto, disponibile in RDF (quindi come Linked Open Data) e accessibile come end point per query SPARQL e API.
- Elaborazione/conversione in RDF, seguendo il modello BIBFRAME e/o altre ontologie di dominio suggerite dal progetto.

Al termine di questi processi, i dati sono pronti per essere indicizzati sul Portale ed essere pubblicati su vari siti, in RDF.

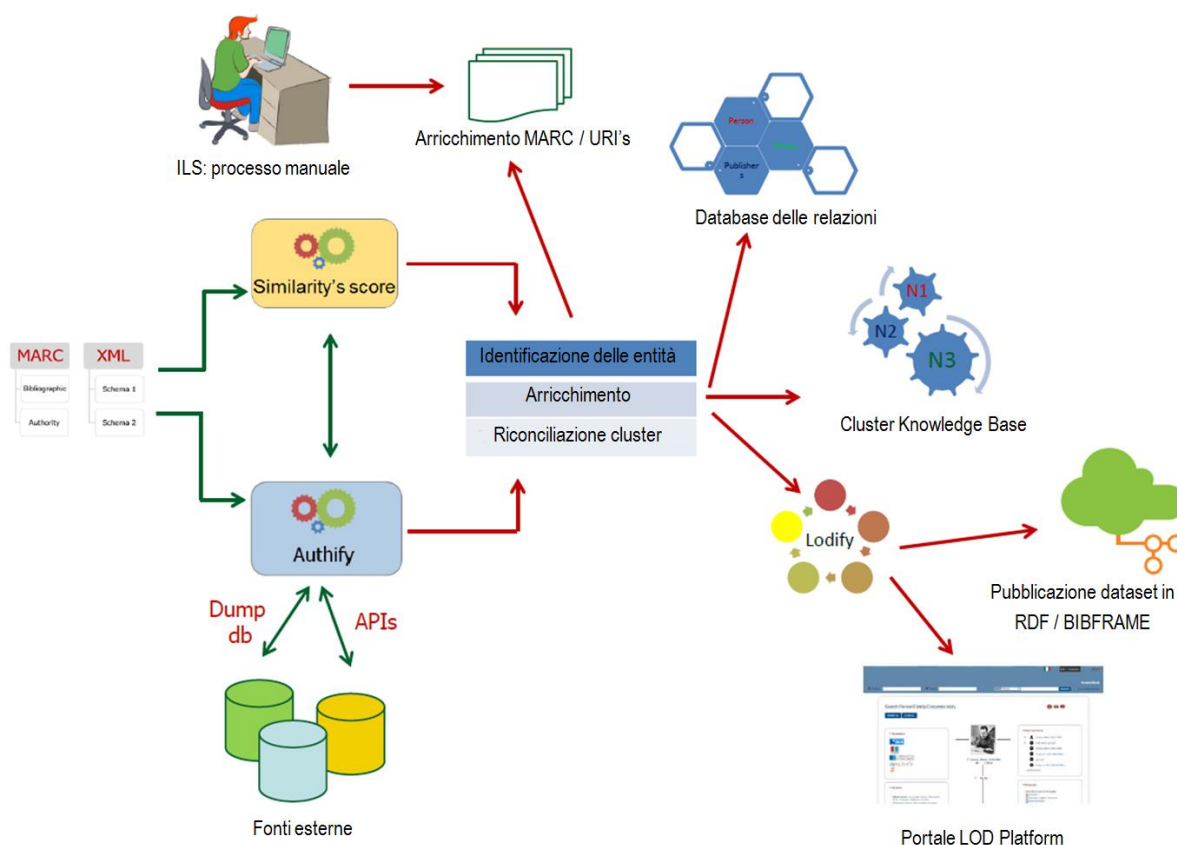


Fig. 1 LOD Platform workflow in un progetto SHARE.

Wikidata: una base di conoscenza *community-driven*

Wikidata è una base di conoscenza libera e collaborativa, rilasciata pubblicamente alla fine di ottobre del 2012, interpretabile da umani e macchine. Ultima nata fra i progetti gestiti dalla Wikimedia Foundation, ha come scopo principale la centralizzazione dei dati in forma strutturata dei progetti Wikimedia e il supporto, come *knowledge base*, a tutti questi progetti, incluse le circa trecento

versioni linguistiche di Wikipedia, con la possibilità di utilizzare i dati anche esternamente grazie alle licenze applicate⁴.

Le entità presenti in Wikidata sono di due tipi: gli *item* e le proprietà, entrambi caratterizzati da un identificatore univoco. Gli *item* sono definiti a partire da etichette, descrizioni e alias, potenzialmente in ognuna delle lingue supportate, e i dati legati a questi item sono rappresentati utilizzando relazioni in forma proprietà-valore.

