Guidelines for using *wikibase.cloud* in a (meta)data management project within libraries

Document type	Deliverable of **Virtual Lab 4, Smart metadata** - *English translation*
Context	**10 Virtual labs for libraries and sustainable development** *organized by* - AIB - Associazione Italiana Biblioteche - Rete delle Reti - EBLIDA
Context references	- Associazione Italiana Biblioteche <https://tinyurl.com/10-lab-AIB>. - Rete delle Reti - EBLIDA - Presentation of the 10 Virtual labs iniziative by Anna Bernabè at the EBLIDA Athens Conference 2022
Authors	- Stefano Bargioni - Giovanni Bergamin - Valentina Lepore
Translated by	- Viviana Vitari - Giovanni Bergamin
Latest version	6 July 2022
Terms of availability	CC BY

Contents

0. Why Metadati intelligenti (smart metadata)?

There is a widespread use of adjectives like *intelligent* or *smart* to form terms such as *artificial intelligence*, *smart cities*, *smartphones*, etc. For the name of Virtual lab # 4, it has been decided to use the same figure of speech – namely the personification: this is why the selected name is *Metadati intelligenti* (*Smart medatada*).

For the purposes of Virtual lab #4, intelligent metadata (or smart metadata) are intended as metadata that contribute to achieve the goals of the 2030 Agenda for Sustainable Development, especially the Goal #4. It is important to note that the organization of knowledge made possible by metadata is a prerequisite for every SDG (Sustainable Development Goal).

The Virtual lab #4was coordinated by Stefano Bargioni, while Giovanni Bergamin and Valentina Lepore acted as facilitators. The purpose of the Virtual lab #4 was to deepen the use of Wikibase as a tool to manage relevant (meta)data for the library services or for GLAM institutions.[1]

Participants (in alphabetical order by name):
1. Angelica Bertellini
2. Anna Molino
3. Cinzia Martone
4. Denise Biagiotti
5. Federica Olivotto
6. Marina Gatti
7. Stefania Puccini

---

[1] <https://en.wikipedia.org/wiki/GLAM_(cultural_heritage)>

The "field trial" method has been adopted, thanks to the availability of a Wikibase instance "smartmeta", installed on wikibase.cloud (initially WBstack).[2]

The results of the overall experimentation are available on

<https://smartmeta.wikibase.cloud/>

All the participants with their own account were able to contribute to the experimentation.

Five work meetings were held and minuted.

1. Why Wikibase and wikibase.cloud matter

Linking two definitions, the first of P. Caplan[3] and the second of K. Coyle[4], we may define the metadata as "structured information according to a certain model (or schema) with the aim of making possible or facilitating a given human activity". For example, catalogue metadata help us find the book we need as well as Google maps metadata can guide us to a chosen destination.

Libraries have a long tradition in this field, although - as many observe - all their potential work in the Web ecosystem is not yet adequately exploited. We are obviously referring to a Web conforming to the vision of *semantic web* and the technological proposal of *linked data*.

The Wikidata initiative was founded in 2012[5] aiming to create a database to collect and control common data of the various language versions of Wikipedia, but also with the aim of serving as a general knowledge base "for the world at large". In other words, it is not

---

[2] See WBStack <https://addshore.com/2019/11/an-introduction-to-wbstack/>.
[3] Priscilla Caplan, *Metadata fundamentals for all librarians*, 2003.
[4] Karen Coyle, *Metadata: data with a purpose*, 2004. <http://www.kcoyle.net/meta_purpose.html>.
[5] Detailed steps <https://addshore.com/2022/02/wikibase-a-history/>.

necessary that a "structured information" respects the canon of "notability"[6] that characterizes Wikipedia: (meta)data can be added in Wikidata, if it refers to a clearly identifiable conceptual or material entity that is described using reliable and publicly available sources. The Wikidata structure has been designed to be fully compliant with the needs of the *semantic web* and the *linked data* technology. In addition Wikidata is designed to be a collaborative tool for everyone.

In 2018, Wikibase, which is the Wikidata software engine, was released open source. In this way not only you can contribute to Wikidata, but you can also install your custom instance of wikibase and populate it with (meta)data related to a specific domain.

The interest of the world of libraries for both Wikidata and Wikibase has grown over time. In libraries there are many initiatives that use Wikidata and Wikibase. IFLA created a work group at the end of 2019 to explore "the integration of Wikidata and Wikibase with library systems and the alignment of Wikidata ontology with library metadata formats such as BIBFRAME, RDA and MARC".

