JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

JLIS.it

Wikidata and authority files: reconciliation and cooperation procedures

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: 06 August 2025; **Accepted:** 19 February 2026; **First Published:** 15 May 2026

ABSTRACT
This article aims to providing a comprehensive overview of the possible interactions between Wikidata and authority files, which can be classified into two main categories: firstly, the methods of reconciliation between Wikidata items and authority records; secondly, the ways in which the Wikidata community can improve Wikidata items using the authority records to which they link, and in which the cataloguers can improve their authority records using the Wikidata items that link to them, in a collaborative perspective. The purpose of this overview is optimising existing workflows and encouraging to establish new ones, so as to foster the reciprocal improvement of Wikidata's and authority files' data. The last part of the article describes a few major collaborations between Wikidata and authority files of national and international relevance, and the main attempts to use Wikibase (Wikidata's software) as a platform to store authority files.

KEYWORDS
Wikidata; Authority control; Entity reconciliation; AuthorityBox; Wikibase.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

1. Introduction

1.1 Historical background

The cooperation between Wikidata (from now onwards WD) and authority files (from now onwards AFs, or singular AF) started in 2013: on March 6, the first properties with datatype "string" (afterwards converted to "external-id") were created in WD, specifically P212 for ISBN-13, P213 for ISNI, and P214 for VIAF.
In the same year the first significant reconciliation of authority records (from now onwards ARs, or singular AR) to WD was performed as part of a project by Maximilian Klein, Wikipedian in Residence at OCLC, and Andrew Gray, Wikipedian in Residence at the British Library: VIAFbot, first proposed in June 2012, was initially programmed to import to English Wikipedia articles the links to the VIAF clusters that already linked to them, but since March 2013 it was adapted to operate on WD (Klein and Kyrios 2013).
In 2016 the International Federation of Library Associations and Institutions (IFLA) endorsed a white paper acknowledging that «Wikidata is becoming an important tool for synchronizing across identifiers like Virtual International Authority File (VIAF) and ORCID identifiers» (Bartholmei et al. 2016, 10).
In 2019 the Association of Research Libraries (ARL) published a white paper regarding WD, in which a paragraph dealt with «Authority Data and Using Wikidata as a Linking Hub». The white paper encouraged the collaboration between WD and AFs, stating that «Wikidata has the potential to be an important part of the linked data ecosystem for authorities» (Association of Research Libraries 2019, 29). AFs are crucial for WD to connect to the bibliographic resources authored by or containing information about the entities it describes, and they are also valuable sources for the structured data about these entities (e.g. birth and death dates). At the same time, WD can provide to libraries persistent identifiers (PIDs) and a vast amount of data (both factual data and external IDs) that they can use to enrich their ARs and, more generally, their catalogues, providing more context to the users.
In the same period many groups were created to explore the use of WD in GLAM institutions, and especially in libraries, including the LD4 Wikidata Affinity Group[1] in 2019 and the Gruppo Wikidata per Musei, Archivi e Biblioteche (GWMAB)[2] in 2020; a Wikidata Working Group in IFLA has been active from January 2020 to December 2022;[3] in 2020-1 the Program for Cooperative Cataloging (PCC) sponsored a PCC Wikidata Pilot[4] to guide librarians in experimenting the use of WD (cf. Zhang et al. 2023). The above list is not exhaustive and focuses on projects having authority control among their main focuses (cf. Boccone 2022).
Systematic reviews of WD's use by libraries (Tharani 2021) and, more broadly, by GLAM (Galleries, Libraries, Archives, and Museums) institutions (Candela et al. 2024) both outlined a growing interest for WD. In the specific case of libraries, «Integration and interoperability of authority

[1] https://www.wikidata.org/wiki/Wikidata:WikiProject_LD4_Wikidata_Affinity_Group.
[2] https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche.
[3] https://www.wikidata.org/wiki/Wikidata:IFLA_Wikidata_Working_Group.
[4] https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

data using Wikidata emerged as the most popular application area based on the number of articles coded for this category» (Tharani 2021, 3).
The importance of WD in authority control has been further highlighted through comparison with VIAF (Bianchini et al. 2021) and in the perspective of Universal Bibliographic Control (Sardo and Bianchini 2022); WD also allows to extend the concept of authority control to researchers, establishing a connection between library catalogues and research repositories (Bargioni et al. 2023). In many digital humanities projects WD itself is currently used as an AF (Zhao 2023, 868); the procedures for creating personal WD items and personal ARs have also been compared, tracing similarities and differences (Chen and Yuxuan 2024).
Some major collaborations between AFs and WD are described in chapter 5.1.

1.2 Statistics about the present situation

In WD, library AFs constitute a small but significant group of properties: as of now the properties that are "instance of" (P31) "Wikidata property for authority control (libraries)" (Q96776953) or of one its subclasses, which include mainly "Wikidata property for authority control by VIAF member" (Q55586529), are 220 in total,[5] out of 9,569 external-id properties.[6] These properties are used as value in 22.8 M statements[7] contained in 5.6 M unique items,[8] the majority of them (4.2 M) being personal items (i.e. having P31 = Q5, "instance of = human");[9] restricting the queries to VIAF members only, they are used as value in 18.7 M statements[10] contained in 4.6 M unique items[11] (including 3.7 M personal items;[12] as a comparison, in 2020 VIAF contained 22.1 M personal clusters[13]).

