JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12725

JLIS.it

Wikidata: a new perspective towards universal bibliographic control*

Carlo Bianchini[a], **Lucia Sardo**[b]

a) Università degli studi di Pavia, http://orcid.org/0000-0002-6635-6371
b) Università di Bologna. Campus di Ravenna, http://orcid.org/0000-0001-6480-759X

Contact: Carlo Bianchini, carlo.bianchini@unipv.it; Lucia Sardo, lucia.sardo@unibo.it

Received: 10 April 2021; **Accepted:** 22 May 2021; **First Published:** 15 January 2022

ABSTRACT

Traditional UBC provides for the standardization of bibliographic records, the creation of guidelines dedicated to national bibliographic agencies, the creation of the UNIMARC format, and the curation of authority data. Bibliographic Control has deeply evolved since IFLA theorization during the Seventies of the XX Century, due to the availability of a very large range of new bibliographic tools. At the beginning of the XXI century, UBC is quite different and involves new actors. Among these, Wikidata has a background greatly different from that of libraries as institutions: it is not devoted to bibliographic data, nor it is limited to personal authority control, but its value in AC tools like VIAF and National Libraries authority files is undiscussed. After a presentation on how Wikidata items describe and identify bibliographic entities, the authors underline how the existence, use and reuse of Wikidata affect the way the professional community thinks about UBC. Wikidata is a clear example of the need for a new approach to identification and description, that are deeply intertwined. Secondly, from a Wikidata perspective, the relevance of globally preferred and variant access points is lessened. Moreover, descriptions in Wikidata – although conceptually very similar to the traditional one – present differences and potentialities that a traditional description does not have and cannot have. Also from a theoretical perspective, Wikidata offers a pragmatic way to think globally and act locally. In fact, it shows that there is no need for one standardization of practices for establishing the headings and structure of authority records in one international form; instead, users' convenience can be achieved by a technological infrastructure capable to present to each user the information about an entity in its own language and script. Additionally, Wikidata is the most evident example of the distributed and diffused approach of the semantic web to the issue of the universal identification of the entities. Also, Wikidata identification and description show that authority and bibliographic control must be tackled as just a part of the more general topic of the creation of a knowledge graph of all human knowledge by means of linked open data. Lastly, this objective cannot be achieved only by contribution, cooperation, and networking of large national agencies (as in VIAF), as a larger number of stakeholders must be involved to achieve a UBC also including the full indexing of any kind of scientific communication

KEYWORDS

UBC; Universal Bibliographic Control; Wikidata; Metadata; Description; Identification.

* The authors cooperated in the redaction and revision of the article. Nevertheless, each author mainly authored specific parts of the article: Carlo Bianchini: sections 1.2, 2, 4, and 5; Lucia Sardo: sections 1, 3, 4.2, and 5.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12725

1. Introduction: From UBC to the semantic web

The idea of bibliographic control, that is, the idea of being able to have an exhaustive overview of what is published ("the mastery over written and published records which is provided by and for the purposes of bibliography"; Unesco and LC Bibliographical Survey 1950, 1), at least as old as bibliography, took on new connotations in the 1970s, with the birth of IFLA's Universal Bibliographic Control program.

The main objective of universal bibliographic control is the availability of bibliographic records of publications produced in all countries. In the context of the program for the UBC, the emphasis is placed not only on the universality of such control, but also on the standardization of the content of bibliographic records, on the need to have a specific program dedicated to this objective to foster cooperation at the international level and to achieve the goal of a worldwide record of what is published, and on the importance given to the rapid availability of these data.

The program for the UBC has its basis in standardization and in the direct involvement of national bibliographic agencies, a fundamental element for international cooperation: during the 1977 congress devoted to national bibliographies, their tasks were defined to be the documentation of national editorial production, and the drafting of authority records for national authors.

The principles on which universal bibliographic control is based, which have been progressively brought into focus through studies and initiatives, can be summarized as follows: first, the aims of the system are the control and exchange of bibliographic information; worldwide coverage is guaranteed by the cooperation of the national components of the system; the main objective is to make the bibliographic data of all publications universally and promptly available, in an internationally accepted standard format. To achieve this goal, the complete bibliographic record of each publication should be made once only in the country of origin by a national bibliographic agency, in accordance with international standards that permit exchange. In this perspective, the national bibliographic agency, usually established in national libraries, which usually benefit from the mandatory deposit of printed matter, results to be the most appropriate structure for authoritatively identifying and recording the authors and publications of each country, and is responsible for producing a current national bibliography, in which to publish such records as soon as possible, and for distributing such records in various standard formats. The agencies then come to be integrated into an international system and regularly exchange records made.

For UBC principles to be implemented and scaled up, some requirements must be satisfied:

- a canon of principles, standards, and practices governing the creation and structure of catalographic data that is shared on a broad scale must be available.
- each national bibliographic agency must fulfil its responsibilities in a manner that is inclusive of and consistent with accepted standards.
- an infrastructure is needed to support the efficient exchange of data among national bibliographic agencies.

While IFLA's work on satisfactorily scaling up the UBC concept regarding bibliographic records has been successful, regarding authorities it has been largely driven by the recognition of the need to deal with these three critical factors:

- the standardization of practices for establishing the headings and structure of authorities.

- the promotion of national responsibilities for the creation and "dissemination" of authority records.
- the planning of an infrastructure that supports the effective international exchange of authority records.

The results of the program's work for the UBC are there for all to see: the publication of ISBD, the creation of the UNIMARC format, the publication of authority lists and tools for controlling the forms of personal and collective names, and the Guidelines for the National Bibliographic Agencies and the National Bibliography. Basically, all the work concerning the standardization of bibliographic descriptions and authority records had its basis in the concept of Universal Bibliographic Control.

1.1 UBC for bibliographic and authority data

From the point of view of the UBC, the ISBD standard was first developed, which had the double function of establishing which data were relevant for bibliographic description and in which order they should be presented. This was the first time that such a standardization effort was undertaken; it preceded the formalization of the program but provided for it as a *conditio sine qua non* for its dissemination.[1]

In the same period MARC, a machine-readable format, was created for the exchange of cataloguing information. IFLA, too, considered it essential to develop an international MARC format capable of supporting the exchange of bibliographic data, which is why the development of the UNIMARC format was undertaken, both for bibliographic and authority data. All of this was born, we recall, in a context where catalogs were paper-based, and national needs trumped those of internationalization, especially about the choice of name form for access points.

As Gorman summarizes, "In sum, arriving at a standard set of elements in a standard order and delimited in a standard manner was in the mutual interest of the effort to achieve an international standard for bibliographic description (what became the ISBD); MARC; [sic] and the use of both, each in accord with the other, in achieving national and international standardization, cooperation, and sharing; leading, ultimately, to Universal Bibliographic Control" (Gorman 2014, 826-827).

1.2 Wikidata: a tool of bibliographic interest in the semantic web

In 2011, the Library Linked Data Incubator Group, a working group with the aim "to help increase global interoperability of library data on the Web",[2] published its final report. It was focused on what libraries can do for the semantic web and what the semantic web can do for libraries, and it underlined that libraries had created and curated a relevant amount of rich data that can "help reduce redundancy of bibliographic descriptions on the Web by clearly identifying key entities that are shared across Linked Data" (W3C Incubator Group 2011). The report offered a new perspective on thinking about the relevance, scope, and purpose of Universal

[1] For an overview about the origins of ISBD, see (Anderson 1974; Gorman 2014).

[2] https://www.w3.org/2005/Incubator/lld/

Bibliographic Control (UBC), beside to "make universally and promptly available, in a form which is internationally acceptable, basic bibliographic data on all publications in all countries" (Anderson 1974, 11).

Since the publication of the Report, many tools with a top-down approach have been developed for the identification of entities (people, locations, works, and expressions) such as VIAF, ISNI or ORCID. The top-down approach of these tools reflects the role assigned to the national agencies by UBC. Nevertheless, some authors suggest that, in the semantic web, "building collaborative authority registries linked to standardised identifiers is one of the fundamental cornerstones of the new Universal Bibliographic Control" (Illien and Bourdon 2014, 15) and that "a better mix of bottom-up and top-down methodologies" is needed to support all those who wish to think globally and act locally (Dunsire and Willer 2014, 11).

Since 2012, Wikidata has developed as a new global actor of the semantic web with both a bottom-up and very inclusive approach. Wikidata is a freely available hosted platform that anyone – including libraries – can use to create, publish, and reuse LOD (Allison-Cassin and Scott 2018). Its main goal and function are to work as a central storage for many Wikimedia projects, but it is also used in external services, for example in VIAF or in the Google Knowledge graph (Vrandečić and Krötzsch 2014), for the enrichment of the quality of bibliographic records (Nguyen, Dinneen, and Luczak-Roesch 2020), and for bibliometrics projects and tools (Lemus-Rojas and Odell 2018; Nielsen, Mietchen, and Willighagen 2017; Hernández-Cazorla, Ramírez-Sánchez, and Rodríguez-Herrera 2019; Seidlmayer et al. 2020; Mietchen and Rasberry 2020). Moreover, in the last years, the Wikidata role as an important tool for identifying entities has been increasingly reconsidered (Association of Research Libraries 2019, 27; van Veen 2019; Linked Data for Production 2020).[3]

2. Identification in Wikidata

As Wikidata is a central storage for all Wikimedia projects, it aims to record data about any kind of item (i.e., entity) and property relevant for all its projects. For example, items of Wikidata can be geographical places, administrative units, events, architectonic objects, any entity of interest for the user, and, of course, any 'res' provided for by IFLA LRM model.

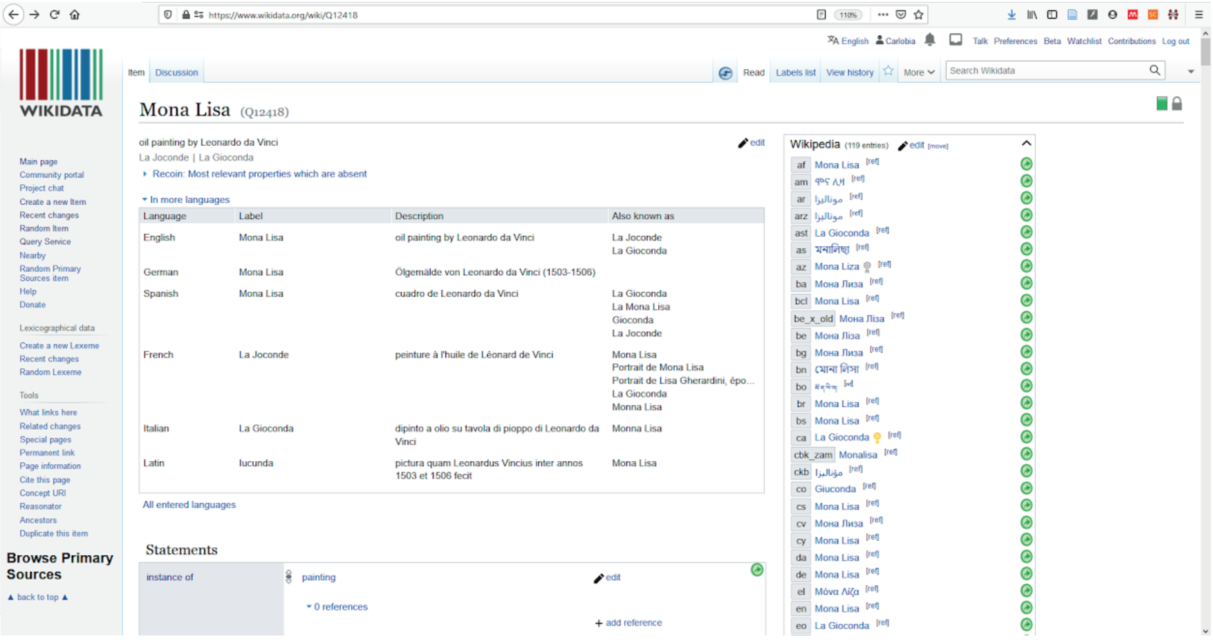
In fact, Wikidata shows a relevant interest for the bibliographic universe. Statistics show that Wikidata records about 91 million of items, 31,5% (ca 22,5 million) of which are scholar articles, and nearly 9% (ca 6.376.000) of the existing items are of human type (Q5). Anyway, this class includes any kind of humans, such as kings, politicians, football players and so on, and not just authors of literary or scientific works. Nevertheless, items representing an authority record (not just of humans) can be estimated to be around 6,3 million.[4]

For identification purposes, Wikidata assigns to each item both an URI – for example, https://www.wikidata.org/wiki/Q12418 – and a label, a description, and one or more aliases (figure 1).