Le proprietà definiscono le relazioni per gli item, utilizzando come valori item stessi, stringhe, URL, date, coordinate, ecc.⁵ Per ognuno di questi valori, è possibile aggiungere qualificatori, che forniscono informazioni di contesto, e fonti a supporto dei valori dichiarati (fig. 2).

Come in Wikipedia, la creazione e l'arricchimento di dati e documentazione avviene collaborativamente e gli *editors* di Wikidata sono inoltre liberi di proporre la creazione di nuove proprietà⁶ relative all'integrazione di uno specifico *data model* o a identificatori esterni ricavati da fonti online, utilizzabili in uno o più domini, connotando in questo modo una base di conoscenza con una struttura in continua evoluzione.

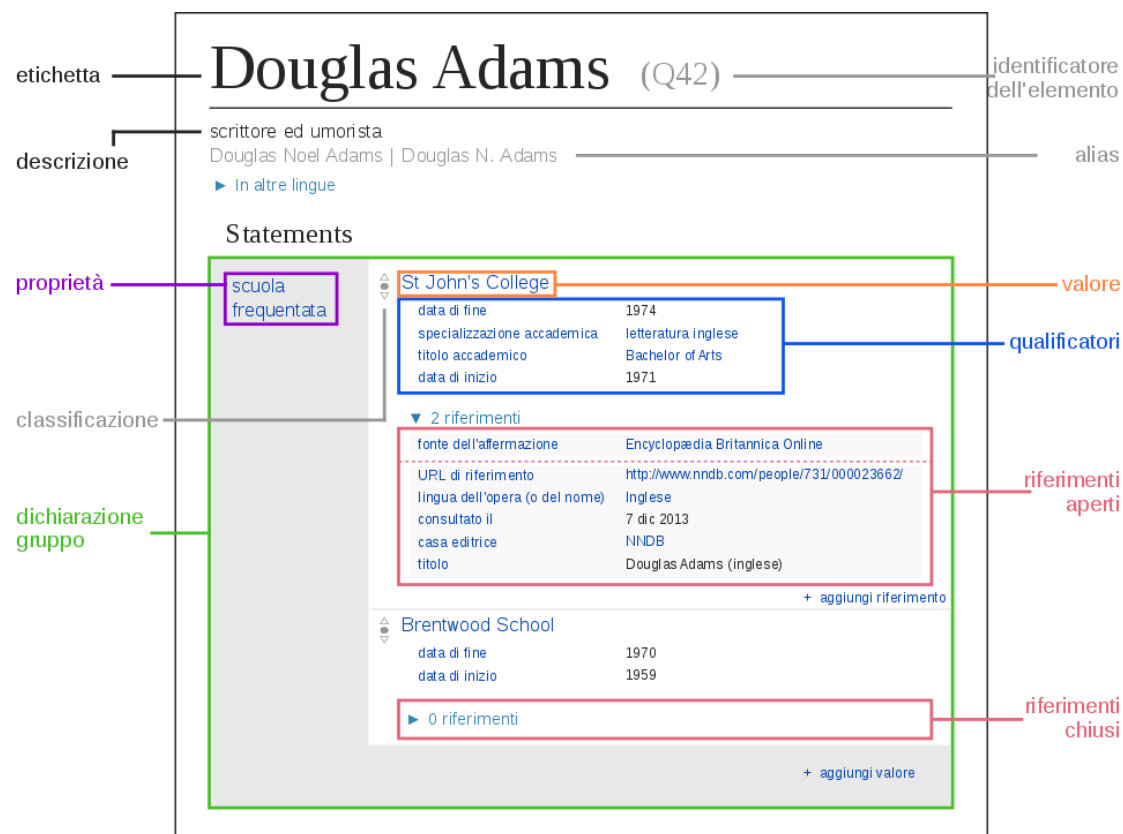


Fig. 2 Rappresentazione del *data model* di Wikidata. Di Charlie Kritschmar (WMDE), Sabas88 [CC0], attraverso Wikimedia Commons

SHARE Catalogue - Wikidata: le fasi operative

Una delle caratteristiche dei progetti SHARE è l'ampia capacità di integrazione con fonti e progetti esterni, sia per i processi di identificazione e riconciliazione delle entità, sia per i processi di arricchimento dei dati. Uno degli esperimenti più interessanti di questa forma di cooperazione tra

⁴ Cfr. Luca Martinelli, «Wikidata: la soluzione wikimediana ai linked open data», *AIB studi* 56, n. 1.

⁵ Help:Data type, https://www.wikidata.org/wiki/Help:Data_type.