[5] https://qlever.dev/wikidata/LI6Z15.
[6] https://qlever.dev/wikidata/4avd1b.
[7] https://qlever.dev/wikidata/wR9jD3.
[8] https://qlever.dev/wikidata/lNhZmi.
[9] https://qlever.dev/wikidata/ZwukeP.
[10] https://qlever.dev/wikidata/DrgpXS.
[11] https://qlever.dev/wikidata/oTzGAR.
[12] https://qlever.dev/wikidata/k0ZRWv.
[13] https://catalogo.pusc.it/beyond_viaf/.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

Among the aforementioned 220 properties, the ten used as values in the highest number of distinct items are:

Table 1. The ten properties linking to library AFs used as values in the highest number of distinct items. The number of distinct items using the properties comes from a SPARQL query,[14] whilst the number of users having added them at least once as main statement (from their creation and in the month of June 2025) comes from Wikidata Navel Gazer[15]

property	VIAF code	items (2024-10-12)	items (2025-07-21)	users (total)	users (June 2025)
P214	(VIAF)	3,444,466	3,649,450	10,439	615
P227	DNB	1,843,039	2,426,148	5,314	318
P213	ISNI	1,699,662	2,216,034	6,840	424
P244	LC	1,635,913	1,686,829	6,750	427
P269	SUDOC	982,160	1,007,193	3,302	255
P8189	J9U	823,439	835,048	1,304	127
P691	NKC	783,355	813,295	1,420	70
P268	BNF	701,258	723,772	4,019	298
P1207	NUKAT	632,937	687,557	1,645	155
P1006	NTA	559,425	576,486	1,649	152

2. Reconciliation

The premise for a comparison between the data in WD and the data in an AF is that WD items and ARs should have been reconciled firstly.

2.1 Interlinking

WD items can link to ARs through apposite properties: if an AF is publicly accessible and assigns a URI to each AR, an external-id property can be proposed[16] and, after a community discussion of minimum 7 days, it is created if there is consensus (i.e. if no objection is raised, or when objections have been adequately answered).
ARs can link to WD items through the 024 field ("other standard identifier") in MARC 21 format or the 017 field ("other identifier") in UNIMARC format; in both cases $a is used for the ID and $2 for the source indication "Wikidata", or a chosen abbreviation.
In principle, the correspondence between ARs and WD items should be biunivocal (1:1), but exceptions are possible, e.g. in the case of pseudonyms: WD usually has one item for the author as a person and for all their pseudonyms (but exceptions are made wherever one Wikimedia project has

[14] https://qlever.dev/wikidata/Uupogc.
[15] https://bambots.brucemyers.com/NavelGazer.php.
[16] https://www.wikidata.org/wiki/Wikidata:Property_proposal.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

distinct pages for the person and for some of their pseudonyms); similarly, certain cataloguing rules prescribe having only one AR for each physical person (e.g. REICAT, used by SBN), whilst others (e.g. RDA) prescribe creating distinct ARs for the person and for each of their pseudonyms if the person has used both their real name and one or more pseudonyms in their works (De Monaco 2025, 7–21; Chou 2025). In cases like this the WD property can apply to the "single-value constraint" (Q19474404) the qualifier "separator" (P4155), which allows to define specific qualifier-properties to be used as separators, in order to mark manually the cases in which the non-biunivocal correspondence is intentional and thus does not constitute a mistake in WD and/or in the AF.[17]
Apart from exceptions, the other cases of non-biunivocal correspondence between ARs and WD items are usually identification issues on one or both sides; undertaking the reconciliation of an AF with WD is usually a perfect occasion to find them (De Monaco 2025, 49–51; De Monaco and Pellizzari di San Girolamo 2025, file 'Tracciamento errori riscontrati in SBN.xlsx').

2.2 Available tools

WD items and ARs can be reconciled manually, adding in an AR a link to a WD item through the 024 MARC 21 field or the 017 UNIMARC field and adding in a WD item the link to an AR through its apposite property. This could be a standard practice for the cataloguers while manually creating new ARs or editing existing ARs; if the link is manually created in one of the two ways, an automatic system can be established to periodically add the reciprocal link (see e.g. NKC in chapter 5.1).
To import to WD a pre-existing list of matches between WD items and ARs, QuickStatements (from now onwards QS) is the most convenient choice;[18] the tool can be used by autoconfirmed users[19] and allows to enter these matches with a CSV syntax at a rate of usually many tens per minute.
Matching in a completely manual way with WD an AF consisting of tens or hundreds of thousands of ARs is obviously not feasible. For this reason, in order to match old ARs with WD items, two reconciliation tools can be used: Mix'n'match (from now onwards MnM)[20] and OpenRefine (from now onwards OR)[21]. Both these tools, given a set of entries, try to match them with WD, suggesting one or more matches whenever possible for each entry and allowing to create new WD items for the others; the user can always revise these judgements, confirming or refusing them.
Apart from these similarities, MnM and OR have many differences. MnM is an online platform used to reconcile WD items to other (usually online) datasets: the datasets are uploaded as catalogs,[22]

[17] See e.g. https://www.wikidata.org/wiki/Property:P1368#P2302.
[18] https://quickstatements.toolforge.org/ (manual: https://www.wikidata.org/wiki/Help:QuickStatements). Since 2024, a new version of QS with additional features is also available, QS 3.0 (https://qs-dev.toolforge.org/; manual: https://meta.wikimedia.org/wiki/QuickStatements_3.0/Documentation).
[19] https://www.wikidata.org/wiki/Wikidata:Autoconfirmed_users.
[20] https://mix-n-match.toolforge.org/ (the manual, https://meta.wikimedia.org/wiki/Mix%27n%27match/Manual, is significantly out-of-date as of now).
[21] Software downloadable from https://openrefine.org/, also containing documentation; a manual in WD is also available (https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine).
[22] E.g. https://mix-n-match.toolforge.org/#/catalog/4129.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

which can be either scraped through an apposite MnM function, or uploaded through a CSV/TSV file or a public Google Spreadsheet; MnM catalogs are public, so all WD users can cooperate to their reconciliation (MnM can also be used directly inside WD items thanks to an apposite gadget[23]); a significant strength of MnM is containing thousands of catalogs, so it is a very good tool to gain an understanding of e.g. conflicting homonyms among many AFs and other sources, and it also allows to create new items combining data from many sources. OR is a software, allowing to clean-up datasets and to reconcile them with WD items; its reconciliation functions and its capability of creating high-quality new items (e.g. with statements containing qualifiers and elaborate references) are more sophisticated than those of MnM: however, in OR a single user works on a single dataset, without the possibility of crossing one's own dataset with many other catalogs as in MnM.
Consequently, MnM tends to be more effective for projects where it is useful to have a reconciliation performed on big amounts of data by potentially many users in a long span of time, and in cases where a majority of the entries already exist as WD items, whilst OR is surely more suitable for smaller projects managed by one user, and in cases requiring the majority of entries to be created as new WD items.