[3] https://www.wikidata.org/wiki/Wikidata:Wikidata_for_authority_control

[4] Personal items with a VIAF identifier are about 2,3 million, but the number of personal items containing at least one identifier of any VIAF source (such as ISNI, LC, GND etc.) are about 6,3 million.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12725



Mona Lisa (Q12418)

oil painting by Leonardo da Vinci
La Joconde | La Gioconda

Recoin: Most relevant properties which are absent

In more languages

Language	Label	Description	Also known as
English	Mona Lisa	oil painting by Leonardo da Vinci	La Joconde La Gioconda
German	Mona Lisa	Ölgemälde von Leonardo da Vinci (1503-1506)	
Spanish	Mona Lisa	cuadro de Leonardo da Vinci	La Gioconda La Mona Lisa Gioconda La Joconde
French	La Joconde	peinture à l'huile de Léonard de Vinci	Mona Lisa Portrait de Mona Lisa Portrait de Lisa Gherardini, épo... La Gioconda Monna Lisa
Italian	La Gioconda	dipinto a olio su tavola di pioppo di Leonardo da Vinci	Monna Lisa
Latin	Iucunda	pictura quam Leonardus Vincius inter annos 1503 et 1506 fecit	Mona Lisa

All entered languages

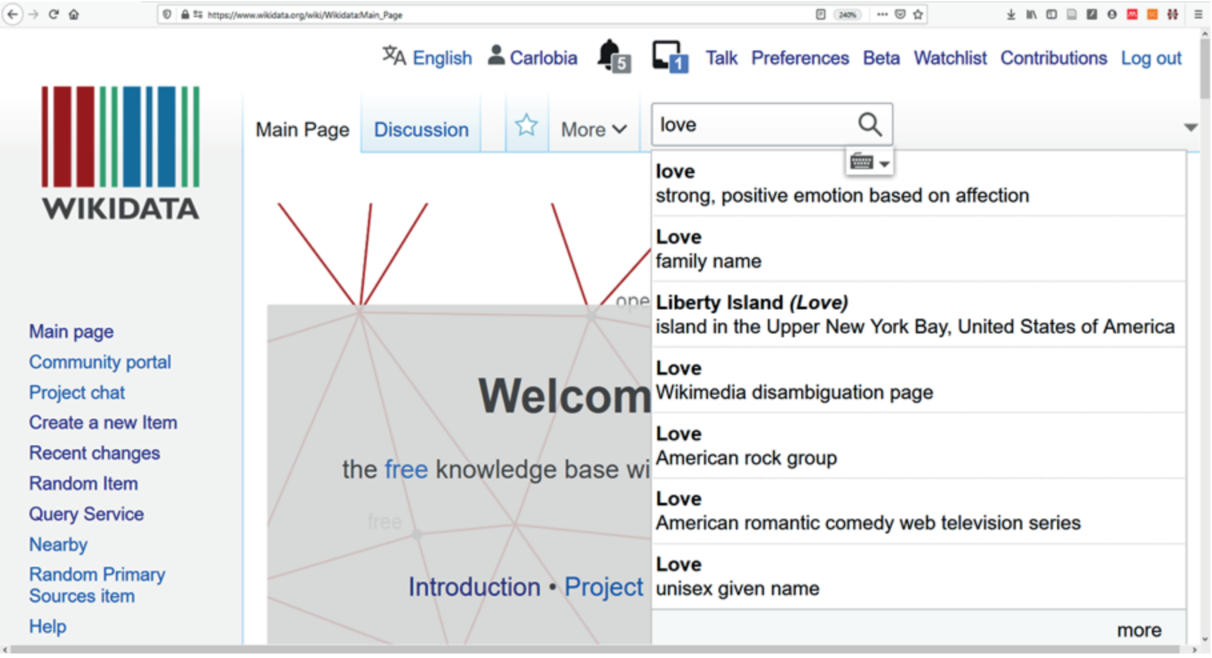
Wikipedia (119 entries)

af	Mona Lisa [ref]
am	ሞና ሊዛ [ref]
ar	موناليزا [ref]
arz	موناليزا [ref]
ast	La Gioconda [ref]
as	মনালিছা [ref]
az	Mona Liza [ref]
ba	Мона Лиза [ref]
bcl	Mona Lisa [ref]
be_x_old	Мона Ліза [ref]
be	Мона Ліза [ref]
bg	Мона Лиза [ref]
bn	মোনা লিসা [ref]
bo	[illegible] [ref]
br	Mona Lisa [ref]
bs	Mona Lisa [ref]
ca	La Gioconda [ref]
cbk_zam	Monalisa [ref]
ckb	مۆنالیزا [ref]
co	Giuconda [ref]
cs	Mona Lisa [ref]
cv	Мона Лиза [ref]
cy	Mona Lisa [ref]
da	Mona Lisa [ref]
de	Mona Lisa [ref]
el	Μόνα Λίζα [ref]
en	Mona Lisa [ref]
eo	La Gioconda [ref]

Fig. 1. Example of the main identifying parts (label, description, aliases) of a Wikidata item

The label is the first data element, and it can be considered as the preferred form of the name for the represented entity. In fact, it can be expressed in any existing language and any registered user is enabled to visualize the label in his/her own language and script, if available. Moreover, the preferred form of the name is expressed directly by the user that usually creates the item. So, preferred forms in Wikidata are *literally* founded on common usage and on the convenience of the user provided for by ICP (IFLA Cataloguing Section and IFLA Meeting of Experts on an International Cataloguing Code 2016), and not on these principles *interpreted by a code* of national or international rules! It must be noted that multiple languages and scripts are available for the very same entity, and not for a cluster of nationally created forms like in VIAF.

The second element is the description of the item, that is a short phrase to describe the item. It is in free language and it is useful to quickly distinguish an item from any other item with the same label (for example, "Love"; figure 2), i.e., to disambiguate homonyms.



love
strong, positive emotion based on affection

Love
family name

Liberty Island *(Love)*
island in the Upper New York Bay, United States of America

Love
Wikimedia disambiguation page

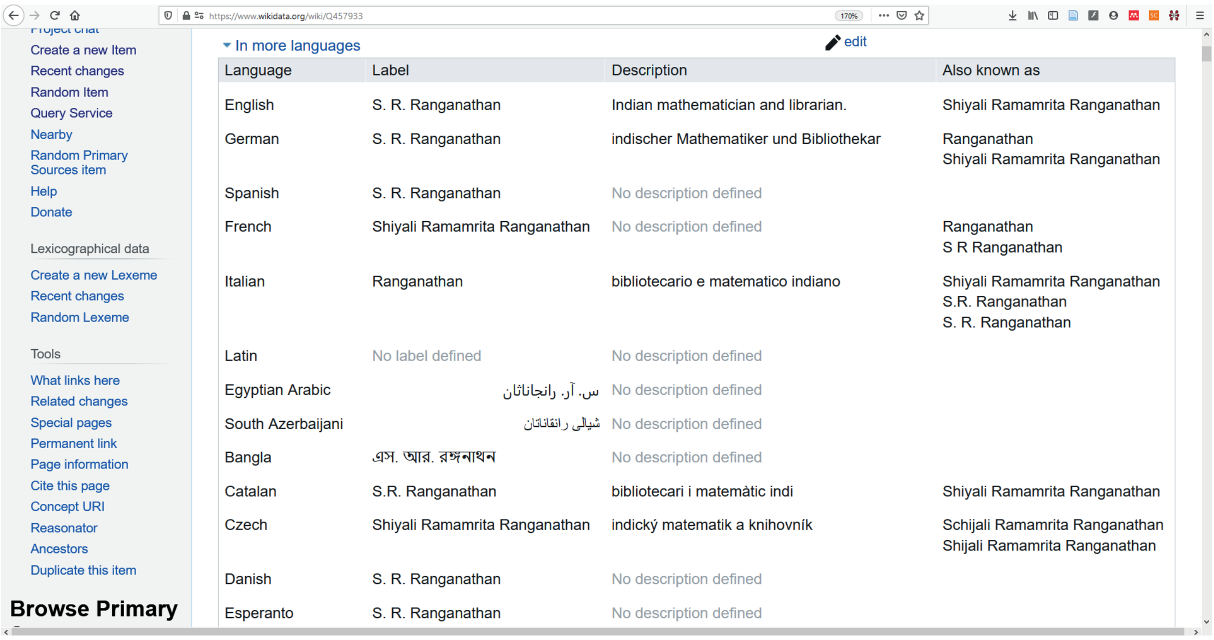
Love
American rock group

Love
American romantic comedy web television series

Love
unisex given name

Fig. 2. Descriptions helps to disambiguate items with the same label "Love"

The third element for the quick item identification in Wikidata are the aliases, that are variant forms of the name in one specific language and script (as variant forms of the name in any other language are provided in the form of preferred and variant names in those languages; figure 3).



Language	Label	Description	Also known as
English	S. R. Ranganathan	Indian mathematician and librarian.	Shiyali Ramamrita Ranganathan
German	S. R. Ranganathan	indischer Mathematiker und Bibliothekar	Ranganathan Shiyali Ramamrita Ranganathan
Spanish	S. R. Ranganathan	No description defined	
French	Shiyali Ramamrita Ranganathan	No description defined	Ranganathan S R Ranganathan
Italian	Ranganathan	bibliotecario e matematico indiano	Shiyali Ramamrita Ranganathan S.R. Ranganathan S. R. Ranganathan
Latin	No label defined	No description defined	
Egyptian Arabic	س. آر. رانجاناثان	No description defined	
South Azerbaijani	شیالی رانقاناتان	No description defined	
Bangla	এস. আর. রঙ্গনাথন	No description defined	
Catalan	S.R. Ranganathan	bibliotecari i matemàtic indi	Shiyali Ramamrita Ranganathan
Czech	Shiyali Ramamrita Ranganathan	indický matematik a knihovník	Schijali Ramamrita Ranganathan Shijali Ramamrita Ranganathan
Danish	S. R. Ranganathan	No description defined	
Esperanto	S. R. Ranganathan	No description defined	

Fig. 3. Aliases available in multiple languages and scripts for S.R. Ranganathan

While the unique identification of the entity is based on a neutral URI (for example: https://www.wikidata.org/wiki/Q1334284), both labels and aliases – in any available language and script – work as access points. This pragmatical approach overcomes the theorical ICP and RDA distinction between preferred and variant access points.

