Beyond Wikipedia

Cultural heritage and LOD on the Wikimedia platforms

Alicia Fagerving
alicia.fagerving@wikimedia.se

WIKIMEDIA
SVERIGE

April 1, 2021
Linnæus University

Wikimedia Sverige

- A non-profit organization based in Sweden
- We work towards making knowledge accessible to more people, for free, and under free licenses
- More about us: wikimedia.se

- We use the Wikimedia platforms as our most important resource
- We support educational and cultural heritage institutions in their work with free knowledge

WIKIMEDIA
SVERIGE

Into Wikidata!

Wiki?

The **Wikimedia movement**: that's all of us!

MediaWiki: the software behind everything

Wikimedia Foundation: runs and develops the websites

Local chapters

wikimedia.se/engagera-dig

GLAM

Galleries, **L**ibraries, **A**rchives, **M**useums

ABM: **A**rkiv, **B**ibliotek, **M**useer

Wikidata

Launched in 2012 – a young project that is developing rapidly.

Over 90 million data items.

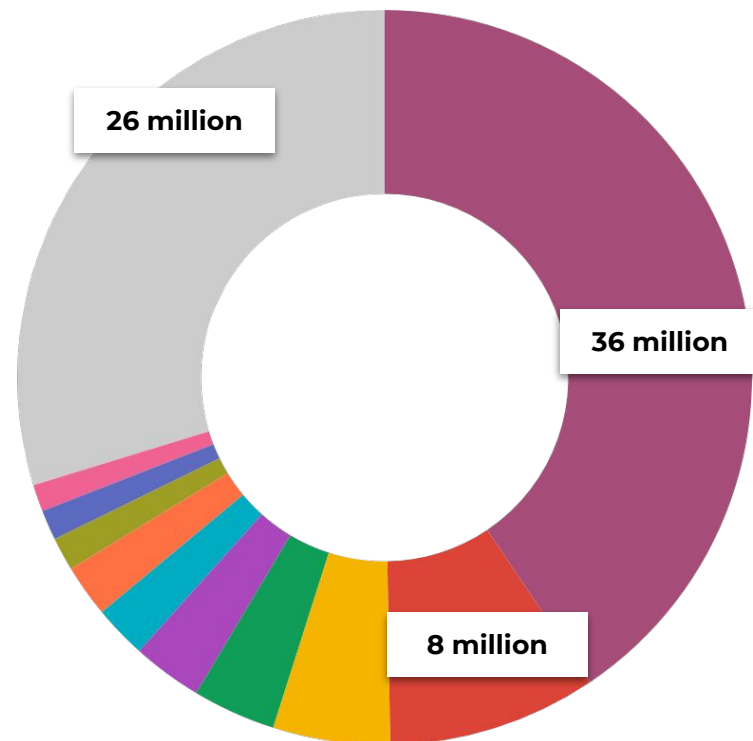Aiming to collect the sum of human knowledge in a structured and linked form.

What does Wikidata actually contain?

The goal of Wikidata is to collect the sum of human knowledge – in a structured, linked form.

With 90 million items, you might think we are close to achieving this…

- scholarly article (Q13442814)
- human (Q5)
- Wikimedia category (Q4167836)
- star (Q523)
- taxon (Q16521)
- galaxy (Q318)
- review article (Q7318358)
- Wikimedia disambiguation page (Q4167410)
- gene (Q7187)
- chemical compound (Q11173)
- other

26 million

36 million

8 million

WIKIMEDIA
S V E R I G E

Source: SPARQL query, 2020-10-12

https://w.wiki/gVM

Scholarly articles on Wikidata

These items were created by importing data from open databases of bibliographic metadata.

Importing data is a major part of the work Wikidata community does.

Wikidata is licensed CC0, this means other datasets with a compatible license can be imported.

PubMed was a major source, leading to an overrepresentation of medicine/life sciences.

WIKIMEDIA
SVERIGE

Multiple Metal-Metal Bonds in Iron-Chromium Complexes (Q59464209)

scientific article published on 11 March 2013

▾ In more languages
Configure

Language	Label	Description	Also known as
English	Multiple Metal-Metal Bonds in Iron-Chromium Complexes	scientific article published on 11 March 2013	
Swedish	No label defined	No description defined	
Finnish	No label defined	No description defined	
Tornedalen Finnish	No label defined	No description defined	

All entered languages

Statements

instance of	scholarly article	✎ edit
	▸ 1 reference	
	+ add value	

title	Multiple metal-metal bonds in iron-chromium complexes (English)	✎ edit
	▸ 1 reference	
	+ add value	

author	Laura Gagliardi	✎ edit
	series ordinal 5	
	▸ 1 reference	
	Connie C. Lu	✎ edit
	series ordinal 6	
	▸ 1 reference	
	+ add value	

Scholia

scholia.toolforge.org

Wikidata profiles of scholars, organizations, research topics, publications and related concepts.