3. Identification issues in authority records and Wikidata items

When reconciling ARs and WD items, their quality plays a significant role in making the process easy or difficult, sometimes impossible. In assessing the quality of an entry (AR or WD item), three parameters are taken into account:

1. identification: the entry is clearly self-identified,[24] not self-identified[25] or confusedly self-identified;[26]
2. incoming links: the links to the entry are only appropriate, or partly inappropriate (note: the distinction appropriate/inappropriate is based on the entry being clearly self-identified; if it is not self-identified, the check should instead establish if the incoming links relate to one entity or 2(+) entities);
3. duplicates: the entry has no duplicates, or 1(+) duplicates.

Parameters 1 and 3 apply to both ARs and WD items; parameter 2 usually applies to ARs, since an AR usually exists if it is connected to at least one bibliographic record (from now onwards BR, or plural BRs), but rarely to WD items, since a WD item can exist without any incoming link from other WD items.
On the basis of these parameters, the following table (Table 2) classifies the identification issues that can affect ARs and WD items.

[23] https://www.wikidata.org/wiki/User:Magnus_Manske/mixnmatch_gadget.js.
[24] The entry contains sufficient information to make the described entity clearly identifiable: usually the dates of birth and death, or the dates of activity, and the occupation(s).
[25] The entry contains insufficient information to make the described entity clearly identifiable (in the most common case, it contains only a name).
[26] The entry contains incoherent information, with some data pointing to an entity A and some other to an entity B, so that it is unclear if the entry describes A or B.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

Table 2. Table of possible identification issues in ARs and WD items

Situation code and name	Description in ARs	Description in WD items
S0: "no issue"	clearly self-identified AR with all the appropriate BRs connected	clearly self-identified WD item with all the appropriate links from other WD items
S1: "wrong connections"	clearly self-identified AR but with 1(+) inappropriate BR(s) connected	clearly self-identified WD item but with 1(+) inappropriate link(s) from other WD items
S2: "weakly identified"	not self-identified AR; identification rests upon the connected BR(s), which relate to only one entity	(theoretically impossible in WD: according to the notability criteria,[27] items must always be "clearly identifiable" in themselves; if they are not, they should be fixed or deleted)
S3: "unidentified"	not self-identified AR; identification cannot rest upon the connected BR(s), since they relate to 2(+) entities	
S4: "conflated"	confusedly self-identified AR	confusedly self-identified WD item
SD: "duplicates"	2(+) ARs regarding the same entity	2(+) WD items regarding the same entity
* SD0: "duplicates, no (other) issues"	2(+) clearly self-identified ARs regarding the same entity with no inappropriate BRs connected	2(+) clearly self-identified WD items regarding the same entity with no inappropriate links from other WD items
* SD1: "duplicates, with wrong connections"	2(+) clearly self-identified ARs regarding the same entity with 1(+) inappropriate BR(s) connected	2(+) clearly self-identified WD items regarding the same entity with 1(+) inappropriate link(s) from other WD items
* SD2: "duplicates, weakly identified"	2(+) not self-identified ARs which, according to their identification based on the connected BR(s), relate to only one entity	*same as for S2 and S3*

In WD items, given that most of them do not have any links from other WD items, S1 is rare, as well as S2 and S3 (which are theoretically impossible, but may happen when items are based on sources which contain very few identifying details about them). So basically in WD the focus of the user can be mainly on the item itself, since the incoming links are mostly absent.
Conversely, ARs usually have at least one connected BR, so their identification is a twofold issue, focusing both on the AR itself and on its connected BR(s). S0 is the standard to which all ARs should point; an S0-AR degrades to S1 when 1(+) inappropriate BR(s) are connected to it, but since it contains good identifying information the inappropriate BR(s) can be easily found and removed, restoring the AR to S0. The rare situation S4 happens when identifying information is poorly assembled and mixes 2(+) entities which are in fact distinct, making the AR completely unusable. S3 is the status of a catalogue where authority control is not implemented, or poorly applied. S2 stays in the middle, i.e. a catalogue where BR(s) are correctly attributed but ARs do not contain

[27] https://www.wikidata.org/wiki/Wikidata:Notability.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

identifying information; S2 is not a mistake in itself, but can easily degrade into SD2 and/or S3 when new BR(s) are wrongly attributed due to insufficient guidance given by AR(s) themselves.

4. Five ways of reciprocal improvement

As said, both WD items and ARs can contain mistakes (i.e. cases S1, S3, S4, SD) and incomplete entries (i.e. case S2), and these mistakes are likely to emerge when someone tries to reconcile WD with an AF. However, while a random WD user can edit directly WD items to solve problems of identification, improving them to S0, the same is not true for AFs, where only the employees of the institution(s) creating it, or users authorized by the institution(s), can make edits, so their mistakes cannot be solved directly by most WD users (cf. Pellizzari di San Girolamo 2024a, para. 5.2).
This chapter explores how the collaboration between WD and AFs (i.e. between WD users and AFs' editors) can lead to increasing the percentage of both WD items and ARs reaching S0. WD, while allowing AFs to reduce their amount of mistakes, also constitutes for them a good way to increase their incoming and outcoming links (both to other AFs and to different kinds of resources), thus improving their position in the Linked Data Web.