All the remaining properties are registered after the identification elements described above. Nevertheless, they are logically divided into main parts: properties and identifiers. While *properties* are traditionally associated with the *descriptive* goal of the data (see below § no. 3), all the other external identifiers respond to the need of the fourth linked data principle stated by Tim Berners-Lee: "Include links to other URIs so that [users] can discover more things" (Berners-Lee 2006).

External identifiers have the goal to interlink the URI of the item of Wikidata with any other identifiable entity described in the semantic web.[5] For this reason, Wikidata is more and more recognised for its relevance in the identification of semantic web entities (Association of Research Libraries, 2020; Linked Data for Production, 2020; van Veen 2020). Enriching data with Wikidata ids allows to discover other sources of data and information available in the semantic web.

To create a link towards an external identifier, a specific property must be created in Wikidata to define that identifier. So, it is possible to know how many identifiers are available for different kinds of entities. In figure 4, created by Simon Cobb (Cobb 2019, 5), the number of identifiers associated with each kind of entity are shown.

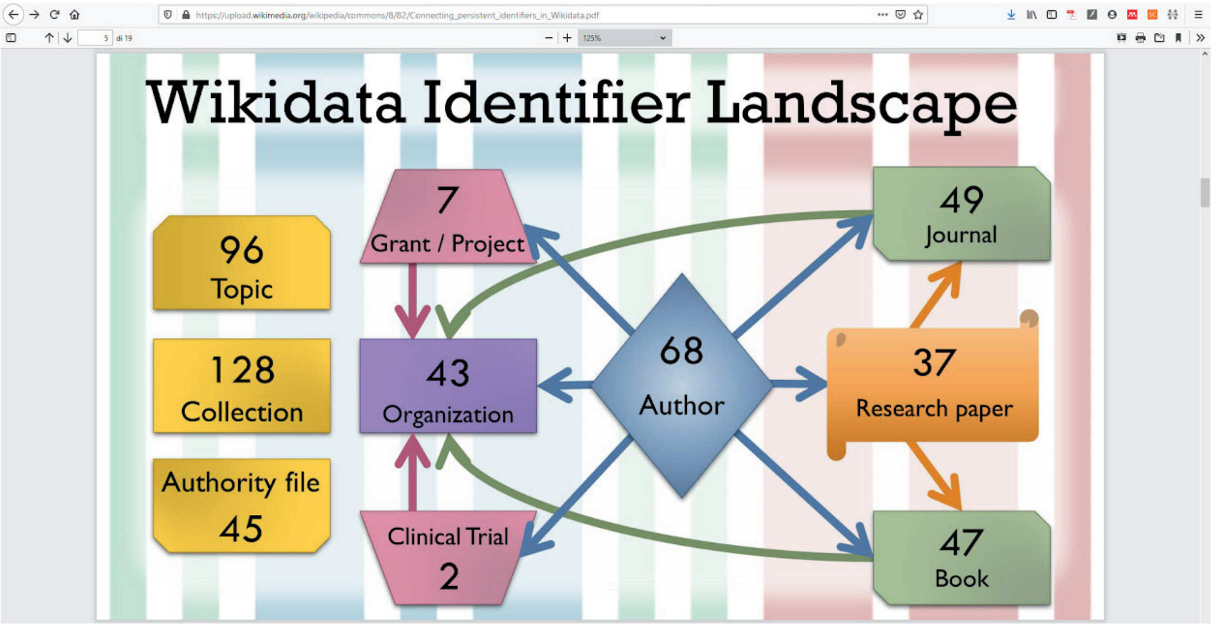

https://upload.wikimedia.org/wikipedia/commons/8/82/Connecting_persistent_identifiers_in_Wikidata.pdf
5
di 19
125%
Wikidata Identifier Landscape
96
Topic
128
Collection
Authority file
45
7
Grant / Project
43
Organization
Clinical Trial
2
68
Author
49
Journal
37
Research paper
47
Book

Fig. 4. Number of identifiers in Wikidata for each kind of entity (by Cobb 2019)

[5] Entity Explosion is a very interesting tool to understand the potential uses of this WD function for the navigation in our discovery tools; see https://chrome.google.com/webstore/detail/entity-explosion/bbcffeclligkmfiocanodamdjclgejcn.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12725

Most frequently used identifiers in Wikidata are: PubMedID (60,152,490), DOI (26,816,446), PMCID (11,339,676), SIMBAD (8,159,240) and VIAF (6,050,830).[6]
Actually, a major role of Wikidata as a hub for identification in the semantic web is recognised by the VIAF. In fact, VIAF uses Wikidata as an 'other data provider', i.e., a provider of data other than a National bibliographic agency. Among the Wikidata items with a VIAF identifier, most common identifiers registered in the personal items are, in decrescent order: ISNI (1,136,260; 18%); DBN (1,012,493; 16%); LC (983,206; 15%); NTA (480,580; 7%); SUDOC (431,919; 6,8%) and BNF (428,792; 6,8%) (Bargioni, Bianchini, and Pellizzari 2021, table 5). The relationship between Wikidata and VIAF is very strong. Wikidata uses property constraints to discover possible inconsistencies in statements both within Wikidata and in the external sources.[7] So, Wikidata users can check the issues and try to fix them, but any external service can take advantage of this characteristic too.
Identification in Wikidata is a process oriented to the quality of data. First, Wikidata explicitly requires – with the second notability criterion – that each item refers to "a clearly identifiable conceptual or material entity. The entity must be notable, in the sense that it can be described using serious and publicly available references".[8] For example, notability prevents Wikidata from accepting isolated clusters formed by VIAF based on a single contributor identifier. Second, clusters of identifiers in a Wikidata item are created by common users and not by automatically performed matches. Matches may be performed semi-automatically (by means of tools such as OpenRefine[9] or Mix'n'match[10]) but human control is always required. Moreover, as in authority work, references are mandatory for each triple and reference sources include encyclopedias, biographical dictionaries, scientific books and articles, in addition to VIAF and other national libraries authority data.
More and more Wikidata initiatives are oriented to improve the quality of authors' data. An example is offered by the bots: a bot in Wikidata "is a program that is allowed to upload large scale data and that is quality controlled by the community" (Siedlmayer 2020). For example, during SWIB 2020 OrcBot was presented: it is a tool created to take advantage of ORCID ids to improve the recording of the property that links the author items to their respective papers, based on ORCID Ids. The Enhancing author items process and issues of reconciliation between ORCID and Wikidata were the focus of the talk "Author items in Wikidata"' at the WikiCite Virtual conference 2020 by Simon Cobb – wikimedian in residence at the National Library of Wales.[11]

3. Description in Wikidata

The "traditional" bibliographical description, marked by descriptive areas in ISBD, and fields and subfields of the MARC format, was firstly challenged by the birth of electronic catalogs, or rather

[6] https://www.wikidata.org/wiki/Wikidata:Database_reports/List_of_properties/all, visited 15 December 2020.

[7] Wikidata helps in identifying issues by two approaches: unique value violations and single value violations. A detailed description of both the approaches and their practical relevance as a quality control tool applied to VIAF is available in (Bargioni, Bianchini, and Pellizzari 2021).

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

[9] https://openrefine.org/.

[10] https://mix-n-match.toolforge.org/. See also (Agenjo-Bullón and Hernández-Carrascal 2020).

[11] https://upload.wikimedia.org/wikipedia/commons/7/79/Author_items_in_Wikidata.pdf

by the new, previously impossible, opportunity to search in any part of the description. Moreover, the electronic catalog challenges the need for a layout made up of descriptive areas, because the flag format of visualization, highlighting the metadata/data structure, overcome the value of the order of citation and the semantics of punctuation.

With the evolution of electronic catalogs and the gradual occurrence of major commercial players (at this stage) into the world of libraries and catalogs, we have therefore come to have tools that allow research and access to different databases, produced by different parties with different purposes. But in this situation the standardization of description is progressively fraying and being lost, in favor of a greater speed in the availability of information about resources, often produced directly by those who produce and make available the resource themselves, whether in analog or digital format.

However, this advantage is detrimental on the search side, as it increases the noise in catalogs and discovery tools, and an insufficiently skilled user may find it difficult to disentangle the results obtained.

The impetuous technological evolution of the 21st century, together with a reflection on the functions and object of cataloguing that has led to radical changes in the way we approach resources, is beginning to show the consequences of all this in the cataloguing world. The semantic web and linked data are influencing the ways in which bibliographic data (in the broadest sense) are created, shared, and made available to potential users. In this phase, moreover, no-profit stakeholders outside the world of libraries are becoming increasingly present. An example is Wikidata, created and implemented by a community of volunteers with different training and mindset than those traditionally linked to the book professions, and by volunteers who deal with bibliographic data management.

On the side of libraries, traditional standards are losing their central role in bibliographic description. The static and linear MARC format, subject to many criticisms for many years, is giving way, with difficulty, to different models, such as BIBFRAME, which offers greater flexibility in line with the developments of the semantic web and linked data.

The ISBD format, still used as a standard for describing resources in some cataloging standards, is marking time for the moment. It remains the basis for both the practice and the teaching of cataloging in some settings and situations, but it cannot be denied that we are moving towards other newer and wider ways of describing, like RDA (Resource Description and Access).

RDA, in its first version still linked to AACR2, despite its innovativeness; but, in the new official version from December 2020 has radically changed the way to approach the description of resources. It uses IFLA LRM as a basis for its implementation, with some adaptations that the editors have considered essential for the "practical" needs of the library community, in line with what is expressed in the model, namely that implementations and changes are possible while respecting the basic structure.

On the one hand, RDA allows different levels of description with different types of data encoding (from the mere literal transcription to the use of IRI), on the other hand, it enables libraries using or wanting to use it to adopt different possibilities of use and implementation, with the only constraint to remain "faithful" to the framework and the general choices proposed by the standard and therefore to be interoperable with other realities that use it.

Anyway, an ethical problem must be highlighted: ISBD is a free descriptive standard, while RDA is not; if ISBD disappears, what will remain to those who cannot afford access to RDA?

