Using Wikidata integration on the Wikimedia projects to enhance GLAM-WIKI content sharing

Mike Peel & João Alexandre Peschanski
Wiki Movimento Brasil
(MP is also WMF Board of Trustees, but not wearing that hat for this presentation!)
GLAM conundrum

Releasing media content from GLAM on Commons has been really successful and important over many years.

However, **it's important it doesn't only exist on Commons in a vacuum.** Instead it must be integrated into the rest of the Wikimedia projects so that it is visible and widely used.
GLAM methodology

- Institutional outreach and validation
- Legal analysis and copyright guidelines
- Digitization and upload pipeline
- Content use initiatives and metrics
Through Wikidata, content that GLAMs share can automatically be reused and made visible across many projects in many languages.
Wikidata is an open collaborative knowledge base

It is a fast-growing Wikimedia project, especially relevant for academic and cultural contributions.

Welcome to Wikidata

the free knowledge base with 107,613,852 data items that anyone can edit.

Introduction • Project Chat • Community Portal • Help

Want to help translate? Translate the missing messages.

(# of items will already be out-of-date!)
GLAM Wiki 2023 (Q18978159) QIDs

Labels

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
<th>Also known as</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>GLAM Wiki 2023</td>
<td>No description defined</td>
<td>GLAM Wiki Conference 2023</td>
</tr>
<tr>
<td>Spanish</td>
<td>Conferencia GLAM Wiki 2023</td>
<td>No description defined</td>
<td>GLAM Wiki Conference 2023</td>
</tr>
<tr>
<td>French</td>
<td>GLAM Wiki Conference 2023</td>
<td>No description defined</td>
<td>GLAM Wiki Conference 2023</td>
</tr>
<tr>
<td>Portuguese</td>
<td>No label defined</td>
<td>No description defined</td>
<td>GLAM Wiki 2023</td>
</tr>
</tbody>
</table>

Linked values

instance of

log image

Properties

Media

Interwikis

Wikipedia

Wikibooks

Wikinews

Wikiquote

Wikisource

Wikiversity

Wikivoyage

Wiktionary

Multilingual sites

commons

meta
The growth capacity of the Wikimedia movement is directly related to its capacity of sharing common values across stakeholders.

Alves et al, *The Technical Infrastructure of Cultural Initiatives on Wikimedia, 2022*

ESSACHESS (PDF available online)
Started in 2018, bot-deployed in 4.7 million Commons categories!

Small infobox to add context for category contents, without hiding category contents

Using Wikidata information only (no local definitions) - completely multilingual code: `{{Wikidata Infobox}}`

Images on Wikidata shown alongside Commons category - are best images selected? (+ night image, map, ...?)

Authority control at the bottom (none for FIC - please improve it!)

Continuous development with community input and suggestions
Гіантський український радіотелескоп [сюжет]

Незвичайно великий радіотелескоп, що буде відкритий 1984 року в Південній Америці. Якщо він працюватиме, то буде одним з найкращих у світі.

**Lydia Pintscher**

*Deutsche Informatikerin*

**Lydia Pintscher** 2015

**Medium hochladen**

**Geburtsdatum:** 20. Juli 1984

**Land der Staatsangehörigkeit:** Deutschland

**Besuchte Bildungseinrichtung:** Karlsruher Institut für Technologie (Diplom, Informatik, 2000–2011)

**Tätigkeit:** Produktmanager (Wikidata), Informatiker, Wikimedianer, Forscher

**Arbeitgeber:** Wikimedia Deutschland

**Mitglied von:** KDE

**Arbeitsgebiet:** Wikidata

**Sprache:** Deutsch

**Auszeichnung:** Academy Award - Best Non-Application Contribution (2012)

**Offizielle Website:**

**Normdatei:**

**Ändern**

**Download**

**Kommentare:**

**Derivative:**

**License:**

**Attribution:**

**Creative Commons**

**Nominierung:**

**Sitenlink:**

**Create new Wikidata item**

**Upload media**

**Extension: Article Placeholders**

*Mediawiki extension to create automatic article placeholders*

**Authority control**

*Q1676493*

**Reasonator**

*Petticus* • Schlottla • Statistics • OpenStreetMap • Locator tool • Wikibooksde • Ключи зложені

**MediaWiki**

*Search deployed*

**NO WIKIDATA ID FOUND!**

*Search for Ukrainian Wikiversity on Wikidata*
- More complicated on Wikipedias
- Many different infoboxes covering different topics, different layouts etc.
- Consensus varies between Wikipedias (wide deployment on Catalan, tricky on English, others mixed).
- Some Wikipedias require referenced Wikidata information only (please add your references!)
- Automatically update when content is added to Wikidata. Add a picture on Wikidata -> it immediately appears in infoboxes across many languages.
- Improves cross-Wikimedia integration.
- Can quickly create lists using Wikidata information, based on Wikidata queries, that are transposed on (some) mainspace Wikipedias
- Whatever you can fetch using a Wikidata query, you can turn into a list (Listeriabot)
- As Wikidata updates, so does the list
- Works across different languages using the same dataset
Wikidata queries