⁶ Wikidata:Property proposal, https://www.wikidata.org/wiki/Wikidata:Property_proposal.

progetti è quello rappresentato dalla integrazione con Wikidata, con la finalità di valorizzazione reciproca del patrimonio informativo, come sarà meglio spiegato nel seguito di questo documento.

Nella fase preliminare della sperimentazione il gruppo di lavoro SHARE, grazie al supporto di utenti esperti, ha utilizzato le modalità previste dalla comunità di Wikidata per consentire l'interazione dei due progetti, formulando una *proposal*⁷ per la creazione di una proprietà legata agli identificatori per *Agent* (persone/famiglie/enti) di SHARE Catalogue⁸. L'approvazione della *proposal* ha permesso di elaborare un workflow utile a pianificare il caricamento dei dati in stati di avanzamento definiti.

Innanzitutto è stata necessaria un'estrazione dei dati complessivi di SHARE Catalogue in formato tabulare, formato facilmente utilizzabile per allargare la base dei bibliotecari partecipanti che negli incontri di autoformazione hanno potuto fare pratica con semplici inserimenti manuali⁹, contestualmente confrontandosi con la ricca documentazione presente in Wikidata, le norme di rilevanza stabilite dalla comunità¹⁰ e aspetti tecnicamente più complessi come le violazioni dei requisiti di vincolo del database per specifiche proprietà¹¹.

Nella fase successiva il gruppo di lavoro SHARE ha determinato con maggiore precisione gli strumenti da utilizzare, definendo due modalità operative:

- il caricamento in batch di dati attraverso il tool QuickStatements¹² (fig. 3);
- la riconciliazione, tuttora in corso, delle entità tramite OpenRefine¹³, in particolare per tutte quelle che non contenevano un link a Wikidata in SHARE Catalogue come fonte esterna benché esistesse un item corrispondente (fig. 4-5).

⁷ La *proposal* per la creazione di una proprietà riferita a SHARE Catalogue è reperibile all'indirizzo https://www.wikidata.org/wiki/Wikidata:Property_proposal/SHARE_Catalogue_author_ID.

⁸ Tra gli output più interessanti dei progetti SHARE c'è la produzione della *Cluster Knowledge Base* (CKB), un dataset autorevole, risultato del processo di elaborazione e arricchimento dei dati con fonti esterne per ogni entità definita. Tra le entità di maggior interesse la CKB include gli *Agent* (cluster di nomi composti dalle forme, autorizzate e varianti, dei nomi di Persone, Famiglie, Enti) e i *Work* (cluster di titoli, punti di accesso autorizzato e forme varianti per i titoli delle Opere).

⁹ «Compared to classic methods of generating LOD [...] Wikidata's editing interface presents a comparatively simple key-value approach for creating and editing statements». Stacy Allison-Cassin e Dan Scott, «Wikidata: a platform for your library's linked open data», *The Code4Lib Journal*, n. 40.

¹⁰ Wikidata:Notability, <https://www.wikidata.org/wiki/Wikidata:Notability>.

¹¹ «[...] the platform introduces the basic ontological principle that properties and classes can have additional requirements without frustrating new contributors at the point of data entry». Allison-Cassin e Scott, «Wikidata: a platform for your library's linked open data». Cfr. inoltre Help:Property constraints portal, https://www.wikidata.org/wiki/Help:Property_constraints_portal.

¹² QuickStatements 2, <https://tools.wmflabs.org/quickstatements/>.

¹³ Open Refine, <http://openrefine.org/>.

1	done	Emilio Reus [Q8775574]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325060"
2	done	George Reese [Q5543776]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325069"
3	done	Eva Grlić [Q5415058]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325071"
4	done	Ugo Intini [Q3547806]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325082"
5	done	Giovanni Pieraccini [Q3767938]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325084"
6	done	Umberto Veronesi [Q780378]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325087"
7	done	Christine Chivallon [Q18744687]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325090"
8	done	Michael Borgolte [Q1927023]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325095"
9	done	Manuel Hueriga [Q3080940]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325097"
10	done	Michel Biron [Q3309022]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325099"
11	done	Lajos Thanhoffer [Q1318976]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325102"
12	done	Gaspar Lucas Hidalgo [Q5875822]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325113"
13	done	Aldo Leopold [Q971382]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325120"
14	done	Paul Groussac [Q2550159]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325121"
15	done	Giovanni Targioni-Tozzetti [Q3768207]	ADD	Statement	SHARE Catalogue author ID [P3987] : "325122"

Fig. 3 Caricamento dei valori di SHARE Catalogue in Wikidata tramite Quickstatements.