4.1 Data round-tripping

WD users, noticing mistakes in ARs, try reporting them, in a process known in WD as "data round-tripping",[28] usually through a webform or an e-mail provided by the institution(s) managing the AF (and indicated in the property itself through P10923, "error-report URL or e-mail"), or contacting directly another WD user with editing access to it;[29] unfortunately, for some AFs there are no known mistake reporting methods (Pellizzari di San Girolamo 2025).

Table 3. Table of the usual management strategies of ARs' identification issues in WD items

Situation in AR	Strategy in WD: add the AR ID(s)
S0: "no issue"	
S1: "wrong connections"	and try reporting the wrong BR(s)
S2: "weakly identified"	
S3: "unidentified"	with deprecated rank and the qualifier P2241 ("reason for deprecated rank") = Q14946528 ("conflation"), and try reporting the conflation
S4: "conflated"	
SD0: "duplicates, no (other) issues"	
SD1: "duplicates, with wrong connections"	and try reporting the wrong BR(s)
SD2: "duplicates, weakly identified"	

[28] https://www.wikidata.org/wiki/Wikidata:Data_round-tripping.
[29] https://www.wikidata.org/wiki/Wikidata:WikiProject_Authority_control/Error_reporting_procedures.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

From WD's perspective (shown in Table 3), adding AR ID(s) is always useful, also when they do not have S0 quality, because it both makes their mistakes more evident and it helps avoiding possible mistakes by other WD users. Specifically, there is no need to send manual reports for SD0 and SD2, as these cases can be listed automatically (see chapter 4.2). S1 and SD1 do not need high-priority mistake reports, but their inappropriate BR(s) should anyway be reported as they are likely to cause confusion in other AFs and especially to lead to VIAF conflations. Manual reports are absolutely necessary only for S3 and S4, since these ARs cannot be linked from WD unless properly fixed; until they are fixed, they must be deprecated to avoid other WD users considering them reliable ARs.

4.2 Constraint violations and mismatches

WD uses constraints to assess potentially wrong statement values;[30] a constraint violation (from now onwards CV, or plural CVs) is visualized through an icon that warns users to check strange values and correct them if needed; lists of CVs can be obtained through SPARQL queries automatically generated by the Property documentation template[31] at the top of the talk page of each property.[32] In order to check identification issues of WD items and ARs, two types of CVs are particularly relevant: unique-value CVs, i.e. the same external ID linked from 2(+) WD items, and single-value CVs, i.e. the same WD item linking to 2(+) external IDs in the same database.

Table 4. Table indicating how unique-value and single-value CVs can be used to find duplications and conflations in WD items and ARs (adapted from Pellizzari di San Girolamo 2023, 14)

	WD items	ARs
duplication (= SD)	unique-value CV	single-value CV
conflation (= S3/S4)	unique-value CV single-value CV	unique-value CV single-value CV

As Table 4 shows, unique-value and single-value CVs always show mistakes in WD items, ARs or both (apart from the few cases of normal non-biunivocal correspondence between them; see chapter 2.1); when mistakes are solved, the cases disappear from the lists of CVs. However, as WD users can usually solve only mistakes in WD items, the cases depending from mistakes in ARs tend to remain in the lists of CVs for a long time and to make them more difficult to use; in particular, the high amount of SD cases in ARs tends to flood the lists of single-value CVs, making very difficult to find through them conflations in WD items and in ARs.
Since the lists of CVs are the most powerful tool for both WD users and AFs' editors to discover identification issues, monitoring and emptying them constantly is the best way to keep them readable and usable; ideally, AFs' editors should be involved as WD users in fixing the mistakes shown by the lists of CVs both in WD items and in their ARs. In 2024 the WikiProject Disambiguation

[30] https://www.wikidata.org/wiki/Help:Property_constraints_portal.
[31] https://www.wikidata.org/wiki/Template:Property_documentation.
[32] E.g. https://www.wikidata.org/wiki/Property_talk:P396.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

has been created in WD to connect AFs' editors active in WD and to make easy reporting to them the most complex cases of conflations affecting multiple AFs.[33]
A query for deprecated statements should be used to find previously-marked conflations; this query is also generated by the Property documentation template; as in the case of other queries, it is important to gradually normalize in WD the AR IDs once fixed, so that the query does not always show the same unsolved results.

4.3 Manual editing on Wikidata

SPARQL queries in WD can be used to find items already connected to a certain AF but with significant information missing (e.g. items without VIAF,[34] or without birthdate,[35] or without occupation,[36] or with conflicting birthdates,[37] or with unreferenced birthdate[38]); for AFs' editors, using these queries to find and improve these items has a lower priority in comparison with the periodical check on the lists of CVs (see chapter 4.2), but can still have two important aims: firstly, making easily available good quality authority data, so that AFs' editors in other countries can use them to improve their ARs to S0; secondly, being able to make reliable statistics about the authors described in the AF (e.g. regarding their distribution by birthplaces, deathplaces, lifespan) through WD SPARQL queries (De Monaco 2025, 60–9).

4.4 Massive enrichment from Wikidata

WD items can be used as a source to import data into an AF; the data can be copied directly into the ARs and/or displayed in real time through the API, e.g. through an AuthorityBox (Bargioni 2020). Examples of both enrichments are mentioned in chapter 5.1.
Both solutions have relevant pros and cons that should be evaluated. From a technical point of view, copying data into the ARs requires being able to massively edit ARs in the backend, whilst displaying data from WD means changing the frontend visualization of ARs. This second task could be difficult or impossible when using a proprietary ILS.
Copying the data directly into the ARs has the advantage of definitively improving the quality of ARs (given that only referenced data are taken from WD), particularly improving S2 ARs to S0, so as to avoid them degrading into S3; however, this data import is static (i.e. it does not consider further updates and/or improvements made in WD items, unless repeated periodically) and adapting WD's data to the data format used in the AF might be challenging, especially if the ARs have a low data granularity (i.e. a lot of information is expressed by a textual note, instead of being segmented in different fields for dates, places, occupation(s) etc.) and thus it is necessary to concatenate the data in readable phrases.

