

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

Camillo Carlo Pellizzari di San Girolamo^(a)

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

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

ABSTRACT

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

KEYWORDS

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

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

ABSTRACT

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

PAROLE CHIAVE

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

1. Introduzione

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

2. Storia della riconciliazione con Wikidata

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

2.1 Nascita e stabilizzazione della P396 (2013-2014)

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

¹ <https://www.wikidata.org/>.

² <https://opac.sbn.it/>.

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

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

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

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

⁷ <https://viaf.org/>.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.2 Evoluzione della riconciliazione (2015-2021)

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

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

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

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

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

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

²⁷ Es. <http://id.sbn.it/BVE0391786>.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

⁴⁴ <https://nuovo-opac.sbn.it/>.

⁴⁵ <https://opac.sbn.it/opere>.

⁴⁶ <https://opac.sbn.it/luoghi>.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

⁶⁶ Esemplicando con il VID CFIV103849:

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

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

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

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

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

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

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

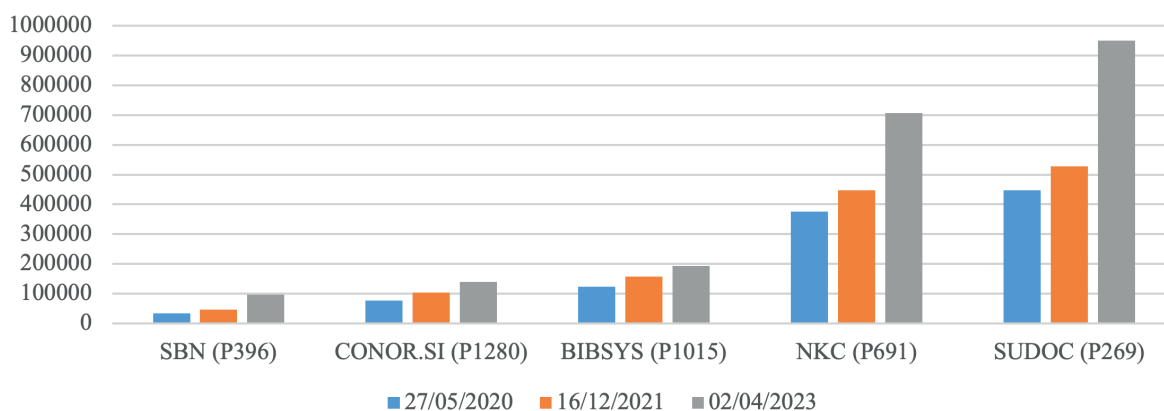


Tabella 1: numero di valori presenti in Wikidata (come dichiarazioni) per cinque proprietà di authority file europei membri del VIAF (fonte: template Number of main statements by property)

3. Conclusioni

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

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

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

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

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

Riferimenti bibliografici

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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