OpenRefine OR_SHARE ID_1419-2667_WD NULL [Link Permanente](#)

Faccette / Filtri Annulla / Rifai 4 / 4

Ricarica Resetta tutto Rimuovi tutti

label: judgment cambia inverte reset

2 choices Ordina per: nome quantità

matched 97 exclude

none 0/0 exclude

Faccetta per quantità alternative

label: best candidate's score cambia reset

15.00 — 101.00

Numerico Non-numeric Blank Error

97 0 0 0

97 matching righe (706 total)

Vista a: righe records Mostra: 5 10 25 50 righe

« primo < precedente 1 - 50

	Tutti	URI	n_cluster_id	label	Wikidata_ID
3.		http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1422	1422	Henry Herman Hausner <small>Scegli un nuovo match</small>	Q59628109
12.		http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1442	1442	Walter John Moore <small>Scegli un nuovo match</small>	Q59629776
15.		http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1454	1454	Theodore I Kamins <small>Scegli un nuovo match</small>	Q55300558
24.		http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1474	1474	Wilbur L. Nelson <small>Scegli un nuovo match</small>	Q16648037
39.		http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1499	1499	Guido Oberti <small>Scegli un nuovo match</small>	Q56629413
43.		http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1505	1505	Otto Oldenberg <small>Scegli un nuovo match</small>	Q59533736
44.		http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1508	1508	Luigi Olivieri <small>Scegli un nuovo match</small>	Q3839950
53.		http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1525	1525	Bernard Oudin <small>Scegli un nuovo match</small>	Q56855756

Fig. 4 L'esito del lavoro in OpenRefine su un subset di 706 record di SHARE Catalogue, di cui 97 sono stati riconciliati con i corrispondenti item di Wikidata.

Wilbur L. Nelson (Q16648037)

identificativo autore
SHARE Catalogue
(P3987)

1474

1 note

URL di riferimento
(P854)

http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1474

Guido Oberti (Q56629413)

identificativo autore
SHARE Catalogue
(P3987)

1499

1 note

URL di riferimento
(P854)

http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=1499

Fig. 5 L'anteprima in OpenRefine delle modifiche che saranno apportate in Wikidata (funzione disponibile a partire dalla versione 3.0).

In parallelo il gruppo di lavoro ha dedicato un segmento del progetto ad arricchire gli item di Wikidata interessati dagli identificatori di SHARE Catalogue con label e descrizioni in diverse lingue, e soprattutto con ulteriori *external id*, allo scopo di migliorare i risultati di specifiche indagini effettuate tramite il *query service* di Wikidata. Per i dati di autorità in particolare, è stato utilizzato il potenziale combinato delle query SPARQL con la funzione *SPARQL-based matches* del tool

Mix'n'match¹⁴, grazie alla quale è possibile confermare gli abbinamenti automatici relativamente a specifici authority¹⁵. Questa attività è servita a limitare la quantità di item che avessero l'identificatore di SHARE Catalogue come unico valore esterno per Wikidata¹⁶ e a monitorare la compresenza di altri identificatori, attualmente non utilizzati come fonti esterne per SHARE Catalogue, ma di grande utilità per espandere il livello di dettaglio delle entità e per l'accesso a informazioni controllate. Come descritto nella tabella seguente, in alcuni casi questa compresenza in Wikidata riguarda oltre l'ottanta per cento degli item che hanno un identificatore per SHARE.

Tot. occorrenze SHARE Catalogue author ID (Identificatore in Wikidata: P3987)				70.128
Nome identificatore	Identificatore proprietà in Wikidata	Occorrenze tot. in Wikidata	Quantità in item con P3987	% in item con P3987
SUDOC authorities ID	P269	413.301	59.814	85,29%
NTA identifier	P1006	382.277	58.710	83,72%
GND ID	P227	737.155	52.014	74,17%
BNE ID	P950	140.240	18.595	26,52%
SBN author ID ¹⁷	P369	30.791	9.968	14,21%

Tab. 1 Dati aggiornati al 4 febbraio 2019

Analisi dei dati e scenari evolutivi

Per la realizzazione del progetto è stato determinante stabilire cosa differenziasse Wikidata nel nostro interesse rispetto ad altre *knowledge base*, quali pratiche passate o in corso potessero esserci d'aiuto¹⁸ e cosa volessimo ottenere da una maggiore interoperabilità.

Di fatto due aspetti ci sono sembrati fortemente incoraggianti. Il primo riguarda sicuramente la comunità di Wikidata e la modalità diffusa e costruttiva di creazione e controllo dei contenuti, frutto di costante conversazione fra gli utenti. Il secondo è la grande ricchezza, regolarmente incrementata da utenti umani e da *bot*, di informazioni in Wikidata, spesso referenziate, soprattutto di natura extra-bibliografica¹⁹ e quanto questa ricchezza garantisca forme di analisi e possibilità di combinazione dei dati non immaginabili in precedenza.