At the end of 2020, Deutsche Bibliothek and Wikimedia Deutschland published the *Wikilibrary manifesto*[7] that proposes collaboration between libraries and the wiki world for an open network for art, culture and sciences based on linked data. The proposal is still open to all institutions interested in a data ecosystem based on FAIR principles.[8] The first section of the Manifesto sets forth principles; the second part is dedicated to the measures that make principles enforceable. In particular, the Wikibase project is identified as a promising technical infrastructure to store, edit and exchange (meta)data. On June 25th, 2021 the Manifesto was also signed by IFLA. [9]

---

[6] Notability in Wikipedia: <https://en.wikipedia.org/wiki/Wikipedia:Notability>; notability in Wikidata: <https://www.wikidata.org/wiki/Wikidata:Notability/en>
[7] <https://www.wikimedia.de/the-wikilibrary-manifesto/>
[8] <https://www.go-fair.org/fair-principles/>
[9] <https://www.ifla.org/node/93952>

A significant shift in the way Wikibase is managed came in October 2021 at the WikidataCon 2021, with the announcement by Wikimedia Deutschland (WMDE) of the launch of wikibase.cloud.

The *Linked Open Data/Strategy 2021* proposal[10] presents the new development strategy for linked open data.

This is the vision inspiring the WMDE strategy:

> «We work towards a future in which people share the power to collect and organize the data that shapes humanity's understanding of the world.
>
> Diverse communities around the world participate in Wikidata and in a network of specialized Wikibases, co-creating an open and free global knowledge graph in a thriving Linked Open Data web.
>
> This is the basis upon which people, companies and institutions of all sizes can generate new insights, build new apps and services, and change the world for the better».[11]
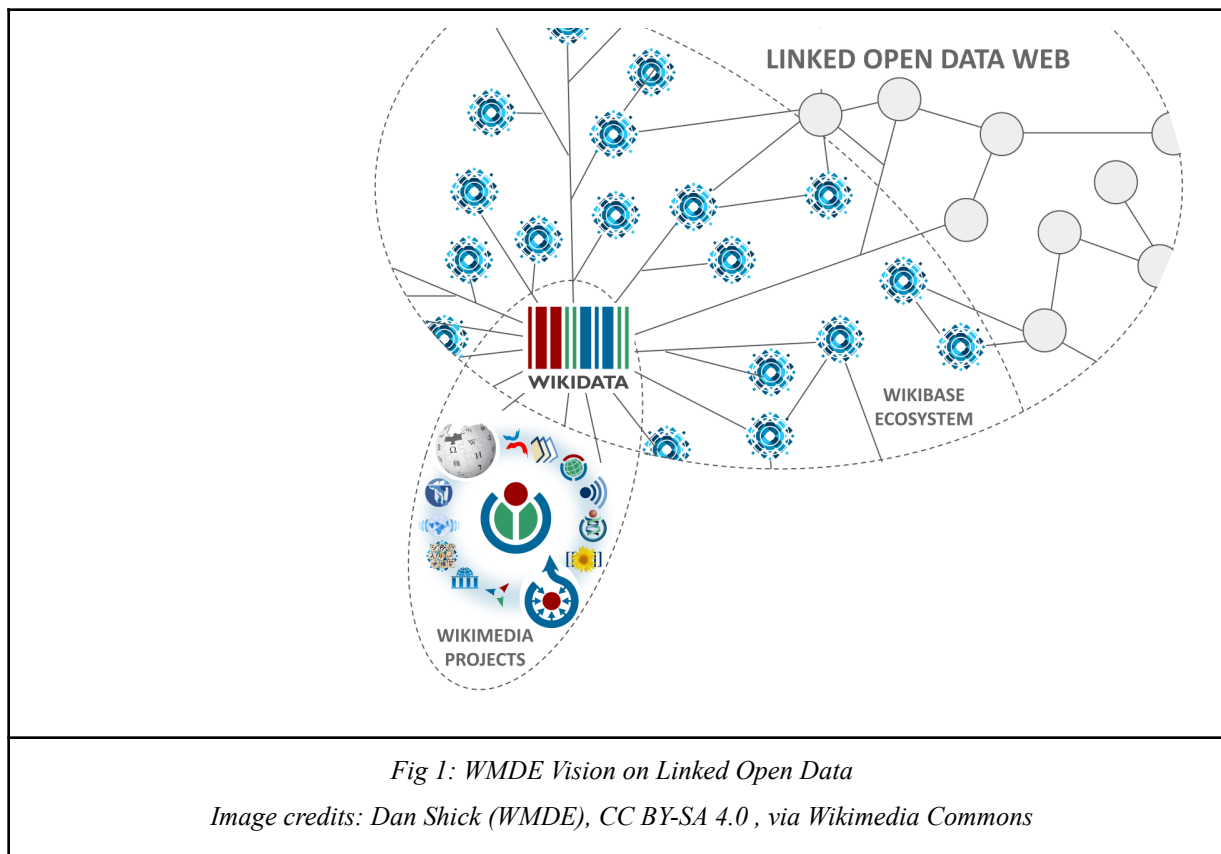
It shows two strategic lines:

1. the enhancement of Wikidata and its role of providing "general purpose data about the world" (not to mention the clearly emerged awareness of the *WikidataCon 2021* about the impossibility of managing *all (meta)data in the world* with a single service);[12]
2. the promotion of widespread creation of a network of specialist Wikibase instances by providing:
   - Wikibase suite: the opportunity for everyone to install, manage and customize their own expert instance of wikibase;
   - Wikibase.cloud: the opportunity for everyone to enjoy a Wikibase instance in Saas mode, directly managed by WMDE in the cloud.

---

[10] <https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021>
[11] < https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021/Joint_Vision>
[12] <https://tinyurl.com/3zt8z2bm>

Fig 1: WMDE Vision on Linked Open Data
Image credits: Dan Shick (WMDE), CC BY-SA 4.0 , via Wikimedia Commons

Wikidata and independent wikibase instances (*suite* and *cloud*) will form the *wikibase ecosystem* which is a candidate to be a fundamental part of the linked open data web.

As planned, from February 2022 the Wikibase.cloud service is operational.[13] From June 2022 it is possible to book, in order to get an invitation.

2. Using Wikibase today

Without the claim of completeness, here are some data on the Wikibase initiative[14]:

---

[13] <https://meta.wikimedia.org/wiki/Wikibase/Wikibase.cloud>
[14] Source: <https://tinyurl.com/3hcpdnpe>.

● Wikibase for scientific research: about 63 projects in progress;

● Wikibase in general: 148 users and 510 Wikibase instances on WBstack (now wikibase.cloud);

● WMDE dedicates 14 full-time people to the Wikibase project;

● The community interested in the Wikibase development has over 200 members on Telegram Wikibase group and 182 developers contribute directly to Github;

● A group, Wikibase Stakeholder Group, currently made up of 18 organizations and 24 personal members, collaborates directly in the development, planning and growth of the Wikibase initiative.

Here is a selected list of ongoing implementations:

1. Enslaved
   ● "Search over numerous datasets and reconstruct the lives of people involved in the historical slave trade. Browse interconnected data, generate visualizations, and explore short biographies of enslaved and freed peoples."
   ● Wiki interface

2. Artbase
   ● «Welcome to the Rhizome ArtBase. This is an archive of born-digital artworks from 1999 to the present day. A small sample of artworks are shown below; you can also browse the archive by date or by artist name. »
   ● Notes:
      o It is based on a customization of MediaWiki for the user interface;
      o Let's see, for example, the SPARQL query Artworks in the ArtBase with more than one artist, including the artist names (visualized as a graph with images).

3. Kohesio

"Kohesio: discover EU projects in your region"

- The Wikibase installation that supports Kohesio

4. Wikibase in the GLAM world

- The FNE (BnF and Abes) Wikibase project for the national authority file.
- The NOEMI (BnF) project uses Wikibase as an internal cataloging tool (over 300 cataloguers). The acronym NOEMI stands for "Nouer les Œuvres, Expressions, Manifestations et Items";
- GND: 1) recording of "How to open Authority Control system - The GND & Wikibase (WikidataCon 2021)" 2) slides 30.6.2022;
- Lexbib: «Welcome to LexBib wikibase at Elexis. LexBib is a digital bibliography project for the domain of Lexicography and Dictionary Research. General aims are LOD-ification of publication metadata, and content description of metalexicographic article full texts, as a service to the lexicographic community. LexBib wikibase brings together LexBib Zotero bibliographical data and a vocabulary of lexicographic terms, LexVoc. We are working on extending the scope of LexBib towards Dictionaries.»
- Smithsonian Libraries and Archives & Wikidata: Using Linked Open Data to Connect Smithsonian Information.