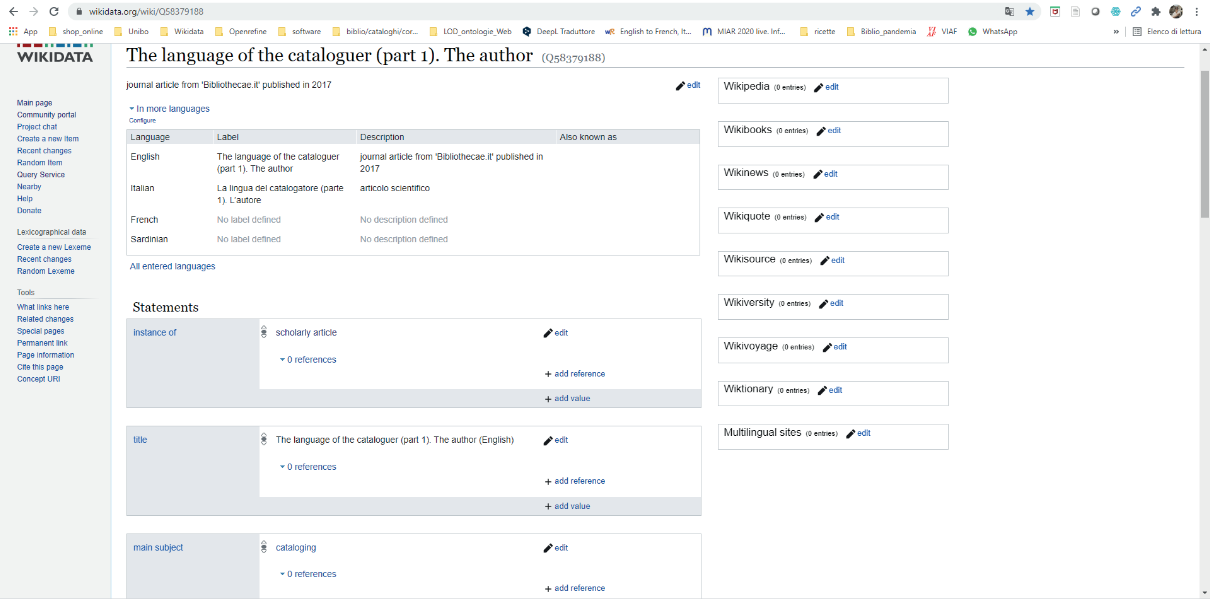
The description in Wikidata, on the other hand, although conceptually very similar to the traditional one, presents differences and potentialities that a traditional description does not have and cannot have.

As said, conceptually the basic elements of a description are those we are used to in the world of libraries, but we can immediately highlight some important innovations to increase the potential of the description. First, the full implementation of the modelling of the bibliographic universe of IFLA models, with the representation of Works, Expressions, Manifestations, and Items. Secondly, the possibility offered by Wikidata to qualify data. Finally, the possibility of integrating in the description identifiers of different types coming from different sources. While in traditional bibliographic description data qualification was impossible, in Wikidata this is not only feasible, but advisable. In this way you can achieve great advantages for all types of resources that you want to describe. Qualifying is not just specifying the data sources, but their chronological or geographical context; for example, the period of use of a form of a printer's name, a form of a place name, or the language used is a major advantage and a potential that has yet to be fully exploited.

Another aspect relevant to the concept of UBC is the possibility of going in depth in the description of resources. By this we mean that if the UBC very often stopped at the monographic level, in Wikidata instead it is possible to find descriptions of journal articles, or "sheets" of conference proceedings or miscellany. Indeed, perhaps because this lack is significant in traditional catalogues, these types of resources represent a very high percentage of the items in Wikidata.

Certainly, some aspects need to be improved, such as the correct attribution of properties to the right level and the creation of relationships, for example, between works, expressions, events, and items, but the potential is great and the foundations are sufficiently solid to be able to think of continuing the construction of a valid tool for a new vision of UBC, not tied to national conditioning or commercial logic.

The challenges that will have to be faced, at another level, will instead be those related to the use, and the visualization/reuse of these data, but this is not the place to delve into the matter.



The language of the cataloguer (part 1). The author (Q58379188)

journal article from 'Bibliothecae.it' published in 2017 edit

In more languages
Configure

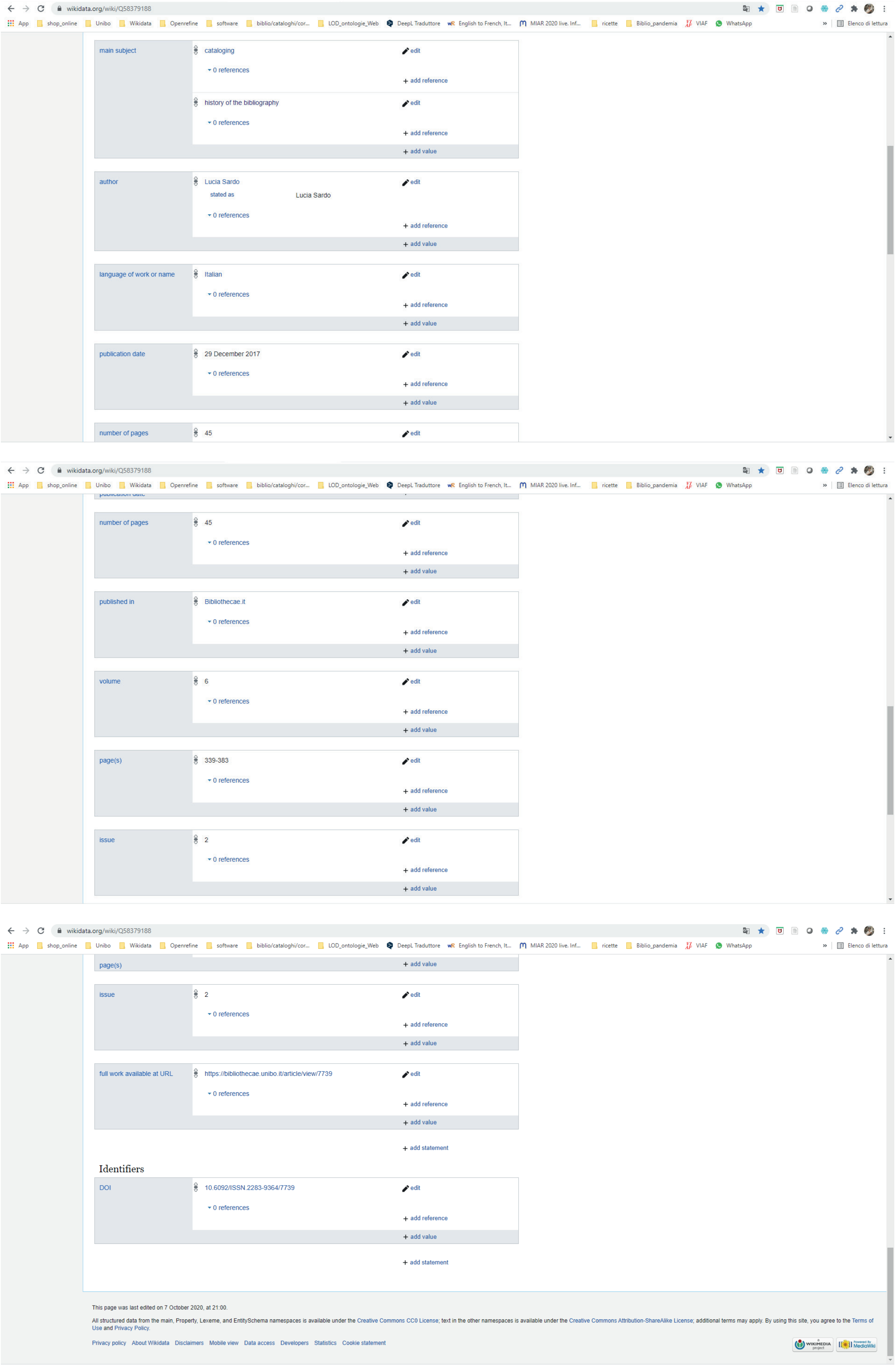
Language	Label	Description	Also known as
English	The language of the cataloguer (part 1). The author	journal article from 'Bibliothecae.it' published in 2017	
Italian	La lingua del catalogatore (parte 1). L'autore	articolo scientifico	
French	No label defined	No description defined	
Sardinian	No label defined	No description defined	

All entered languages

Statements

instance of	scholarly article 0 references	edit + add reference + add value
title	The language of the cataloguer (part 1). The author (English) 0 references	edit + add reference + add value
main subject	cataloging 0 references	edit + add reference

Wikipedia (0 entries) edit
Wikibooks (0 entries) edit
Wikinews (0 entries) edit
Wikiquote (0 entries) edit
Wikisource (0 entries) edit
Wikiversity (0 entries) edit
Wikivoyage (0 entries) edit
Wiktionary (0 entries) edit
Multilingual sites (0 entries) edit



main subject — cataloging — 0 references — edit — add reference
history of the bibliography — 0 references — edit — add reference — add value

author — Lucia Sardo — edit
stated as Lucia Sardo
0 references — add reference — add value

language of work or name — Italian — 0 references — edit — add reference — add value

publication date — 29 December 2017 — 0 references — edit — add reference — add value

number of pages — 45 — edit

number of pages — 45 — 0 references — edit — add reference — add value

published in — Bibliothecae.it — 0 references — edit — add reference — add value

volume — 6 — 0 references — edit — add reference — add value

page(s) — 339-383 — 0 references — edit — add reference — add value

issue — 2 — 0 references — edit — add reference — add value

page(s) — add value

issue — 2 — 0 references — edit — add reference — add value

full work available at URL — https://bibliothecae.unibo.it/article/view/7739 — 0 references — edit — add reference — add value

add statement

Identifiers

DOI — 10.6092/ISSN.2283-9364/7739 — 0 references — edit — add reference — add value

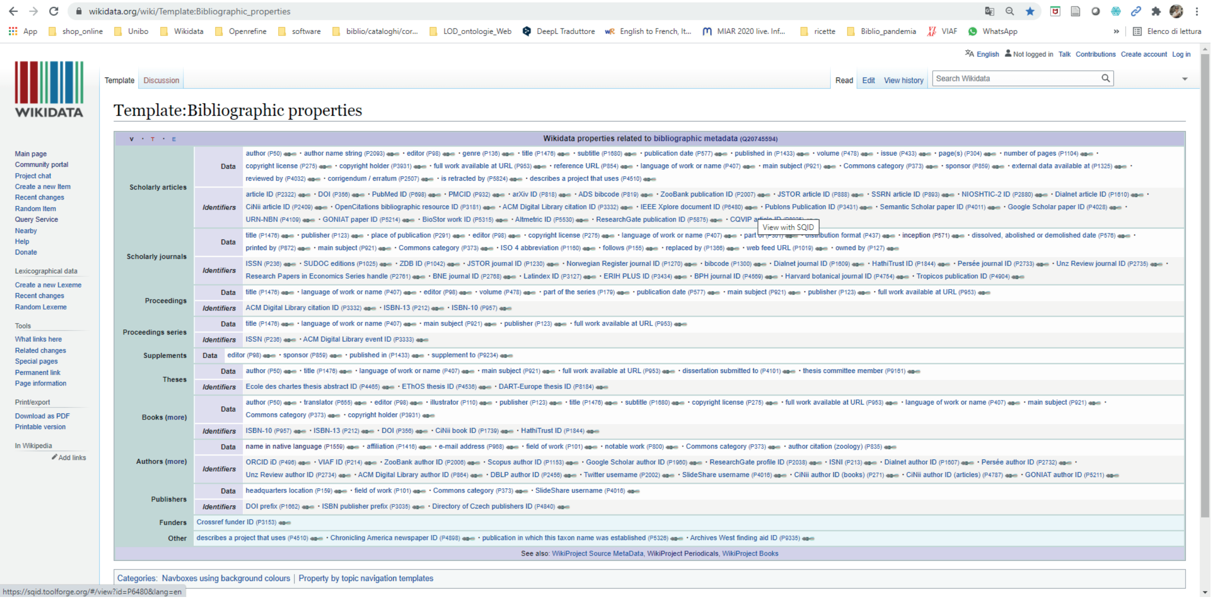
add statement

This page was last edited on 7 October 2020, at 21:00.