Scholia is part of the WikiCite initiative - wikicite.org

organization | location | topic

Linnaeus University (Q1751928)

Linnaeus University (LNU) (Swedish: Linnéuniversitetet) is a state university in the Småland region of Sweden. It has two campuses, one in Växjö and one in Kalmar. Linnaeus University was established in 2010 by a merger of former Växjö University and Kalmar University (Högskolan i Kalmar), and is named in honour of the Swedish botanist Carl Linnaeus. ... (from the English Wikipedia)

Related: Växjö University · University of Gothenburg · Q10549212 · Örebro University · Ministry of Defence · Salle Richelieu · Chalmers University of Technology · Tropenmuseum · Stockholm School of Economics · Kooperativa förbundets arkitektkontor

🐦 @linneuni | ROR 00j9qag85

Employees and affiliated

Past and present employees, affiliated, and members

Show 10 ▾ entries   Search:

Works	Wikis	Researcher	Researcher description	Orcid
125	0	Jonas Waldenström	researcher	0000-0002-1152-4235
79	0	Alf Månsson	researcher	0000-0002-5889-7792
54	0	Cornelius Holtorf	archaeologist	0000-0002-0557-9651
38	0	Daniel Lundin	researcher ORCID ID = 0000-0002-8779-6464	0000-0002-8779-6464
35	0	Janusz Sadowski	researcher ORCID ID = 0000-0002-9495-2648	0000-0002-9495-2648
30	0	Carl Rikard Unelius	researcher	0000-0001-7158-6393
23	1	Stefan Gössling	Swedish academic	
21	0	Anders Bremer	researcher	0000-0001-7865-3480
21	0	António Filipe Teixeira	researcher ORCID ID = 0000-0003-3436-2010	0000-0003-3436-2010

WIKIMEDIA
SVERIGE

Anyone can edit!
... do they?

Whose knowledge?

The average Wikipedian is:

- a white male
- technically inclined
- formally educated
- aged 15–49
- from a majority-Christian country
- from a developed nation
- from the Northern Hemisphere
- likely employed as a white-collar worker or enrolled as a student rather than being employed as a blue-collar worker

WIKIMEDIA SVERIGE

Which languages do Wikidatans speak?

Wikidata editors can add their language skills to their user page.

- Over 50% speak English.
- Most popular other languages: French, German, Spanish, Russian.

Wikidata is multilingual – but English is the de facto language of community discussions and documentation.

Babel user information

fi-N	Tämän käyttäjän äidinkieli on suomi.
en-4	This user has near native speaker knowledge of English.
sv-2	Den här användaren har medelgoda kunskaper i svenska.
fr-2	Cette utilisatrice dispose de connaissances intermédiaires en français.
nb-1	Denne brukeren har grunnleggende kjennskap til norsk (bokmål).
nn-1	Denne brukaren meistrar nynorsk på eit grunnleggjande nivå.

Kaffee, L.-A., & Simperl, E. (2018, August). Analysis of editors' languages in Wikidata. *Proceedings of the 14th International Symposium on Open Collaboration* (pp. 1–5).

WIKIMEDIA
S V E R I G E

Which languages do Wikidatans speak?

Items can have labels in multiple languages; their distribution can tell us something about the editors' knowledge and interests.

- English: 11%
- Dutch: 6%
- French: 6%
- ....
- Swedish: 4%
- Chinese: 1%

The Semantic Web has the potential to improve the language imbalance of the Internet, but suffers from the same problems.

▼ In more languages

Language	Label
English	National Library of Sweden
Danish	Kungliga biblioteket
German	Königliche Bibliothek zu Stockholm
Spanish	Biblioteca Nacional de Suecia
Finnish	Ruotsin kuninkaallinen kirjasto
Polish	Kungliga biblioteket
Russian	Национальная библиотека Швеции
Swedish	Kungliga biblioteket

Kaffee, L.-A., Piscopo, A., Vougiouklis, P., Simperl, E., Carr, L., & Pintscher, L. (2017, August 23). A Glimpse into Babel. *Proceedings of the 13th International Symposium on Open Collaboration* (pp. 1–5)

WIKIMEDIA
S V E R I G E

GLAM institutions and Wikimedia

A lot of GLAMs around the world have experience from working with the Wikimedia projects since before Wikidata became a thing, such as by uploading digitized artworks to Wikimedia Commons.