¹⁴ Mix'n'match, <https://tools.wmflabs.org/mix-n-match/>.

¹⁵ Per il dettaglio cfr. Claudio Forziati e Valeria Lo Castro, «La connessione tra i dati delle biblioteche e il coinvolgimento della comunità: il progetto SHARE Catalogue-Wikidata», *JLIS.it* 9, n. 3 p. 117, fig. 4.

¹⁶ Si veda il risultato della relativa query <http://tinyurl.com/yc3y8bny>.

¹⁷ Gli usi correnti con la compresenza dei due identificatori negli item di Wikidata posso essere visualizzati all'indirizzo <http://tinyurl.com/y9y8z9yj>.

¹⁸ Un breve excursus dei progetti che ci hanno fornito utili spunti teorici e operativi è disponibile in Forziati e Lo Castro, «La connessione tra i dati delle biblioteche e il coinvolgimento della comunità», p. 113-115.

¹⁹ «More than any other factor, however, it is the richness of the data that makes Wikidata unique. Many statements come with provenance information or include additional context data, such as temporal validity; data is strongly connected to external datasets in many domains; and all of the data is multi-lingual by design». Fredo Erxleben et al., «Introducing Wikidata to the Linked Data Web», in *The Semantic Web – ISWC 2014*, Lecture Notes in Computer Science (International Semantic Web Conference, Springer, Cham, 2014), 50–65.

Grazie a questo secondo aspetto, abbiamo affiancato al lavoro di caricamento diverse tipologie d'indagine: dall'analisi quantitativa come mostrato in tabella 1 (benché, come detto, con lo scopo di rafforzare il corredo di fonti esterne per le entità di SHARE), ottenibile con una conoscenza basilare di SPARQL²⁰, con una interrogazione piuttosto semplice tramite il *query service* di Wikidata, come la seguente

```
SELECT (COUNT(DISTINCT ?item) AS ?entities)
WHERE
{
  ?item wdt:P3987 ?share_catalogue_id;
    wdt:P269 ?sudoc_id.
}
```

per passare alla generazione in tempo reale e sempre aggiornata di liste di contenuti assenti in una data versione linguistica di Wikipedia o in Wikimedia Commons²¹, utili a pianificare una tipologia di contributo che, lì dove alimenta questi due progetti coerentemente con le rispettive regole di rilevanza e copyright, arricchisce come in una reazione a catena i progetti come SHARE Catalogue che ne riutilizzano i contenuti²².

Sempre in questa chiave, sono state effettuate interrogazioni che relativamente a uno specifico sottoinsieme di entità di SHARE ci consentono di determinare un accesso più immediato a opere in pubblico dominio presenti in Wikisource, coerentemente con la prospettiva di lavorare per l'evoluzione di un prodotto bibliografico non esclusivamente focalizzato sul possesso dei documenti²³.

Indagini di questo genere ci portano a prospettare uno scenario in trasformazione per il nostro catalogo in cui, utilizzando processi simili ai servizi di visualizzazione integrati nel *query service* di Wikidata, potremo includere dei grafi dei coautori per specifiche entità (fig. 6), l'analisi citazionale per un'opera²⁴, visualizzare su mappa un gruppo di autori aggregati per nazionalità e occupazione²⁵:

²⁰ Un manuale semplice e utile da cui partire è reperibile all'indirizzo <https://en.wikibooks.org/wiki/SPARQL>.

²¹ Il risultato di una query in cui viene generata una lista di mineralogisti presenti in Wikidata, con un valore per SHARE Catalogue, il collegamento a un'immagine in Commons (se presente), senza una voce nella Wikipedia in lingua italiana <http://tinyurl.com/y7ldkqrf>.

²² Una delle modalità più utili è la progettazione di una specifica attività con i GLAM (si vedano ad esempio le pagine di coordinamento nella Wikipedia in italiano all'indirizzo <https://it.wikipedia.org/wiki/Progetto:GLAM>). Fra i progetti territorialmente di riferimento per gli enti partecipanti a SHARE Catalogue vanno segnalati il Progetto GLAM/AIB Campania attivo dal 2015 (https://it.wikipedia.org/wiki/Progetto:GLAM/AIB_Campania, cfr. anche Alessandra Boccone et al., «Valorizzazione dei fondi privati in una biblioteca accademica: divulgazione, comunicazione, ricerca», *Bibliothecae.it* 6, n. 2: 255–84), e il laboratorio wiki che si è svolto nella primavera del 2018 presso la Biblioteca di Area Agraria dell'Università degli studi di Napoli Federico II ([https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteca di agraria Federico II](https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteca_di_agraria_Federico_II)), durante il quale sono state realizzate, fra le altre, la voce in italiano di Nicola Antonio Pedicino, l'incipit della cui biografia è ora visualizzato in SHARE Catalogue all'indirizzo http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=82266 e quella in inglese di Orazio Comes, che è possibile visualizzare modificando la lingua d'interrogazione del catalogo: http://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=69925&l=en.