All structured data from the main, Property, Lexeme, and EntitySchema namespaces is available under the Creative Commons CC0 License; text in the other namespaces is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy.

Privacy policy About Wikidata Disclaimers Mobile view Data access Developers Statistics Cookie statement

Fig. 5. Example of a Wikidata description of a scholarly article. https://www.wikidata.org/wiki/Q58379188

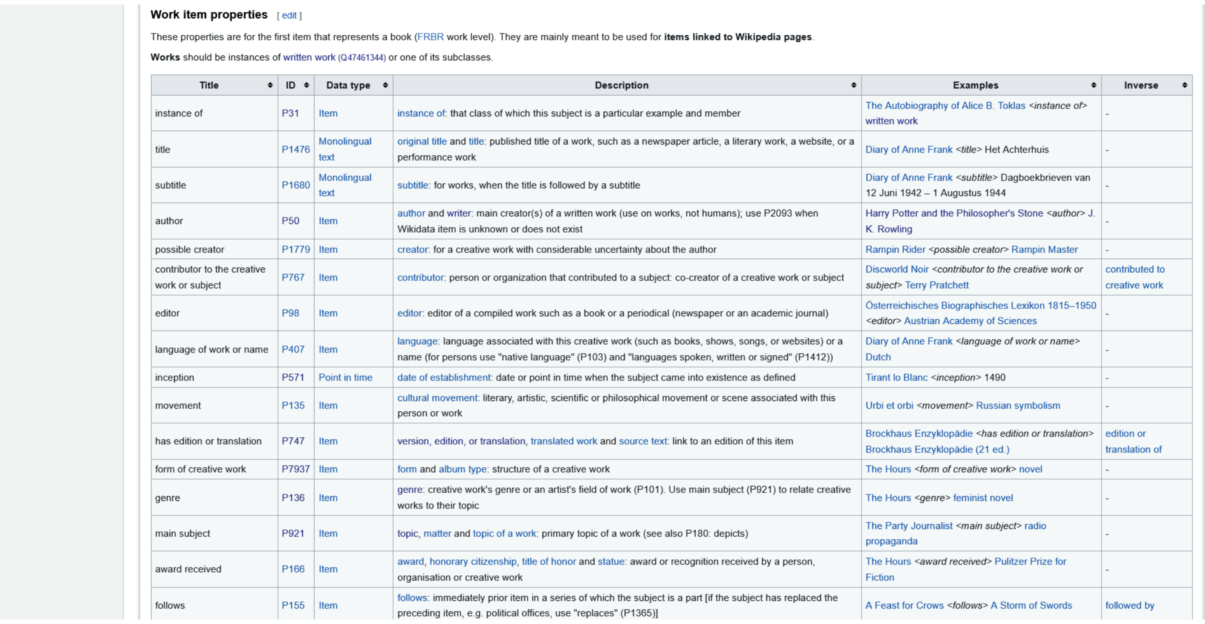


Template:Bibliographic properties

v · t · e		Wikidata properties related to bibliographic metadata (Q20745594)
Scholarly articles	Data	author (P50) · author name string (P2093) · editor (P98) · genre (P136) · title (P1476) · subtitle (P1680) · publication date (P577) · published in (P1433) · volume (P478) · issue (P433) · page(s) (P304) · number of pages (P1104) · copyright license (P275) · copyright holder (P3931) · full work available at URL (P953) · reference URL (P854) · language of work or name (P407) · main subject (P921) · Commons category (P373) · sponsor (P859) · external data available at (P1325) · reviewed by (P4032) · corrigendum / erratum (P2507) · is retracted by (P5824) · describes a project that uses (P4510)
	Identifiers	article ID (P2322) · DOI (P356) · PubMed ID (P698) · PMCID (P932) · arXiv ID (P818) · ADS bibcode (P819) · ZooBank publication ID (P2007) · JSTOR article ID (P888) · SSRN article ID (P893) · NIOSHTIC-2 ID (P2880) · Dialnet article ID (P1610) · CiNii article ID (P2409) · OpenCitations bibliographic resource ID (P3181) · ACM Digital Library citation ID (P3332) · IEEE Xplore document ID (P6480) · Publons Publication ID (P3431) · Semantic Scholar paper ID (P4011) · Google Scholar paper ID (P4028) · URN-NBN (P4109) · GONIAT paper ID (P5214) · BioStor work ID (P5315) · Altmetric ID (P5530) · ResearchGate publication ID (P5875) · CQVIP a[illegible]
Scholarly journals	Data	title (P1476) · publisher (P123) · place of publication (P291) · editor (P98) · copyright license (P275) · language of work or name (P407) · part [illegible] · [illegible]bution format (P437) · inception (P571) · dissolved, abolished or demolished date (P576) · printed by (P872) · main subject (P921) · Commons category (P373) · ISO 4 abbreviation (P1160) · follows (P155) · replaced by (P1366) · web feed URL (P1019) · owned by (P127)
	Identifiers	ISSN (P236) · SUDOC editions (P1025) · ZDB ID (P1042) · JSTOR journal ID (P1230) · Norwegian Register journal ID (P1270) · bibcode (P1300) · Dialnet journal ID (P1609) · HathiTrust ID (P1844) · Persée journal ID (P2733) · Unz Review journal ID (P2735) · Research Papers in Economics Series handle (P2761) · BNE journal ID (P2768) · Latindex ID (P3127) · ERIH PLUS ID (P3434) · BPH journal ID (P4569) · Harvard botanical journal ID (P4754) · Tropicos publication ID (P4904)
Proceedings	Data	title (P1476) · language of work or name (P407) · editor (P98) · volume (P478) · part of the series (P179) · publication date (P577) · main subject (P921) · publisher (P123) · full work available at URL (P953)
	Identifiers	ACM Digital Library citation ID (P3332) · ISBN-13 (P212) · ISBN-10 (P957)
Proceedings series	Data	title (P1476) · language of work or name (P407) · main subject (P921) · publisher (P123) · full work available at URL (P953)
	Identifiers	ISSN (P236) · ACM Digital Library event ID (P3333)
Supplements	Data	editor (P98) · sponsor (P859) · published in (P1433) · supplement to (P9234)
Theses	Data	author (P50) · title (P1476) · language of work or name (P407) · main subject (P921) · full work available at URL (P953) · dissertation submitted to (P4101) · thesis committee member (P9161)
	Identifiers	Ecole des chartes thesis abstract ID (P4465) · EThOS thesis ID (P4536) · DART-Europe thesis ID (P8184)
Books (more)	Data	author (P50) · translator (P655) · editor (P98) · illustrator (P110) · publisher (P123) · title (P1476) · subtitle (P1680) · copyright license (P275) · full work available at URL (P953) · language of work or name (P407) · main subject (P921) · Commons category (P373) · copyright holder (P3931)
	Identifiers	ISBN-10 (P957) · ISBN-13 (P212) · DOI (P356) · CiNii book ID (P1739) · HathiTrust ID (P1844)
Authors (more)	Data	name in native language (P1559) · affiliation (P1416) · e-mail address (P968) · field of work (P101) · notable work (P800) · Commons category (P373) · author citation (zoology) (P835)
	Identifiers	ORCID iD (P496) · VIAF ID (P214) · ZooBank author ID (P2006) · Scopus author ID (P1153) · Google Scholar author ID (P1960) · ResearchGate profile ID (P2038) · ISNI (P213) · Dialnet author ID (P1607) · Persée author ID (P2732) · Unz Review author ID (P2734) · ACM Digital Library author ID (P864) · DBLP author ID (P2456) · Twitter username (P2002) · SlideShare username (P4016) · CiNii author ID (books) (P271) · CiNii author ID (articles) (P4787) · GONIAT author ID (P5211)
Publishers	Data	headquarters location (P159) · field of work (P101) · Commons category (P373) · SlideShare username (P4016)
	Identifiers	DOI prefix (P1662) · ISBN publisher prefix (P3035) · Directory of Czech publishers ID (P4840)
Funders		Crossref funder ID (P3153)
Other		describes a project that uses (P4510) · Chronicling America newspaper ID (P4898) · publication in which this taxon name was established (P5326) · Archives West finding aid ID (P9335)
		See also: WikiProject Source MetaData, WikiProject Periodicals, WikiProject Books

Categories: Navboxes using background colours | Property by topic navigation templates

Fig. 6. Wikidata template with all the properties for the bibliographic description – sample; cfr. https://www.wikidata.org/wiki/Template:Bibliographic_properties


Work item properties [edit]

These properties are for the first item that represents a book (FRBR work level). They are mainly meant to be used for **items linked to Wikipedia pages**.

Works should be instances of written work (Q47461344) or one of its subclasses.