- SPARQL querying is powerful, but writing them in the right format can be tricky
  - Wikidata:Request_a_query can help!
- Only Wikidata content: can't edit the lists locally, and local changes will be lost
- Querying is powerful in general for GLAMs to identify and visualize trends in their collections (web semantics)

query.wikidata.org
Cite Q for automatically generated citations

- Currently millions of Wikidata items for journal articles - part of Wikicite
- Can use this information to build structured references using Wikidata QID
- Can be used in any language!
- Automatically link to DOIs, author names, etc.
- Can auto-update all, e.g., preprint -> publication, or if a paper is withdrawn/replaced

---


<ref>{{Cite Q|Q56603073}}</ref>

---

Structured Data on Commons

- Wikimedia Commons also has structured data
- Defined for each file, includes caption & properties
- Stores information about copyright, camera, location, depicts, etc.
- Used by file page templates, e.g., {{Art Photo}}
- Queryable at https://commons-query.wikimedia.org/ (but need to be logged in)
Wikidata holds many exciting possibilities, and opens the door for a variety of new research opportunities and potential applications across all areas of sciences, technology and cultures.

Evenstein, Nachmias, Investigating the potential of the semantic web for education: Exploring Wikidata as a learning platform. 2023
Education and Information Technologies (PDF available online)
- Information contributed to Wikidata is extracted using queries and added into templates that generate a WikiBook

- A collaborative catalogue, a portal that structures and integrates apps and tools in which contributions happen more easily; it also connects related content
An ecosystem of Wikidata apps

- Wikidata-powered apps are a user-friendly solution for crowdsourcing cataloguing and content structuring.
- Apps can be designed to answer specific institution questions, support cataloguing gaps and support public engagement with collections.
Página inicial  Sobre  Ajuda  Outros aplicativos  Autenticar-se

WIKI MUSEU DO IPIRANGA

How many?

Obra  São Paulo - Jardim da Luz
Criador(a)  Guilherme Goensly
Data  1903
Instância de  collotype technique; printing; postcard
Materiais utilizados  collotype technique

Quantos itens de cada descriptor você vê?

Jardim da Luz: park in São Paulo
human: any member of Homo sapiens, unique extant species of the genus Homo, from embryo to adult
tree: perennial woody plant

CONFIRMAR  PULAR  INÍCIO

Veja essa imagem no Wikimedia Commons

WIKI MUSEU DO IPIRANGA

Com que roupa eu vou?

Faça a sua autenticação clicando no botão 'Autenticar-se' no canto superior direito

Obra  Retrato de Martim Francisco Ribeiro de Andrade
Instância de  pintura
Data  1925
Criador(a)  Oscar Pereira da Silva
Material  tinta a óleo
Versão audível

Adicione elementos descritivos às pessoas retratadas
homem  Veste
Usa

Veja essa imagem no Wikimedia Commons

GLAM WIKI CONFERENCE 2023
Bibliometric and scientific impact dashboard

- Scholarly-related content (publications, author information) can be transformed into dashboards for easy visualization and trends
- Strongly connected to open-science values and practices
- Dashboards can be adapted to fit specific data requests, creating an institutional profile
- [https://scholia.toolforge.org/](https://scholia.toolforge.org/)

Antonio Galves (Q17489997)
Jefferson Antonio Galves was a Brazilian mathematician, professor of the Institute of Mathematics and Statistics of the University of São Paulo (USP) and member of the Brazilian Academy of Sciences. His field of studies was related to statistical models, in particular models that have stochasticity and variable range of memory. Galves was also the leader of Neurosat, a research center established in 2013 at USP that is dedicated to integrating mathematical modeling and theoretical neuroscience. [Read more on English Wikipedia]

[https://orcid.org/0000-0001-6757-716X](https://orcid.org/0000-0001-6757-716X)

NeuroMat (Q18477654)
The Research, Innovation, and Dissemination Center for Neuromathematics is a Brazilian research center established in 2013 at the University of São Paulo that is dedicated to integrating mathematical modeling and theoretical neuroscience. Among the core missions of Neurosat are the creation of a new mathematical system to understanding neural data and the development of neuroscientific open-source computational tools, keeping an active role under the context of open knowledge, open science and scientific dissemination. The research center is headed by Antonio Galves, from USP's Institute of Mathematics and Statistics, and is funded by the São Paulo Research Foundation (FAPESP). As of 2019, the co-principal investigators are Osvaldo Batista Filho (USP), Pablo A. Ferrani (USP/IBA), Fernando da Paixão (UNICAMP), Antônio Carlos Roque (USP), Jorge Stolfi (UNICAMP), and Claudia G. Vargas (UFRJ). Ernst W. Hamburger (USP) was the former director of scientific dissemination. Neurosat's International Advisory Board consists of David R. Brüning, Leonardo G. Cohen (NIH), Markus Diesmann (Jülich), Francesco Guerra, Wojciech Szpankowski (Purdue). [Read more on English Wikipedia]