²³ Si veda il risultato di una query che interroga Wikidata per gli autori con un valore per la proprietà identificativo SHARE Catalogue, nati in Campania e morti fra 1800 e 1947 con il link alle rispettive opere disponibili in Wikisource, all'indirizzo <http://tinyurl.com/yck88qqt>.

²⁴ Si veda come esempio il *citaton graph* reperibile all'indirizzo <http://tinyurl.com/y8sy9wns> per l'articolo di Carlo Bianchini e Mirna Willer «ISBD Resource and Its Description in the Context of the Semantic Web» (con identificatore *Q56752722* in Wikidata) elaborato tramite il query service di Wikidata dal tool Scholia

uno scenario, questo descritto, che faciliterebbe l'analisi del contenuto di SHARE con lo scopo *aggiungere contesto oltre le collezioni*²⁶.

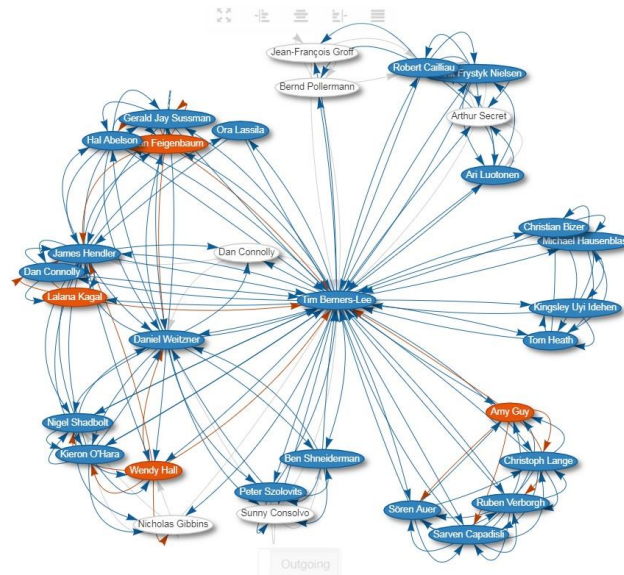


Fig. 6 Grafo dei co-autori per Tim Berners Lee (si veda la query <http://tinyurl.com/yb9bhol4>)

Infine, oltre a fornirci l'orizzonte evolutivo, l'interazione con Wikidata ci consegna una utile prospettiva di miglioramento della *Cluster Knowledge Base* di SHARE Catalogue, facilitando l'attività che potremmo definire di "manutenzione ordinaria" orientata a migliorare la qualità dei dati, grazie ai report periodicamente aggiornati relativi alle violazioni di vincolo del database, utilissimi per conoscere omonimie e duplicazioni dell'authority (fig. 7).

(<https://tools.wmflabs.org/scholia/>). Su Scholia e i suoi potenziali usi per la ricerca accademica si veda la sperimentazione descritta in Mairelys Lemus-Rojas e Jere Odell, «Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project Using Wikidata and Scholia», *Journal of Librarianship and Scholarly Communication* 6, n. 1. A proposito di Scholia cfr. anche Finn Aarup Nielsen, Daniel Mietchen, e Egon Willighagen, «Scholia, Scientometrics and Wikidata», in *European Semantic Web Conference*, (Springer, 2017), 237-259.

²⁵ Una mappa dei luoghi di nascita dei filologi di cittadinanza italiana in Wikidata con un valore per la proprietà identificativo SHARE Catalogue è visualizzabile all'indirizzo <http://tinyurl.com/yb97w9m7>.

²⁶ Cfr. Alex Stinson, «Wikidata in Collections: Building a Universal Language for Connecting GLAM Catalogs», *Medium* (blog), 13 dicembre 2017.