3. What to do with Wikibase and Wikibase.cloud

With a Wikibase installation (based on Wikibase suite or Wikibase.cloud) a project on (meta)data has at its disposal:

1. a graphic interface (GUI) to update a database in collaborative mode, with data version control tools;

2. a database organized as a triplestore, that is specifically built for saving and recovering subject-predicate-object triples. A very interesting peculiarity is that the *Wikibase data model* provides that every statement expressed in the form of a triple can be enriched with qualifiers and/or references.[15] From a technical point of view - as provided by RDF - each triple can be "reified" and play the role of the subject of another semantic triple;[16]

3. a SPARQL endpoint with a graphic interface (GUI) to access to the stored triples;

4. the possibility to search for the data entered through integration of Elasticsearch;

5. APIs to access Wikibase through reading and writing programs;

6. availability of scripts such as PyWikibot, WikidataIntegrator;

7. availability of easy-to-use tools for importing or exporting data (especially massively):

    a. Quickstatements;

    b. Cradle;

    c. OpenRefine reconciliation service (not yet available on wikibase.cloud)[17].

4. The Wikibase (meta)data model

The *Wikibase data model* is the conceptual model that presides over the centralized storage and management of the information of entities that, starting from the Wikidata repository, populate all Wikimedia projects. It should be particularly emphasized, that the *Wikibase data mod*el outlines a method of structuring data applicable to each data type, as long as it is single and disaggregated, as well as expressible as an atomic entity susceptible of interrelation with other atomic entities which can be in turn processed according to the *Wikibase data model*. In short, the *Wikibase data model* can be adopted to describe any

---

[15] This is a fundamental requirement for Wikidata. In order to be used as a secondary source, Wikidata must document its (triple) statements with the qualifiers and the references of a given statement or a triple one. For example (see https://www.wikidata.org/wiki/Q38), the triple statement "Italy has 50,199,700 inhabitants" has a qualifier in 1960 and a reference in World Bank Open Data, while the triple statement "Italy has 60,317,000 inhabitants" has a qualifier in 2020 and a reference in ISTAT.

[16] For Wikibase as a (meta)model, see paragraph #4. For reification see also http://ceur-ws.org/Vol-1457/SSWS2015_paper3.pdf.

[17] Source with integration: <https://tinyurl.com/3hcpdnpe>.

object of any scope of reality, therefore it can also be used for the description of any entity that lies within the library world, or, in a broader sense, of the management of institutional resources coming from GLAM institutions.This is made possible because it is open and not ontologically limited, that is interoperable.

It must be said, to be more precise, that the *Wikibase data model* is a metamodel rather than a conceptual model. That is to say, compared to common data models, the *Wikibase data model* consists of a further level of abstraction, due to which, on the level of representing reality with interrelated concepts, it involves only the definition of the syntactic macrostructure (i.e. the way how entities of interest relate to each other). It does not establish either an ontology or a reference domain, that, at the semantic level, may constraint the contents to be managed.

In other words, the *Wikibase data mod*el only provides instructions on how to manage information on contents that are relevant for structuring meaningful and complete statements. In addition, the contents deemed relevant are identified and reconfigured from time to time by Wikidata contributors in a scalable and bottom-up way; they are described and interrelated by properties that are also not predefined. This allows the maximum semantic interoperability, that is the possibility to formulate and interpret any statement on any object, because the *Wikibase data mode*l does not impose from above, in a top-down mode, any dominant vocabulary.

5. How to use Wikibase in a project

Wikibase is convenient, if the final goal is to publish Linked Open Data. There does not seem to be such a tool freely available that is not linked to a particular commercial supplier and that allows the immediate publication of Linked Open Data.

Normally the tools for publishing LODs are related to a particular domain (they have a predetermined ontology), or they are the result of a recurrent conversion in LOD of (meta)data created using legacy platforms. .

There are open tools to build ontologies such as Protégé. However, there are no tools that allow you to progressively construct the ontology while entering data. This also entails risks to be assessed, but in principle the collaborative construction of an ontology is a reasonable goal: today, the "items" created in Wikidata are around 98 million (with 1.6 billion updates or "edits").[18]

Wikibase has a very low entry barrier, if compared to other tools in the world of Linked Open Data.