[33] https://www.wikidata.org/wiki/Wikidata:WikiProject_Disambiguation.
[34] E.g. https://w.wiki/Hy3i.
[35] E.g. https://w.wiki/Hy3k.
[36] E.g. https://w.wiki/Hy3n.
[37] E.g. https://w.wiki/Enmr.
[38] E.g. https://w.wiki/EnkA.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

Conversely, displaying data in real time has the advantage of using always up-to-date data, but in some cases it can lead to showing incoherent data (i.e. divergencies between what is written in the AR and what is displayed from the corresponding WD item).

4.5 Massive imports to Wikidata

Massively importing data from an AF to WD has basically the same advantages of manually editing WD items (cf. chapter 4.3), but on a bigger scale: the data in the AF become much more visible and reusable for other AFs, and easily analysable through WD SPARQL queries. Massive imports are possible only if the AF has released its data in CC0 license, a choice already made by GND (since 2015),[39] NKC (since 2022),[40] and NTA;[41] otherwise, only one-by-one manual imports are legally possible.[42] The most relevant imports of data from AFs to WD have been made by NKC (cf. chapter 5.1).

5. Some relevant collaborations

5.1 Authority files actively collaborating with Wikidata

This chapter provides brief descriptions of the most relevant ongoing collaborations between AFs and WD.

NKC (P691 in WD): the import of ARs' data into WD items, including the creation of new missing items, began in 2019, as part of a pre-existing collaboration between the National Library of Czech Republic (NLCZ) and Wikimedia Czech Republic (WMCZ); the total uses of P691 climbed from 220,732 at the end of 2018[43] to 264,736 at the end of 2019,[44] 646,660 at the end of 2020,[45] 1,656,882 at the end of 2022;[46] they are currently 2,118,160.[47] The reconciliation was made firstly importing pre-existing matches from Czech Wikipedia, then with semiautomatic methods: VIAF matches were tried but abandoned due to the excessive mistake percentage; the most effective semiautomatic method proved to be combining the name match with a date match (same birthdate, i.e. day of birth, or same year of birth and year of death, or same year of birth and same VIAF); in parallel, MnM is also used for manual reconciliation (Dostál 2021). Newly created ARs can be monitored in WD week by week[48] and, most importantly, a bot[49] imports on a daily basis

[39] https://gnd.network/Webs/gnd/EN/FAQ/faq_node.html; https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/Gesamtabzuege/gesamtabzuege_node.html.
[40] https://www.nkp.cz/en/about-us/professional-activities/open-data.
[41] https://data.bibliotheken.nl/KB/-/stories/NTA.
[42] https://www.wikidata.org/wiki/Wikidata:Data_donation.
[43] https://www.wikidata.org/w/index.php?oldid=823585223&action=edit.
[44] https://www.wikidata.org/w/index.php?oldid=1087497651&action=edit.
[45] https://www.wikidata.org/w/index.php?oldid=1332488380&action=edit.
[46] https://www.wikidata.org/w/index.php?oldid=1799601120&action=edit.
[47] https://www.wikidata.org/w/index.php?oldid=2372675275&action=edit.
[48] https://www.wikidata.org/wiki/Wikidata:WikiProject_Czech_Republic/New_authorities.
[49] https://www.wikidata.org/wiki/User:Frettiebot.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

the link to NKC ARs into WD items already linked from NKC ARs through the 024 MARC 21 field (the same procedure suggested in chapter 2.1), and also converts some ARs' fields into WD statements (mainly field 374 "Occupation" into P106 "occupation" and field 372 "Field of Activity" into P101 "field of work"). As of now, the NKC case represents the most accomplished case of routinary data synchronisation from an AF to WD.

GND (P227 in WD): GND, currently the second most-linked AF in WD (besides only VIAF), has been tightly connected to the Wikimedia community since before its birth (cf. Wikimedia Deutschland 2025): in 2005 German Wikipedia users started collecting mistake reports for PND, one of the AFs merged together in 2012 to create the GND; since 2016 some of these users have been provided editing access to GND with *Katalogisierungslevel* "Tp4";[50] after twenty years, German Wikipedia users are still active in connecting their articles to GND (708,640 articles link to GND as of now[51]) and in sending monthly lists of mistake reports to the Bavarian State Library (BSB) for people died in 1850 or earlier and to the German National Library (DNB) for people died after 1850 or still alive;[52] unfortunately, the backlog of reports yet to be checked is of three years for BSB and one year for DNB as of now.[53]

LCNAF (P244 in WD): since 2023 a bot[54] has been periodically adding the link to LCNAF ARs into WD items already linked from LCNAF ARs through the 024 and/or 670 MARC 21 fields, and generating an internal report that allows the bot maintainers to find and correct some mistakes in LCNAF ARs themselves (Muratsubaki Campany et al. 2024); the Python code of the bot is publicly available on GitHub.[55]

SBN (P396 in WD): the Central Institute for the Union Catalogue of Italian Libraries and Bibliographic Information (ICCU), the institution coordinating the SBN network, has had a collaboration agreement with Wikimedia Italia (WMIT) since 2015; since the renewal of OPAC SBN's interface in 2021, the joint effort of Wikidata users and ICCU to improve the reconciliation of WD items and SBN AFs has intensified, especially thanks to ICCU's decision to make visible the vast majority of ARs (Pellizzari di San Girolamo 2024c). Since 2023 the collaboration has added to its focuses the massive enrichment of SBN ARs with data from WD so as to bring them from S2 to S0; a first group of ca. 16200 ARs has been enriched on July 29, 2024 (Ravelli 2024); moreover, new reconciliation methods, both manual and semiautomatic, have brought a significant growth of P396 values in WD, from 132,946 at the end of 2023[56] to 193,035 as of now.[57] The collaboration is regularly documented in WD.[58]