"Single value" violations [modifica]

Violations count: 22 (+ 3 exceptions)

- Julius Plücker (Q57561): 58213 , 349962 
- Victor Glushkov (Q93031): 104421 , 47224 
- Marvin Minsky (Q204815): 25951 , 49347 
- Su Buqing (Q323594): 59153 , 45407 
- Albert Lautman (Q327198): 48507 , 47664 
- Raimon Panikkar (Q457950): 185983 , 49719 
- Elia Millosevich (Q558995): 56002 , 307640 
- Villa Hügel (Q674670): 400311 , 401158 
- Il'ja Pjateckij-Šapiro (Q718045): 56905 , 47125 
- Boris Dubrovin (Q893614): 66740 , 52477 
- David Landes (Q931541): 11063 , 437839 
- Plan Commission (Q946520): 264146 , 374790 
- Edward Brian Davies (Q1273470): 51945 , 71552 
- Hans Ankum (Q1500441): 204529 , 255006 
- Raoul Gatto (Q1844650): 45932 , 47109 
- Rhode Island School of Design Museum (Q2148186): 208893 , 218896 
- Ibn Sab'in (Q2719051): 624175 , 662437 
- New York University Institute of Fine Arts (Q3339070): 213881 , 217432 
- Pseudo-Callistene (Q3409584): 189854 , 534255 
- Umberto Bartocci (Q4003325): 52881 , 464153 
- Brian Smith (Q4963278): 61336 , 48548 
- Simon Gindikin (Q7518846): 104875 , 57833 

Fig. 7 Report aggiornato al 6 febbraio 2019 delle *Single value violations* per la proprietà SHARE Catalogue author ID (<https://tinyurl.com/y7zg6fv4>).

Conclusioni

L'integrazione dei dati fra SHARE Catalogue e Wikidata rappresenta un modello utile e riproducibile di cooperazione, tanto che un processo analogo è stato recentemente avviato anche per SHARE-VDE, per il quale è stata formulata con successo una *proposal* simile per la creazione di una specifica proprietà²⁷. Ma il lavoro effettuato con SHARE Catalogue, con il portato dei possibili scenari evolutivi nella visualizzazione contestualizzata e dinamica delle informazioni nel front-end, così come la sperimentazione appena iniziata con SHARE-VDE, sono solo una delle possibili soluzioni per approfondire l'interazione con Wikidata.

L'aspetto forse più interessante di questo progetto di integrazione di una base di conoscenza come Wikidata, gestita e arricchita da una comunità aperta, e una base di dati così fortemente e autorevolmente costruita in un ambiente bibliotecario (la CKB prodotta dai progetti SHARE) è proprio la commistione di ambienti molto differenti, in quanto ad origine, ma che nel progetto si arricchiscono vicendevolmente.

Un progetto, quello descritto, in cui il modello di cooperazione di riferimento non è più esclusivamente fra istituzioni simili per storia e missione ma che, partendo da quel modello consolidato, si apre a positive forme di contaminazione che investono il ruolo delle istituzioni come quello dei professionisti, chiamati a sviluppare e condividere approcci e pratiche lavorative in trasformazione²⁸ che si muovono secondo i modelli di riferimento «dell'apertura, della convergenza e

²⁷ La *property proposal* per SHARE-VDE è reperibile all'indirizzo https://www.wikidata.org/wiki/Wikidata:Property_proposal/SHARE-VDE_author_ID.

²⁸ «Fare il bibliotecario oggi, in ogni caso, significa far esplodere la propria sfera di azione e soprattutto pensare in prospettiva: la routine quotidiana si svolge anche in funzione della partecipazione a una rete e, in questo modo, i risultati possono essere più autorevoli e più visibili. La parola chiave deve essere "contaminazione": più ci si apre e si guarda oltre i confini tradizionali della professione, accogliendo stimoli esterni verso nuovi

dell'integrazione di risorse e dati culturali eterogenei»²⁹ per i quali, moltiplicandosi le esperienze, va definendosi un quadro teorico applicabile in contesti differenziati.

Centrale per questi modelli di riferimento sono i linked data come tecnologia, proprio perché 'linguaggio' comune e globale: è grazie a questo framework teorico-pratico in divenire, applicabile al progetto descritto in questo contributo, che emergono forme ibride di reciprocità tra soggetti istituzionali e comunità strutturate (nella condivisione delle regole, negli strumenti utilizzati, nelle modalità di sviluppo tecnologico) ma gerarchicamente libere e informali, come quelle dei progetti Wikimedia.