Title	ID	Data type	Description	Examples	Inverse
instance of	P31	Item	instance of: that class of which this subject is a particular example and member	The Autobiography of Alice B. Toklas *<instance of>* written work	-
title	P1476	Monolingual text	original title and title: published title of a work, such as a newspaper article, a literary work, a website, or a performance work	Diary of Anne Frank *<title>* Het Achterhuis	-
subtitle	P1680	Monolingual text	subtitle: for works, when the title is followed by a subtitle	Diary of Anne Frank *<subtitle>* Dagboekbrieven van 12 Juni 1942 – 1 Augustus 1944	-
author	P50	Item	author and writer: main creator(s) of a written work (use on works, not humans); use P2093 when Wikidata item is unknown or does not exist	Harry Potter and the Philosopher's Stone *<author>* J. K. Rowling	-
possible creator	P1779	Item	creator: for a creative work with considerable uncertainty about the author	Rampin Rider *<possible creator>* Rampin Master	-
contributor to the creative work or subject	P767	Item	contributor: person or organization that contributed to a subject: co-creator of a creative work or subject	Discworld Noir *<contributor to the creative work or subject>* Terry Pratchett	contributed to creative work
editor	P98	Item	editor: editor of a compiled work such as a book or a periodical (newspaper or an academic journal)	Österreichisches Biographisches Lexikon 1815–1950 *<editor>* Austrian Academy of Sciences	-
language of work or name	P407	Item	language: language associated with this creative work (such as books, shows, songs, or websites) or a name (for persons use "native language" (P103) and "languages spoken, written or signed" (P1412))	Diary of Anne Frank *<language of work or name>* Dutch	-
inception	P571	Point in time	date of establishment: date or point in time when the subject came into existence as defined	Tirant lo Blanc *<inception>* 1490	-
movement	P135	Item	cultural movement: literary, artistic, scientific or philosophical movement or scene associated with this person or work	Urbi et orbi *<movement>* Russian symbolism	-
has edition or translation	P747	Item	version, edition, or translation, translated work and source text: link to an edition of this item	Brockhaus Enzyklopädie *<has edition or translation>* Brockhaus Enzyklopädie (21 ed.)	edition or translation of
form of creative work	P7937	Item	form and album type: structure of a creative work	The Hours *<form of creative work>* novel	-
genre	P136	Item	genre: creative work's genre or an artist's field of work (P101). Use main subject (P921) to relate creative works to their topic	The Hours *<genre>* feminist novel	-
main subject	P921	Item	topic, matter and topic of a work: primary topic of a work (see also P180: depicts)	The Party Journalist *<main subject>* radio propaganda	-
award received	P166	Item	award, honorary citizenship, title of honor and statue: award or recognition received by a person, organisation or creative work	The Hours *<award received>* Pulitzer Prize for Fiction	-
follows	P155	Item	follows: immediately prior item in a series of which the subject is a part [if the subject has replaced the preceding item, e.g. political offices, use "replaces" (P1365)]	A Feast for Crows *<follows>* A Storm of Swords	followed by

Fig. 7. Wikidata template with the Work item properties; https://www.wikidata.org/wiki/Wikidata:WikiProject_Books#Work_item_properties

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12725

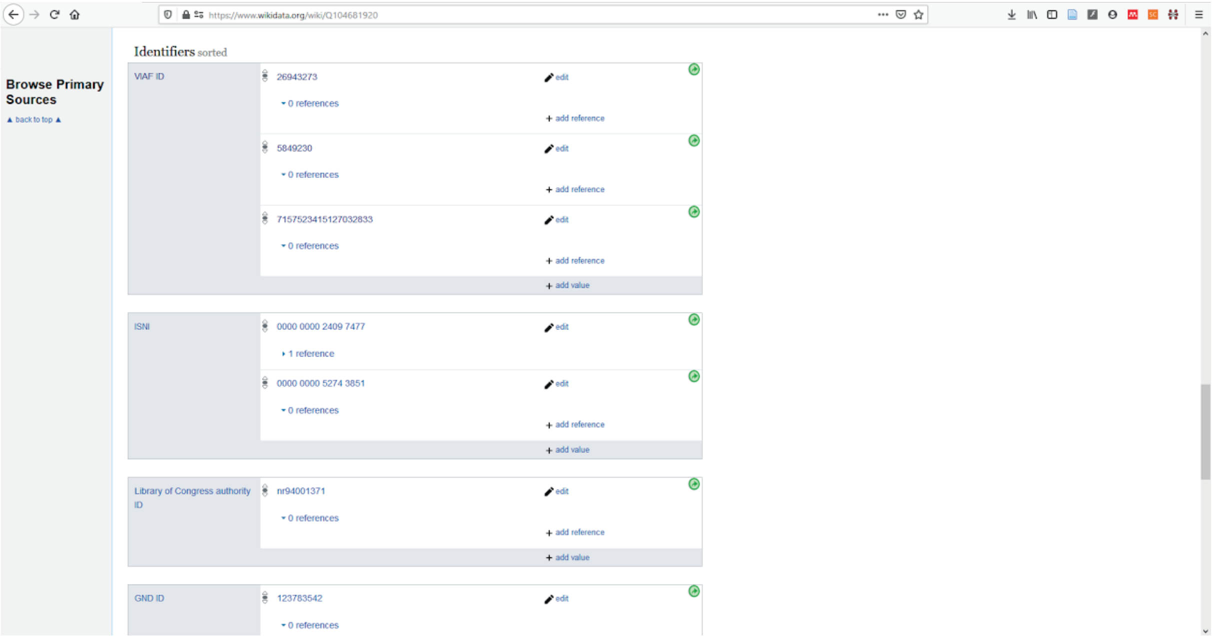
4. Wikidata as a bibliographic tool. Strategies, projects, and tools

4.1 Identification

To understand how the identification process can be improved in Wikidata, it is necessary to distinguish two possible approaches: by identifiers and by properties of the items.

The most important tool of quality control for proper identification of items are identifiers.[12] In fact, every property associated with an external identifier is provided with constraint rules – usually, an external identifier must be associated with only one item and one item must have only one identifier per type. These rules are extremely important for bibliographic control. In fact, they allow to identify possible errors within Wikidata (e.g., a duplicated item), but above all, they show possible mistakes also within the sources of the external identifiers linked to a Wikidata item (e.g., when a duplication of external identifiers occurs in a Wikidata item).

An example of how it works can be useful to understand how much Wikidata can help in the identification work for persons. A quick check of the identifiers associated with "Ferruccio Battolini" shows that three distinct VIAF IDs and two distinct ISNI IDs are associated with the same person (figure 8a).[13] This example is relatively simple, but things get more complicated with classical authors (e.g., poetess Saffo; figure 8b).[14]



https://www.wikidata.org/wiki/Q104681920

Identifiers sorted

VIAF ID	26943273 0 references 5849230 0 references 7157523415127032833 0 references
ISNI	0000 0000 2409 7477 1 reference 0000 0000 5274 3851 0 references
Library of Congress authority ID	nr94001371 0 references
GND ID	123783542 0 references

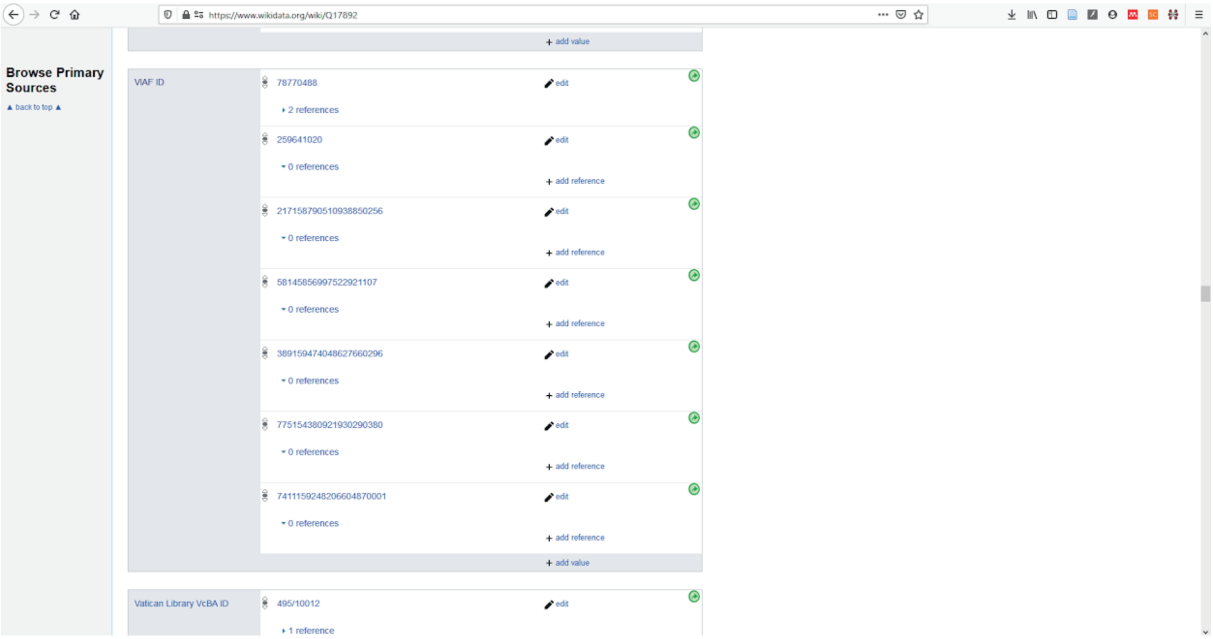
Fig. 8a. Duplicated VIAF and ISNI identifiers for Ferruccio Battolini

[12] See Wikidata Project: https://www.wikidata.org/wiki/Wikidata:WikiProject_Authority_control.

[13] https://www.wikidata.org/wiki/Q104681920.

[14] https://www.wikidata.org/wiki/Q17892.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12725



VIAF ID

78770488
2 references

259641020
0 references

217158790510938850256
0 references

58145856997522921107
0 references

389159474048627660296
0 references

775154380921930290380
0 references

7411159248206604870001
0 references

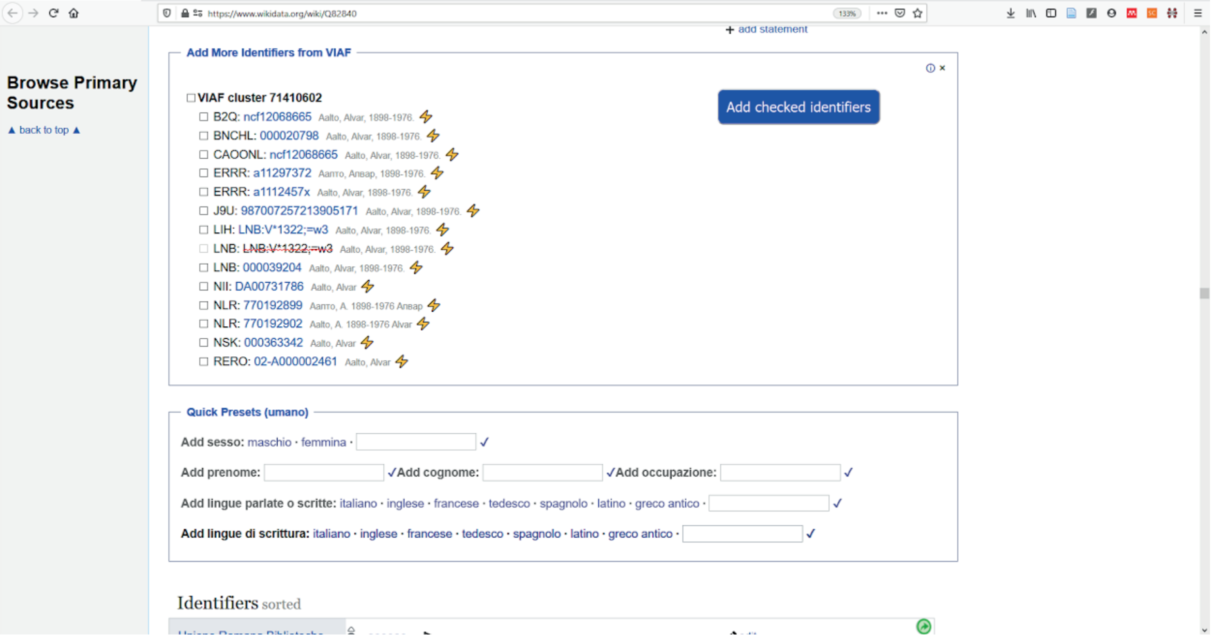
Vatican Library VcBA ID

495/10012
1 reference

Fig. 8b. Duplicated VIAF identifiers for Sappho

As VIAF remains a major source for Wikidata, the community developed specific tools – named gadgets – to improve the reuse of its data in Wikidata items. Gadgets are enhancements of the edit interface for registered users and are very useful for data production.[15] For instance, the gadget *MoreIdentifiers* was created by Stefano Bargioni and Camillo Pellizzari to facilitate the creation of links between Wikidata items and VIAF entities and it enables users to add easily and quickly authority control IDs from VIAF with few edits checking the identifier and clicking on the button (figure 9).[16] Moreover, it enables to know whether an identifier is old or wrong (as it is presented strikethrough in red) and to create a report for any wrong identifier wrongly included in the VIAF cluster, if the case, by means of the thunder icon. A page of identifiers wrongly included in a VIAF cluster is maintained and constantly updated by Wikidata users; alas, it seems not so used by VIAF managers.[17]