[50] https://de.wikipedia.org/wiki/Hilfe:GND.
[51] https://de.wikipedia.org/w/index.php?oldid=258099995.
[52] https://de.wikipedia.org/wiki/Wikipedia:GND/Fehlermeldung.
[53] https://de.wikipedia.org/w/index.php?oldid=257735082.
[54] https://www.wikidata.org/wiki/User:LccnBot.
[55] https://github.com/lcnetdev/lccn-wikidata-bot.
[56] https://www.wikidata.org/w/index.php?oldid=2038639439&action=edit.
[57] https://www.wikidata.org/w/index.php?oldid=2372669213&action=edit.
[58] https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/SBN.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

Parsifal (P12458 in WD): the AFs of some libraries of the network Unione Romana Biblioteche Ecclesiastiche (URBE) have started being reconciled with WD in 2018; since 2019 a growing number of cataloguers have been directly involved in editing WD, reconciling ARs with WD items through MnM and using the lists of CVs to find and correct mistakes in ARs (Pellizzari di San Girolamo 2024b). Since the creation of the collective catalog of URBE libraries, Parsifal, in 2022, the involvement of librarians in WD has further increased, as shown by a learning day on May 14, 2024[59] and a meeting of 30+ librarians on July 3, 2025.[60] The collaboration is regularly documented in WD.[61]

5.2 Authority files using Wikibase

Wikibase is the open-source software used by Wikidata;[62] it consists of a set of MediaWiki extensions that allow to collect structured data in multiple languages with a flexible data model.
The possible use of Wikibase as a platform to store an entire AF has long been investigated. In 2019 OCLC published a report regarding a Wikibase prototype used by librarians to create linked data to describe resources (Godby et al. 2019). In 2020 the DNB in collaboration with Wikimedia Deutschland (WMDE) started the development of "a Wikibase-based GND"[63] and, in the same year, the French National Library (BNF) and the Agence bibliographique de l'enseignement supérieur (ABES) recommended to use Wikibase as the platform for the development of the Fichier National d'Entités (FNE), a new AF merging the two pre-existing ones, BNF and IDREF, as part of the Transition bibliographique project.[64] However, the FNE project has been terminated in 2023, judging Wikibase an unsuitable solution, mainly due to the insufficient speed of massive data upload via API,[65] an issue which has not been overcome yet as of now.[66] The project of a Wikibase-based GND is still ongoing as of now;[67] in 2024, however, the final aim of having the 10 M GND entities in two distinct but synchronised platforms, the present one and the Wikibase instance, was perceived as «far away», and «bulk import and export of data» was among the *desiderata*; presently Wikibase is being used mainly to transform GND documentation into structured data (Fischer 2024). OCLC used Wikibase also in the CONTENTdm Pilot Project (Bahnemann et al. 2021), but finally discarded it for the development of the Shared Entity Management Infrastructure, mainly due to the aforementioned low speed of massive data upload (Phillips and

[59] https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/URBE/Formazione.
[60] https://en.pusc.it/article/wikilib-bibliotecari-romani-e-vaticani; https://www.iccu.sbn.it/it/eventi-novita/novita/LIC-CU-alla-giornata-di-WikiLib/.
[61] https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Parsifal.
[62] https://wikiba.se/; https://www.mediawiki.org/wiki/Wikibase.
[63] https://wiki.dnb.de/spaces/GND/blog/2020/03/07/167019461/Report+GND+meets+Wikibase+2.
[64] https://web.archive.org/web/20200715115414/https://www.transition-bibliographique.fr/wp-content/uploads/2020/07/synthese-preuve-concept-fne.pdf.
[65] https://web.archive.org/web/20231008163559/https://www.transition-bibliographique.fr/2023-09-26-suspension-du-projet-fne-et-perspectives-nouvelles/.
[66] https://phabricator.wikimedia.org/T287164.
[67] https://web.archive.org/web/20250210173054/https://gnd.network/Webs/gnd/DE/Projekte/projekte_node.html.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

Washington 2022, 46); in 2022 the project has been officially launched[68] as WorldCat Entities.[69] Despite this, there are ongoing projects using Wikibase to store authority data.

Czech Republic: a Wikibase instance was created by WMCZ in 2019 with a part of the NKC AF as an intermediate step before importing data into WD (cf. chapter 5.1).[70] Further use of Wikibase for other NLCZ databases is one of the future possible areas of collaboration between NLCZ and WMCZ (Jansová et al. 2024, 8, 15); e.g. in 2024–5 TDKIV (the Czech Explanatory Terminology Database of Library and Information Science), one of the controlled vocabularies curated by NLCZ, has been converted into a Wikibase instance.[71]

Wales: in 2022 the Semantic Name Authority Repository Cymru (SNARC) has been "established in order to provide a central hub for name authority records relating to Wales and in the Welsh language";[72] this Wikibase instance has been created and mainly edited by Jason Evans, National Wikimedian at the National Library of Wales (NLW) since 2017,[73] and has currently more than 139 thousand items.

Greece: the National Library of Greece (NLG) has worked on a proof of concept using a Wikibase instance as backend to store authority data in RDA/RDF, accessible only to NLG staff, and a public platform[74] as frontend to visualize data in a user-friendly way and to query them;[75] the experiment was successful and the project continues to be developed (Zapounidou et al. 2024).