Interessanti sono gli scenari evolutivi delle possibili forme di cooperazione, in tempi in cui il paradigma e le tecnologie dei linked data suggeriscono un passaggio dalla sola *conversione* e *pubblicazione* dei dati alla *creazione* nativa in LOD. Analogamente, il processo di condivisione dei progetti e delle fonti va evolvendo da un iniziale riuso dei dati con la finalità di arricchire l'esperienza della ricerca ad una forma di collaborazione ben più integrata e nativa. In questa direzione va uno dei casi d'uso che il progetto SHARE-Wikidata sta definendo e disegnando, che prevede la possibilità di creare e gestire entità del tipo *Agent* (Persone/Famiglie/Enti) e *Work* contemporaneamente sulla base dati di Wikidata e sulla Cluster Knowledge Base di SHARE: in questa modalità, lo sforzo di analisi e produzione di informazioni autorevoli, che caratterizza il lavoro di ricerca e catalogazione in ambiente LAM, può essere automaticamente trasferito in un ambito tradizionalmente meno controllato, aumentando la qualità del dato; allo stesso modo, le comunità LAM, tendenzialmente più ristrette e autoreferenziali, potranno estendere l'ambito della cooperazione a soggetti esterni e distribuiti, arricchendo e potenziando il gruppo di lavoro e la tipologia e varietà di informazioni. Questa più larga e condivisa integrazione delle esperienze e dei dati esprime bene il valore dell'eterogeneità dei contesti e delle esperienze, in cui ciascun agente non può che arricchirsi del contributo di altri agenti, ancor più se diversi e lontani. L'abbattimento dei silos della conoscenza avviene, così, anche attraverso l'incontro, il riconoscimento e la contaminazione con ambienti così (apparentemente) diversi e la definizione di cicli e modalità di lavoro che possano trarre e offrire il meglio da ciascun contesto e da ciascuna diversa esperienza.

approcci, più si dà profondità al proprio lavoro, si è portati al confronto e quindi a un aggiornamento costante, garantendo un servizio migliore. Chiara Cauzzi et al., «Conoscersi per riconoscersi: la partecipazione come specchio del bibliotecario», *AIB studi* 56, n. 2.

²⁹ Luigi Catalani, «La convergenza delle risorse informative nel Web dei dati e delle relazioni (note sul Convegno AIB CILW 2016, Biblioteca nazionale centrale di Roma, 21 ottobre 2016)», *AIB studi* 57, n. 1.

Bibliografia

- Allison-Cassin, Stacy, e Dan Scott. «Wikidata: a platform for your library's linked open data». *The Code4Lib Journal*, n. 40 (4 maggio 2018). <http://journal.code4lib.org/articles/13424>.
- Boccone, Alessandra, Claudio Forziati, Tania Maio, e Remo Rivelli. «Valorizzazione dei fondi privati in una biblioteca accademica: divulgazione, comunicazione, ricerca». *Bibliothecae.it* 6, n. 2 (29 dicembre 2017): 255–84. <https://doi.org/10.6092/issn.2283-9364/7703>.
- Catalani, Luigi. «La convergenza delle risorse informative nel Web dei dati e delle relazioni (note sul Convegno AIB CILW 2016, Biblioteca nazionale centrale di Roma, 21 ottobre 2016)». *AIB studi* 57, n. 1 (2 maggio 2017). <https://doi.org/10.2426/aibstudi-11614>.
- Cauzzi, Chiara, Valeria De Francesca, Lisa Longhi, e Federica Viazzi. «Conoscersi per riconoscersi: la partecipazione come specchio del bibliotecario». *AIB studi* 56, n. 2 (16 luglio 2016). <https://doi.org/10.2426/aibstudi-11462>.
- Erxleben, Fredo, Michael Günther, Markus Krötzsch, Julian Mendez, e Denny Vrandečić. «Introducing Wikidata to the Linked Data Web». In *The Semantic Web – ISWC 2014*, 50–65. Lecture Notes in Computer Science. Springer, Cham, 2014. https://doi.org/10.1007/978-3-319-11964-9_4.
- Forziati, Claudio, e Valeria Lo Castro. «La connessione tra i dati delle biblioteche e il coinvolgimento della comunità: il progetto SHARE Catalogue-Wikidata». *JLIS.it* 9, n. 3 (15 settembre 2018): 109–20. <https://doi.org/10.4403/jlis.it-12488>.
- Lemus-Rojas, Mairelys, e Jere Odell. «Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project Using Wikidata and Scholia». *Journal of Librarianship and Scholarly Communication* 6, n. 1 (10 dicembre 2018): eP2272. <https://doi.org/10.7710/2162-3309.2272>.
- Martinelli, Luca. «Wikidata: la soluzione wikimediana ai linked open data». *AIB studi* 56, n. 1 (2 marzo 2016). <https://doi.org/10.2426/aibstudi-11434>.
- Nielsen, Finn Aarup, Daniel Mietchen, e Egon Willighagen. «Scholia, Scientometrics and Wikidata». In *European Semantic Web Conference*, 237–259. Springer, 2017. https://doi.org/10.1007/978-3-319-70407-4_36
- Stinson, Alex. «Wikidata in Collections: Building a Universal Language for Connecting GLAM Catalogs». *Medium* (blog), 13 dicembre 2017. <https://medium.com/freely-sharing-the-sum-of-all-knowledge/wikidata-in-collections-building-a-universal-language-for-connecting-glam-catalogs-59b14aa3214c>.