What might be missing from Wikibase:

- A controlled and ergonomic interface for data entry:
  - there are tools such as ShEx and Cradle that offer a customized input interface without the need of external programs;
  - if there are special needs, the input interface can be done with a bespoke application that uses the APIs to access Wikibase through reading and writing programs (see for instance Enslaved and Kohesio).

Basic computer skills are required for the installation of Wikibase, but to use wikibase.cloud (management, configuration, design of new features) training courses are useful (maybe on the same style of the Library Carpentry proposals). There are many possibilities offered on the Net.

In general, it is known that open source software developed in research projects and European projects is at risk when projects end: there is no funding to develop and maintain

---

[18] < https://www.wikidata.org/wiki/Wikidata:Statistics> (1.6.2022).

the software. The choice of designing the software within the Wikibase ecosystem allows you to protect your investment (if you follow the rules of the open source software development the community takes care of it and maintains it).

6. Overview

6.1. The issue

- libraries usually have specific tools to manage metadata relating to a given domain;
- there is a lack of tools to manage the special features of an institution (e.g. types of collections) or to launch special initiatives (e.g. thematic repertories, collection of data documenting events);
- there is a lack of tools to semantically interoperate with other domains.

6.2. Unsustainable alternatives

- to force the existing applications, oriented to the bibliographic metadata domain, to deal with other types of metadata;
- bespoke solutions involve the risk of the loss of overall control.

6.3. The Wikibase.cloud solution

- institutions have no direct costs to acquire and maintain basic technologies: this is provided as a service (Wikibase as a Service);
- it allows to be part of a widespread metadata management ecosystem (federation and use of semantic web technologies);
- wiki model: very low barrier of entry, comments, transparency of changes, possibility of recovery.

6.4. Expected benefits

- good ratio of (human, financial, etc.) resources invested and results;

- in case of problems it could be easier to find solutions given the widespread of wikibase instances;
- sustainable model of linked open data and of respect for *ontodiversity*[19];
- increasing of the cultural participation and of the reuse through one of the largest (meta)data management ecosystems.

6.5. Use cases and types of users

1. a user searching for and getting only web pages as a result (no need to know how wikibase works);
2. a user using SPARQL functionalities;
3. machines (software agents) that query, index data (e.g. search engines) or reuse data for other services, applications and users;
4. a user updating via wikibase interface;
5. a user updating via quickstatements interface (massive data import, modification and deletion);
6. a user updating via freely customizable forms with controls on the data entered (e.g. cradle).

6.6. Cost structure

- wikibase.cloud is free of charge if the data are freely available (each Institution can decide in any case who can update the data);
- optional costs may relate to the customisation of end user interfaces or data entry interfaces;
- if an institution decides not to use wikibase anymore, all metadata entered in a wikibase instance can be downloaded in standard formats and serializations at any time.

---

[19] This neologism is here created on the model of *biodiversity*. Every semantic web ontology is a point of view: Wikibase can host the sharing and reuse of data between different points of view or ontologies.

6.7. Economic sustainability

All investments in the improvement and customization of wikibase can be shared following the best practice of the open source software development, thus improving the overall economic sustainability of all wikibase adopters.

7. Basic references

7.1 Official documentation

- Wikibase official website https://wikiba.se/
- Intro video by Wikimedia Deutschland
  https://www.youtube.com/watch?v=dCAjhjeJpgY
- Wikibase data metamodel
  https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer;
- Wikibase suite https://www.mediawiki.org/wiki/Wikibase/Install;
- Wikibase cloud https://wikibase.cloud

7.2 Other references

Lepore, Valentina. 2019. *Il Wikibase data model per la pubblicazione dei dati bibliografici sul web semantico. Una sperimentazione presso la Biblioteca Nazionale Centrale di Firenze.* II level Master thesis, Università degli studi di Firenze. http://eprints.rclis.org/39494/

Mauri, Luca. 2021. *Wikibase, starting from scratch* https://www.youtube.com/watch?v=7kUUarBN2vM. Come installare da soli la suite Wikibase/ How to install the Wikibase suite yourself.

Rossenova, Lozana. 2022. *Examining Wikidata and Wikibase in the context of research data management applications.* https://tinyurl.com/3hcpdnpe.

Shorland, Adam. 2022. *Wikibase a history*. https://addshore.com/2022/02/wikibase-a-history/