[15] A list of gadgets is available at https://www.wikidata.org/wiki/Wikidata:VIAF/cluster#Gadgets.
[16] https://www.wikidata.org/wiki/User:Bargioni/moreIdentifiers.
[17] https://www.wikidata.org/wiki/Wikidata:VIAF/cluster/conflating_specific_entries.



https://www.wikidata.org/wiki/Q82840

+ add statement

Browse Primary Sources

▲ back to top ▲

Add More Identifiers from VIAF

Add checked identifiers

□ **VIAF cluster 71410602**
- □ B2Q: ncf12068665 Aalto, Alvar, 1898-1976.
- □ BNCHL: 000020798 Aalto, Alvar, 1898-1976.
- □ CAOONL: ncf12068665 Aalto, Alvar, 1898-1976.
- □ ERRR: a11297372 Аалто, Алвар, 1898-1976.
- □ ERRR: a1112457x Aalto, Alvar, 1898-1976.
- □ J9U: 987007257213905171 Aalto, Alvar, 1898-1976.
- □ LIH: LNB:V*1322;=w3 Aalto, Alvar, 1898-1976.
- □ LNB: ~~LNB:V*1322;=w3~~ Aalto, Alvar, 1898-1976.
- □ LNB: 000039204 Aalto, Alvar, 1898-1976.
- □ NII: DA00731786 Aalto, Alvar
- □ NLR: 770192899 Аалто, А. 1898-1976 Алвар
- □ NLR: 770192902 Aalto, A. 1898-1976 Alvar
- □ NSK: 000363342 Aalto, Alvar
- □ RERO: 02-A000002461 Aalto, Alvar

Quick Presets (umano)

Add sesso: maschio · femmina · ✓

Add prenome: ✓**Add cognome:** ✓**Add occupazione:** ✓

Add lingue parlate o scritte: italiano · inglese · francese · tedesco · spagnolo · latino · greco antico · ✓

Add lingue di scrittura: italiano · inglese · francese · tedesco · spagnolo · latino · greco antico · ✓

Identifiers sorted

Fig. 9. Box of the Wikidata gadget *MoreIdentifiers*.

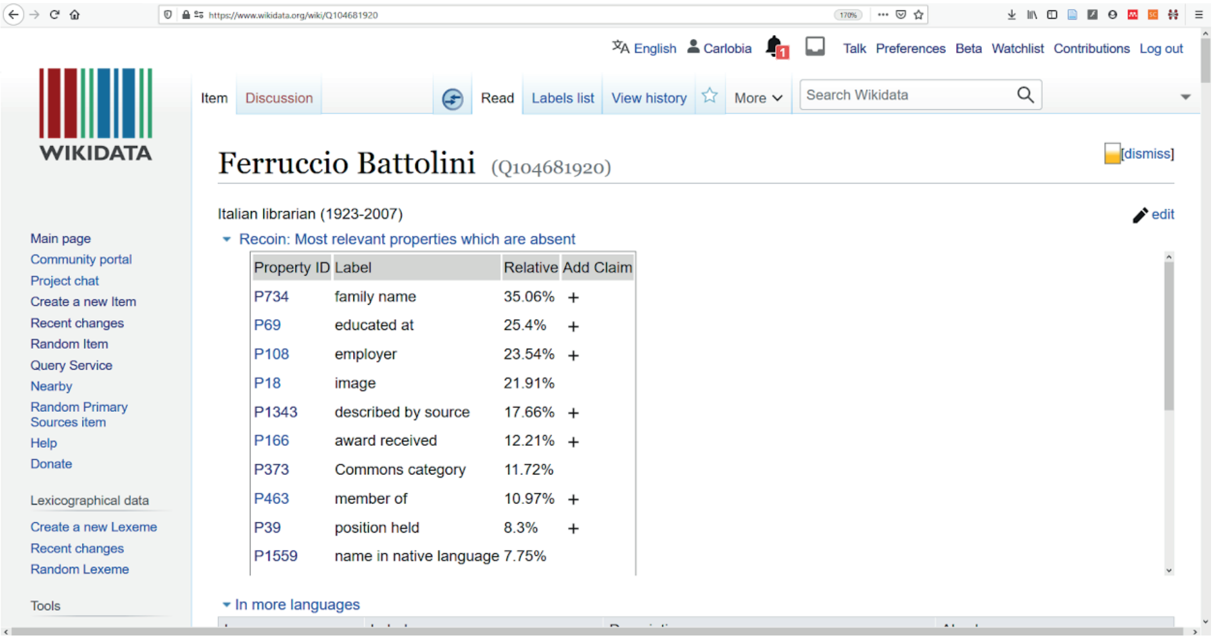
MoreIdentifiers works for any kind of VIAF entity (such as geographic names or corporate names) but it is best useful for personal names.

Properties are key for the identification of items within Wikidata. In this case, identification is based on the matching of several properties. For example, human beings' identification is usually based on the matching of the label, the description, and the dates of birth and death.

In this approach, the more the available properties, the more the probabilities for identification. So, the number of properties of an item is a key issue, because a higher number of properties describing an entity assure a more probable identification – or disambiguation – of the two entities being compared.

For this reason, a dedicated gadget was developed by Wikidata community: *Recoin*, i.e., *Relative Completeness in Wikidata* (figure 10). Recoin is a "script that extends Wikidata entity pages with information about the *relative completeness* of the information" referring to the "extent of information found on an item in comparison with other similar items".[18] Recoin is a tool to help authors of Wikidata to know on which data attention must be focused on; moreover, it is also extremely useful for data consumers to be aware of the degree of information available about an item.

[18] https://www.wikidata.org/wiki/Wikidata:Recoin.



Ferruccio Battolini (Q104681920)

Italian librarian (1923-2007)

Recoin: Most relevant properties which are absent

Property ID	Label	Relative	Add Claim
P734	family name	35.06%	+
P69	educated at	25.4%	+
P108	employer	23.54%	+
P18	image	21.91%	
P1343	described by source	17.66%	+
P166	award received	12.21%	+
P373	Commons category	11.72%	
P463	member of	10.97%	+
P39	position held	8.3%	+
P1559	name in native language	7.75%	

Fig. 10. Example of missing properties highlighted by Recoin

As shown in figure 10, Recoin offers a status indicator icon, ranging from very detailed to very basic, to indicate the relative completeness of the description of an item on a 5-level scale, and a list of the most relevant properties which are not present in the item. Missing properties are detected by a comparison of the properties in that item and the properties most frequently occurring in that class. For example, the properties in an item representing a politician are compared to the most frequently occurring properties of the item belonging to the class ‘politicians’ (Balaraman, Razniewski, and Nutt 2018).

Nevertheless, identification within Wikidata is far from being perfect and can still be improved. Many new items are poorly described because of two main issues: many items are created by semi-automatic processes and, for this reason, data can be incorrect, generic (e.g., string versus author; cf. below),[19] or poor. In addition, at present Wikidata as a semantic web hub, is undoubtedly more oriented towards identifying than describing items.

When data derived from external sources are incorrect, their limit is inherited in the Wikidata item description (as seen above with VIAF identifiers). For example, a large part of the item creation work from sources like ORCID is made by bulk upload from bots; this means that in these cases “errors can persist for many months without being rectified and can be replicated in bulk editing of the description without detection” (Cobb 2020, 3).

External data can result in generic data too. For example, it is possible to import data from Zotero – a Reference Manager Software – to Wikidata, but the authors of the books or the articles are recorded as a *string of characters* (P2093), instead of a relationship between the item and the *author* (P50). And this happens with many other automatic tools, so that about

[19] https://www.wikidata.org/wiki/Wikidata:Database_reports/List_of_properties/all.

135 million of authors are recorded as strings compared to just 20 million recorded as author relationships.

Poor external source data can produce a low number of average statements, and this means a poorer description and a more difficult process of identification of the items (mainly towards other external sources). For example, a study by Simon Cobb shows that items having an ORCID have a low number of statements; so that "the latest author items are sparse in comparison to older items, which have had longer to attract the curatorial efforts of community members" (Cobb 2020, 3).

Anyway, as author item relations allow for much richer analysis, to fix the issue, the special tool *Author disambiguator*[20] was created: it is a tool for editing the authors of works recorded in Wikidata and for assisting in converting "those strings into links to author items as efficiently and easily as possible".[21]

Another issue for authority control in Wikidata is that the data harvesting process is not structured; initiatives to upload data are very much, and sometimes based on semi-automatic tools, but there is not a clear overall design nor strategy, as typical of bottom-up approaches.

At the end of his analysis, Simon Cobb suggests a few steps to improve identification process for Wikidata items that can be applied to any kind of item and of external source; major suggestions are:

- Seek community consensus on minimum acceptable standard for author items created by bot imports.
- Define author data requirements for a variety of use cases.
- Review and validate data in existing author items.
- Organise an online workshop to facilitate discussion and collaboration between interested members of the Wikidata editor community and other stakeholders within and outside the Wikimedia Foundation projects.
- Establish a WikiProject special interest group (SIG) to focus on the improvement and maintenance of author items" (Cobb 2020, 10).

4.2 Description

The improvement of description in Wikidata can be approached from at least two points of view: the description of a remarkable variety of items in Wikidata, and the more traditional description of bibliographic resources.

In the first case (item description) the issues are certainly more intertwined with the problems of identification, because to have good descriptions, a correct identification of the entity is necessary and therefore the number and quality of properties necessary and sufficient for the description itself must be defined.

In the second case, instead, the improvement would require the growth of the available resources and their univocal identification when possible (by means of identifiers such as ISBN, ISSN, DOI). Problems arise for all the resources that do not have an identifier univocally assigned by an internationally recognized agency, but that have only identifiers assigned by the world of libraries (BID; etc.).

[20] https://author-disambiguator.toolforge.org/ .

[21] https://www.wikidata.org/wiki/Wikidata:Tools/Author_Disambiguator. See also (Smith 2020)

A key issue is the quality of the source metadata, as digitized resources or "digital libraries" show when collecting the product of digitization from different sources (one example for all, Internet Archive). In such situations, the critical issues concern the description of the "less standardized" events and the need of a proper identification of the expressions and works on the one hand, and on the other, a trustful reconstruction of the physical presentation of the events, to facilitate more specialized or bibliographic research.

The description approach provided by RDA, i.e., based on identifiers and IRIs, is particularly effective in a context like Wikidata, and could lead to a significant growth of the number of described resources, and to an enhancement in the quality of the descriptions, as well as to the correct identification of the various entities.