[http://neuromat.numec.prp.usp.br/](http://neuromat.numec.prp.usp.br/)
- Interactive timelines of nearly everything!
- Wikidata back-end, interactive display
- [http://histropedia.com](http://histropedia.com), example: [http://histropedia.com/timeline/41trtpg9bg0t/Botanists-active-in-South-America](http://histropedia.com/timeline/41trtpg9bg0t/Botanists-active-in-South-America)
Monuments database is now on Wikidata
Easy to expand with new monuments, bringing lists e.g., from different municipalities, together for the first time
Automatic monument lists & maps, etc.
Using {{MonumentID}} on Commons - can discover uploads with no image on Wikidata
Wikidata connection leads to more diverse contributions and image uses; possibly the largest cultural heritage repository ever made
Roundtripping online repositories and Wikidata

(Info from Dalton Martins / IBRAM)

- Wikidata can be integrated into online GLAM repositories (i.e., Tainacan in Brasil) in a federated data circuit

- We are facing four main challenges for the development of the circuit ([FAPESP 21/06767-8]):
  - Data reconciliation
  - Workflow for publishing data (creating items and uploading images via Tainacan)
  - Monitoring workflow and contributions
  - Monitoring data reuse (indicators, metrics)
• Regular labs focused on different Wikidata topics (much more detail than covered here!)
• Initial presentation + afternoon workshop
• 38 so far, in either Portuguese or English: https://www.wikidata.org/wiki/Wikidata:Wiki data_Labs (including recordings!)
• 39th on Tuesday 21st November, hybrid (São Paulo/online) on the Wikimedia ecosystem: https://w.wiki/85J$
Getting content into Wikidata

In general:

- Pywikibot (Python library) - https://www.mediawiki.org/wiki/Manual:Pywikibot
- OpenRefine (see other talks about this!)

For journals:

- Zotero to export to QuickStatements format, then import to Wikidata (See https://w.wiki/pi2) - https://www.wikidata.org/wiki/Wikidata:Zotero
- Import using SourceMD using PMID or DOI (use the old version; not ISBN or ORCID at the moment) https://www.wikidata.org/wiki/Wikidata:SourceMD
Can also 'gamify' adding info to Wikidata, see: https://wikidata-game.toolforge.org/distributed/

Latin America in Wikidata Contest 2023 - GLAM

In the 2023 edition of the Latin America in Wikidata contest, we invite you to contribute to Wikidata by improving content related to cultural heritage in Latin America.

There are two game modes. In this case, you will have to complete the days in which the heritage institutions are open to the public.

If you know the correct answer but it does not appear among the options, you can touch the name of the institution to go to the element in Wikidata and edit this property manually. The manual editions in the account for the contest but are a great contribution.

The contest is from November 14, 2023 at 00:01 UTC on December 14, 2023 at 23:59 UTC, until the items are closed available. It arises from the collaboration between Wikimedia Argentina, Wikimedistas de Bolivia, Wikimedia Chile, Wikimedia Colombia, Wikimedia Mexico, Wikimedistas de Uruguay and Wikimedia Venezuela.

Duas Rodas Museum [Q6694401]

Museu Duas Rodas

Museum

Sitio web oficial

http://www.museuduasrodas.com.br

<table>
<thead>
<tr>
<th>lunes a jueves</th>
<th>lunes a viernes</th>
<th>lunes a sábado</th>
<th>viernes</th>
<th>sábado</th>
<th>domingo</th>
<th>lunes</th>
<th>martes a viernes</th>
</tr>
</thead>
<tbody>
<tr>
<td>martes a domingo</td>
<td>sábado y domingo</td>
<td>Siguiente</td>
<td>Ninguna opción es la correcta</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Getting content out of Wikidata

On-wiki:

- Lua Scribunto interface
  - https://www.mediawiki.org/wiki/Extension:Wikibase_Client/Lua
- Can access directly in Lua codes, or in templates
  - {{#statements:P1476|from=Q191380}} will display "Notre-Dame de Paris"
- Module:WikidataIB - flexible Wikidata module

Off-wiki:

- API: https://www.mediawiki.org/wiki/API:Main_page
- Wikidata Query Service: https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service
- Pywikibot (Python library) - https://www.mediawiki.org/wiki/Manual:Pywikibot
What we still need

- Better Wikidata-Wikimedia integration that is easier for more people to use
- Wikidata metrics - usage, quantifying improvements by activities
- More training opportunities to share Wikidata knowledge and how-to
- More scholarly publications - when you do something cool with Wikidata, publish about it and share best-practices with others!
- Conversations around GLAM and Wikidata ethos, for example, around decolonisation data modeling (Indigenous Artists and Wikidata)
Questions? And, what are we missing? What are possible future opportunities for Wikidata integration?
Thank You :-}