Nigeria: in 2025 the Wikibase instance National Library of Nigeria Semantic Name Authority Repository (NLN SNAR) has been created as part of the Integrating Wikimedia Projects into African Libraries' Ecosystem project (IWIPALE Project) of African Library and Information Associations and Institutions (AfLIA), supported by the Knowledge Equity Fund of Wikimedia Foundation (WMF) and implemented by the National Library of Nigeria (NLN) and WMDE; it has currently more than 3 thousand items.[76]

6. Conclusions

WD, given its crucial importance in the linked data ecosystem, has become increasingly important for AFs, both for providing PIDs to which ARs can link and for extracting data to enrich ARs; at the same time, WD constantly uses AFs as reliable sources for its data and as authoritative IDs

[68] https://www.oclc.org/en/worldcat/entities.html.
[69] https://entities.oclc.org/worldcat/entity.
[70] https://autority.wikimedia.cz/.
[71] https://tdkiv.wikibase.cloud/.
[72] https://snarc-llgc.wikibase.cloud/.
[73] https://www.library.wales/about-nlw/press-and-news/press-releases/2017-press-releases/the-national-library-of-wales-appoint-uks-first-permanent-wikimedian.
[74] https://data.nlg.gr/.
[75] https://data.nlg.gr/query.
[76] https://nlnsnar.wikibase.cloud/.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

for disambiguating its entities (chapter 1). Reconciliation tools, especially MnM and OR, simplify the process of matching ARs and WD items (chapter 2). However, quality issues on both sides can sometimes make it difficult to establish matches: if the identity of the subject of an AR and/or an item is unclear, a match can be established only after it has been clarified (chapter 3). Establishing good procedures to receive mistake reports is crucial for AFs to improve their connection with WD; cataloguers can use WD lists of CVs and self-made queries to find mistakes in both WD items and their ARs; data from WD can be reused in AFs either displaying them through the API or copying them directly into the ARs; vice versa, AFs can also massively import their data into WD, provided that they are released in CC0 license (chapter 4). Relevant collaborations between AFs and WD already exist in Czech Republic, Germany, US, and Italy; the use of Wikibase, WD's software, to store entire AFs has been experimented with mixed results, but presently promising projects exist in Czech Republic, Wales, Greece, and Nigeria (chapter 5). The main challenges in the next years will be involving more cataloguers in WD editing, establishing stable ways of synchronising AFs and WD (including accurate checks on non-biunivocal links) and spreading the use of CC0 license among AFs, so that more ARs can be imported to WD, improving its coverage of the persons described in AFs across the world.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

References*

Association of Research Libraries. 2019. 'ARL White Paper on Wikidata: Opportunities and Recommendations'. https://www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf.

Bahnemann, Greta, Michael Carroll, Paul Clough, Mario Einaudi, Chatham Ewing, Jeff Mixter, Jason Roy, Holly Tomren, Bruce Washburn, and Elliot Williams. 2021. *Transforming Metadata into Linked Data to Improve Digital Collection Discoverability: A CONTENTdm Pilot Project*. OH: OCLC Research. https://doi.org/10.25333/fzcv-0851.

Bargioni, Stefano. 2020. 'From Authority Enrichment to AuthorityBox : Applying RDA in a Koha Environment.' *JLIS.it* 11 (1): 175–89. https://doi.org/10.4403/jlis.it-12595.

Bargioni, Stefano, Carlo Bianchini, and Camillo Carlo Pellizzari di San Girolamo. 2023. 'IRIS e Wikidata: Un progetto per una migliore valorizzazione e fruizione degli archivi della ricerca scientifica italiana.' In *Guardando oltre i confini. Partire dalla tradizione per costruire il futuro delle biblioteche. Studi e testimonianze per i 70 anni di Mauro Guerrini*, edited by Giovanni Bergamin and Tiziana Possemato, 31–45. Roma: Associazione italiana biblioteche.

Bartholmei, Stephan, Rachel Franks, James Heilman, Mylee Joseph, Vicki McDonald, Anna Raunik, Mia Ridge, and Mark Robertson. 2016. 'Opportunities for Academic and Research Libraries and Wikipedia: A Discussion Paper.' https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/topics/info-society/iflawikipediaopportunitiesforacademicandresearchlibraries.pdf.

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

Boccone, Alessandra. 2022. 'The Role of the Wikidata Librarian in a Renewed Bibliographical Universe: "Next Generation Metadata", next Generation Librarians.' *JLIS.it* 13 (2): 45–57. https://doi.org/10.36253/jlis.it-460.

Candela, Gustavo, Mirjam Cuper, Olga Holownia, Nele Gabriëls, Milena Dobreva, and Mahendra Mahey. 2024. 'A Systematic Review of Wikidata in GLAM Institutions: A Labs Approach.' In *Linking Theory and Practice of Digital Libraries*, edited by Apostolos Antonacopoulos, Annika Hinze, Benjamin Piwowarski, Mickaël Coustaty, Giorgio Maria Di Nunzio, Francesco Gelati, and Nicholas Vanderschantz, 34–50. Cham: Springer. https://doi.org/10.1007/978-3-031-72440-4_4.

Chen, Chen, and Zhong Yuxuan. 2024. 'A Comparative Study on the Approaches of Name Authority Control and Wikidata Identity Management.' In *Knowledge Organization for Resilience in Times of Crisis: Challenges and Opportunities*, edited by Wei Lu, Thiago Henrique Bragato Barros, Lu An, and Daniel Martínez-Ávila, 63–74. Baden-Baden: Ergon. https://doi.org/10.5771/9783987400476-63.

Chou, Charlene. 2025. 'Linking Challenges for Pseudonyms: The Studies of VIAF, ISNI, Wikidata, Meridian, and Beyond.' *Cataloging & Classification Quarterly* (2): 1–45. https://doi.org/10.1080/01639374.2025.2502329.

* All the links have been last retrieved on July 21, 2025 and rechecked on February 23, 2026; all the references to "now" and "currently" used in the article refer to July 21, 2025.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

De Monaco, Sara. 'Il Dizionario degli scrittori italiani contemporanei pseudonimi in Wikidata. Metodologia e risultati'. (Tesi di laurea magistrale, Università di Pisa, 2025), https://etd.adm.unipi.it/t/etd-05132025-160428.