Compared to this, the advantages related to the possibility of making a more granular and detailed bibliographic control than library catalogs are certainly notable (articles, miscellanea perusal, etc.); the possibility of inserting identifiers related to the catalogs of the major libraries or library systems worldwide also allows to satisfy the user function *to obtain*, which is often what most users want as a result of a search.

Finally, we remember that description and identification issues are inevitably intertwined.

5. Suggestions from Wikidata for the UBC

Wikidata is a Wikimedia tool to meet the needs of Wikimedia platforms, but it has relevant bibliographic features that can help to better understand the future of the Universal Bibliographic Control. In fact, Wikidata offers a completely new approach to data management that involves the way in which our community thinks and operates the Universal Bibliographic Control, both from a practical and theoretical perspective.

Wikidata is not designed as a bibliographic tool, and it is not oriented, nor limited, to bibliographic resources. For this reason, even if a data schema is available as Wikidata property page for works, editions, scientific articles, serials and so on, the quality and completeness of bibliographic data are usually high, but not certain. In fact, the number and the quality of the identifiers and properties recorded in Wikidata items are very varying, and the oldest items are usually more well-structured than the most recent ones; anyway, many gadget and tools (*MoreIdentifiers*, *Recoin*, *Author Disambiguator*, etc.) are available to improve them. Furthermore, while its bottom-up approach is a major asset in a global environment in which the role of great national bibliographic agencies is unable to fulfil the requirements of UBC, it is also a limit for the lack of a clear overall strategy of implementation of authority and bibliographic data.

Anyway, from a practical perspective, Wikidata is a clear example of the need for a new approach to identification and description, that are intertwined. First, a change in the workflow and in the mindset is required to the cataloguer, because a basic and even problematic identification must precede the description of the item.

Secondly, the relevance of globally preferred and variant access points is lessened; in fact, they remain relevant just in a local environment, and in a specific context defined by a particular set of rules. While labels and aliases pragmatically meet the requirements of making data accessible for any users' search, the identification function – a pillar of UBC – is assured by international identifiers, among which Wikidata ID is more and more significant. Moreover, the description in Wikidata – although conceptually very similar to the traditional one – presents differences and potentialities that a traditional description does not have and cannot have. First, the full implementation of the modelling of the bibliographic universe of IFLA models is available, with the representation of Works, Expressions, Manifestations, and Items. Third, the possibility of integrating in the description identifiers of different types coming from different sources, above all from the library field. Last, but not least, the possibility offered by Wikidata to qualify data. For instance, the chance to specify the period of use of a form of a printer's name, or of a place name, or of the used language is a major advantage and a potential that has yet to be fully exploited.

Another relevant point in Wikidata practical perspective is its major value as a *de facto* infrastructure to support the efficient exchange of bibliographic data among users, especially those who are not national bibliographic agencies. Wikidata is already a major hub of the semantic web also for bibliographic purposes. Moreover, Wikidata can record and disseminate bibliographic data of analytic descriptions, such as scholarly articles or chapters of books. Finally, Wikidata upgrades the concept of authority work, including reference both to the main international library catalogs and to local library catalogs and to a wider variety of reference sources (such as encyclopedias, dictionaries, and biographical repertories).

From a theoretical perspective, Wikidata offers a pragmatic way to think globally and act locally. For instance, it suggests looking at authority data as just a part of a wider perspective in which we produce and record data. For instance, it helps to recognize that an 'author' is just a person with a typed relationship toward a work, or a subject is any kind of entity with another typed relationship with a work. Authors and subjects, in a sense, do not exist in 'nature', but they become meaningful only in a bibliographic data perspective, and they must be expressed by a relation of authorship or aboutness between entities.

Furthermore, it shows that there is no need for one standardization of practices for establishing the headings and structure of authority records in one international form; instead, users' convenience can be achieved by a technological infrastructure capable to present to each user the information about an entity in its own language and script. Wikidata is the most evident example of the distributed and diffused approach of the semantic web to the issue of the universal identification of the entities. Moreover, thanks to a bottom-up and co-operative approach, Wikidata fulfils the requirements of International cataloguing Principles of common usage and convenience of the user by means of the users themselves.

There are other two relevant points about the contribution of Wikidata to the theoretical framework of the Universal Bibliographic Control. The first is that authority and bibliographic control must be tackled as just a part of the more general topic of the creation of a knowledge graph of all human knowledge by means of linked open data. In fact, Wikidata and the Semantic Web record data for any kind of item and not just for entities of bibliographic interest. In this new context, data for the achievement of the Universal Bibliographic Control and data, information, resources controlled by the Universal Bibliographic Control are perfectly integrated in one structure. The second is that this objective cannot be achieved only by contribution, cooperation, and networking of large National Agencies, as a larger number of stakeholders must be involved to achieve a UBC also including the full indexing of any kind of scientific communication.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12725

Bibliographic references

Agenjo-Bullón, Xavier, and Francisca Hernández-Carrascal. 2020. 'Wikipedia, Wikidata y Mix'n'match'. *Anuario ThinkEPI* 14. https://doi.org/10/ghbj6t.

Allison-Cassin, Stacy, and Dan Scott. 2018. 'Wikidata: A Platform for Your Library's Linked Open Data'. *Code4Lib Journal*, 4 May 2018. https://journal.code4lib.org/articles/13424.

Anderson, Dorothy. 1974. *Universal Bibliographic Control. A Long Term Policy - A Plan for Action*. Munchen: Verlag Dokumentation.

Association of Research Libraries. 2019. *ARL White Paper on Wikidata. Opportunities and Recommendations*.

Balaraman, Vevake, Simon Razniewski, and Werner Nutt. 2018. 'Recoin: Relative Completeness in Wikidata'. In *WWW '18 Companion: The 2018 WebConference Companion*, April 23–27, 2018, Lyon, France. New York, NY, USA: ACM. https://doi.org/10.1145/3184558.3191641.

Bargioni, Stefano, Carlo Bianchini, and Camillo Pellizzari. 2021. 'Beyond VIAF. Wikidata as a Complementary Tool for Authority Control in Libraries'. *Information Technology and Libraries* 40 (2). https://doi.org/10.6017/ital.v40i2.12959

Berners-Lee, Tim. 2006. 'Linked Data - Design Issues'. 27-7-2006. 2006. http://www.w3.org/DesignIssues/LinkedData.html.

Cobb, Simon. 2019. 'Connecting Persistent Identifiers in Wikidata'. In *Portland PID Workshop, 6th May 2019*. https://upload.wikimedia.org/wikipedia/commons/8/82/Connecting_persistent_identifiers_in_Wikidata.pdf.

———. 2020. 'Author items in Wikidata'. Presented at the WikiCiteVirtual Conference, October 26. https://upload.wikimedia.org/wikipedia/commons/c/cc/WikiCite_Virtual_Conference_2020_-_Author_items_in_Wikidata_-_Slides.pdf.

Dunsire, Gordon, and Mirna Willer. 2014. 'The Local in the Global: Universal Bibliographic Control from the Bottom Up'. In *IFLA WLIC 2014*. Lyon, France. http://library.ifla.org/817/.

Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Knudson Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, et al. 2020. 'Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage'. OCLC. 4 May 2020. https://doi.org/10.25333/faq3-ax08.

Gorman, Michael. 2014. 'The Origins and Making of the ISBD: A Personal History, 1966–1978'. *Cataloging & Classification Quarterly* 52 (8): 821–34. https://doi.org/10.1080/01639374.2014.929604.

Hernández-Cazorla, Iván, Manuel Ramírez-Sánchez, and Gregorio Rodríguez-Herrera. 2019. 'Wikidata, WikiCite y Scholia Como Herramientas Para Un Corpus de Datos Bibliográficos Enlazados. Curación y Estructuración de La Producción Científica de Los Investigadores Del IATEXT'. *PRISMA.COM* 40 (2019): 78–87.

IFLA Cataloguing Section and IFLA Meeting of Experts on an International Cataloguing Code. 2016. *Statement of International Cataloguing Principles (ICP)*. Den Haag: IFLA.

Illien, Gildas, and Françoise Bourdon. 2014. 'A la recherche du temps perdu, retour vers le futur: CBU 2.0'. In *IFLA WLIC 2014*. Lyon, France. http://library.ifla.org/956/.

Lemus-Rojas, Mairelys, and Jere D. Odell. 2018. 'Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project Using Wikidata and Scholia'. *Journal of Librarianship and Scholarly Communication* 6. https://doi.org/10.7710/2162-3309.2272.

Linked Data for Production. 2020. 'Wikidata as a hub for identifiers'. Google Docs. 11 June 2020. https://docs.google.com/presentation/d/1jWz3_nCf5rdd-7ejETGlfv99UV2PnD1v/edit?usp=embed_facebook.

Mietchen, Daniel, and Lane Rasberry. 2020. 'Presenting Scholia. A Scholarly Profiling Tool'. Presented at the LD4 Wikidata Affinity Group, August 11. https://docs.google.com/presentation/d/1jJbYSnYSDh36-LxzSpedFyWUzusZAjuBbP-y46ji-0w/edit#slide=id.g35f391192_00.

Nguyen, Ba Xuan, Jesse David Dinneen, and Markus Luczak-Roesch. 2020. 'A Novel Method for Resolving and Completing Authors' Country Affiliation Data in Bibliographic Records'. *Journal of Data and Information Science* 5 (3): 97–115. https://doi.org/10/ghsnkn.

Nielsen, Finn Årup, Daniel Mietchen, and Egon Willighagen. 2017. 'Scholia, Scientometrics and Wikidata'. In *The Semantic Web: ESWC 2017 Satellite Events*, edited by Eva Blomqvist, Katja Hose, Heiko Paulheim, Agnieszka Ławrynowicz, Fabio Ciravegna, and Olaf Hartig, 10577:237–59. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-70407-4_36.

Seidlmayer, Eva, Jakob Voß, Tetyana Melnychuk, Lukas Galke, Klaus Tochtermann, Carsten Schultz, and Konrad Forstner. 2020. 'ORCID for Wikidata – Data Enrichment for Scientometric Applications'. In *Proceedings of The 1st Wikidata Workshop*. https://wikidataworkshop.github.io/papers/Wikidata_Workshop_2020_paper_9.pdf.

Smith, Arthur P. 2020. 'Author Disambiguation'. In *WikiCite 2020 Virtual conference*. https://upload.wikimedia.org/wikipedia/commons/3/38/WikiCite_2020_Author_items.webm.

Unesco/LC Bibliographical Survey. 1950. *Bibliographical Services: Their Present State and Possibilities of Improvement*. Washington: Library of Congress.

Veen, Theo van. 2019. 'Wikidata: From "an" Identifier to "the" Identifier'. *Information Technology and Libraries (Online)* 38 (2): 72–81. https://doi.org/10/ghbj62.

Vrandečić, Denny, and Markus Krötzsch. 2014. 'Wikidata: A Free Collaborative Knowledgebase'. *Communications of the ACM* 57 (10): 78–85. https://doi.org/10/gftnsk.

W3C Incubator Group. 2011. 'Library Linked Data Incubator Group Final Report'. http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/.