De Monaco, Sara, and Camillo Carlo Pellizzari di San Girolamo. 2025. 'Il Dizionario degli scrittori italiani contemporanei pseudonimi in Wikidata. Metodologia e risultati (dati)'. Zenodo. https://doi.org/10.5281/ZENODO.15474929.

Dostál, Vojtěch, 'Bringing Czech Authority Files into 21st Century: Integration with Wikidata' (presentation, Data Quality Days, online, September 10, 2021). https://www.youtube.com/watch?v=JKqZTVisHC4.

Fischer, Barbara, 'Why We Use Wikibase at the German National Library. What It Takes to Collaboratively Produce Reliable Authority Files' (presentation, Wikidata Days Lisbon, Lisbon, Portugal, November 14, 2024). https://commons.wikimedia.org/wiki/File:Why_we_use_Wikibase_at_the_German_National_Library.pdf.

Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Knudson Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Craig Thomas, and Holly Tomren. 2019. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*. https://doi.org/10.25333/faq3-ax08.

Jansová, Linda, Lenka Maixnerová, and Petra Šťastná. 2024. 'Library Data in Wikimedia Projects: Case Study from the Czech Republic.' *O-Bib. Das Offene Bibliotheksjournal / Herausgeber VDB* 11 (4): 1–19. https://doi.org/10.5282/O-BIB/6081.

Klein, Maximilian, and Alex Kyrios. 2013. 'VIAFbot and the Integration of Library Data on Wikipedia.' *Code4lib Journal* 22. https://journal.code4lib.org/articles/8964.

Muratsubaki Campany, Mary, Kelly Davis, Sasha Frizzell, Elizabeth Plantz, and Crystal Yragui, 'LCCNbot: Creating Connections between the Library of Congress Name Authority File (LCNAF) & Wikidata' (presentation, LD4 Conference, online, October 8, 2024). https://www.youtube.com/watch?v=T-xKLnOqJ-4.

Pellizzari di San Girolamo, Camillo Carlo, 'Conflations and Duplications' (presentation., Data Modelling Days, December 1, 2023). https://commons.wikimedia.org/wiki/File:DMD2023_-_Conflations_and_duplications.pdf.

Pellizzari di San Girolamo, Camillo Carlo. 2024a. 'Conflations and Duplications in Wikidata Items: Causes, Detection, Solutions, and Issues.' In *Proceedings of the Wikidata Workshop 2023 Co-Located with 22nd International Semantic Web Conference (ISWC 2023)*. https://ceur-ws.org/Vol-3640/paper4.pdf.

Pellizzari di San Girolamo, Camillo Carlo. 2024b. 'Storia della collaborazione tra Wikidata e le biblioteche della Rete URBE nel controllo di autorità.' In *Parsifal. Un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*, edited by Silvano Danieli, 147–62. Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0356-2.17.

JLIS.it vol. 17, no. 2 (May 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-698

Pellizzari di San Girolamo, Camillo Carlo. 2024c. 'Riconciliare le voci di autorità in SBN con Wikidata. Progressi e prospettive dopo un decennio di lavoro (2013-2023)'. *JLIS.it* 15 (1): 33–44. https://doi.org/10.36253/jlis.it-573.

Pellizzari di San Girolamo, Camillo Carlo, 'Data Round-Tripping with Authority Files. An Overview and Some Examples' (presentation, Data Reuse Days, online, February 27, 2025). https://www.wikidata.org/wiki/File:Data_round-tripping_with_authority_files_(Data_Reuse_Days_2025).pdf.

Phillips, Michael, and Anne Washington. 2022. 'Shared Entity Management Infrastructure di OCLC.' *Biblioteche oggi* 40 (6): 44–51. https://doi.org/10.3302/0392-8586-202206-044-1.

Ravelli, Elena, 'Authority file di SBN e Wikidata. Un progetto per l'arricchimento dei dati del Catalogo nazionale' (presentation, Wikidata Days Bologna, Bologna, Italy, November 8, 2024). https://commons.wikimedia.org/wiki/File:Authority_file_di_SBN_e_Wikidata_-_Wikidata_Days_Bologna_2024.pdf.

Sardo, Lucia, and Carlo Bianchini. 2022. 'Wikidata: A New Perspective towards Universal Bibliographic Control.' *JLIS.it* 13 (1): 291–311. https://doi.org/10.4403/jlis.it-12725.

Tharani, Karim. 2021. 'Much More than a Mere Technology: A Systematic Review of Wikidata in Libraries.' *The Journal of Academic Librarianship* 47 (2): 102326. https://doi.org/10.1016/j.acalib.2021.102326.

Wikimedia Deutschland, 'GND in Wikidata' (presentation, April 14, 2025). https://commons.wikimedia.org/wiki/File:GND_in_Wikidata.pdf.

Zapounidou, Sofia, Lazaros Ioannidis, Michalis Gerolimos, Eftychia Koufakou, and Charalampos Bratsas. 2024. 'Entity Management Using RDA and Wikibase: A Case Study at the National Library of Greece.' *Journal of Library Metadata* 24 (2): 111–31. https://doi.org/10.1080/19386389.2024.2307208.

Zhang, Erica, Paromita Biswas, and Iman Dagher. 2023. 'Reflections on the PCC Wikidata Pilot at UCLA Library: Undertaking the PCC Learning Objectives.' *Cataloging & Classification Quarterly* 61 (7–8): 735–72. https://doi.org/10.1080/01639374.2023.2269416.

Zhao, Fudie. 2023. 'A Systematic Review of Wikidata in Digital Humanities Projects.' *Digital Scholarship in the Humanities* 38 (2): 852–74. https://doi.org/10.1093/llc/fqac083.