The emergence and development of Wikidata, has created new possibilities to not only add their collections to Wikimedia's ecosystem of open knowledge, but also to create connections with other institutions' data, and data contributed by volunteers.

WIKIMEDIA
SVERIGE

Museum collections on Wikidata

Several museums have made a decision to upload their collection data to Wikidata, or to release openly licensed datasets that volunteers can work with.

The Nationalmuseum in Stockholm has worked with Wikimedia Sverige to upload high quality images to Wikimedia Commons and the metadata about the artworks to Wikidata.

Items with Nationalmuseum Artwork ID

```
SELECT ?item ?itemLabel ?natmus_id ?image WHERE {
  ?item wdt:P2539 ?natmus_id.
  OPTIONAL { ?item wdt:P18 ?image. }
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "sv|,en".
  }
}
```

item	itemLabel	natmus_id	image
🔍 wd:Q18537473	Q18537473	25644	
🔍 wd:Q18573169	Procession at the Feast of Corpus Christi in Rome	18223	🖼 commons:Procession at the Feast of Corpus Christi in Rome (Egron Lundgren) - Nationalmuseum - 18223.tif
🔍 wd:Q18573186	Cupid	18246	🖼 commons:Cupid (Fredric Westin) - Nationalmuseum - 18246.tif
🔍 wd:Q18573562	Queen Marie Antoinette of France and two of her Children Walking in The Park of Trianon	18035	🖼 commons:Queen Marie Antoinette of France and two of her Children Walking in The Park of Trianon - Nationalmuseum - 18035.tif
🔍 wd:Q18573677	Christ and the Woman taken in Adultery NM 1128	18131	🖼 commons:Christ and the Woman taken in Adultery - Nationalmuseum - 18131.tif
🔍 wd:Q18574301	The Council of the Society of Artists	18606	🖼 commons:The Council of the Society of Artists (Richard Bergh) - Nationalmuseum - 18606.tif
🔍 wd:Q18577540	Portrait of Pope Clement XI	16162	🖼 commons:Clemens XI, 1649-1721, påve - Nationalmuseum - 16162.tif
🔍 wd:Q18578927	Q18578927	40239	🖼 commons:Självporträtt (Egron Lundgren) - Nationalmuseum - 40239.tif
🔍 wd:Q18578930	Selfportrait	19297	🖼 commons:Selfportrait (Martin van Meytens d.y.) - Nationalmuseum - 19297.tif

https://www.wikidata.org/wiki/Q18573562

Artworks that depict a person

```
SELECT ?item ?itemLabel ?image WHERE {
  ?item wdt:P180 wd:Q47365.
  OPTIONAL { ?item wdt:P18 ?image. }
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "sv,en".
  }
}
```

https://w.wiki/39aq

Crotos

www.zone47.com/crotos/

A search and display engine for artworks powered by Wikidata and Wikimedia Commons.

WIKIMEDIA
SVERIGE

The Sum of All Paintings

https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings

A WikiProject (dedicated community space) to collect and improve the data about the world's artworks and coordinate mass imports.

Background [edit]

On Wikipedia we have a lot of experience breaking down large amounts of work into smaller chunks that can be completed. Examples of this are the windmill project on the Dutch Wikipedia and the many lists of historic buildings that were created for Wiki Loves Monuments (Q1353202). For paintings we can use the same strategy: We break it down into smaller parts that can be handled and then we can utilize Wikidata to celebrate the sum of all paintings through linked open data!

Add a painting! See also this video by Shonagon

Scope [edit]

All or everything is too broad and has no limit. This limit is to be able to handle the processing of the data – let's not overwhelm ourselves (yet). The scope can always be expanded – in the future we will have the sum of all sculptures, drawings, you get the picture! The scope of this project is notable paintings. A painting is considered notable if one of these options is true:

- The painting is in a notable collection of a museum, library, archive (e.g. a painting in the Metropolitan Museum of Art)
- The painting is made by a notable artist (e.g. Van Gogh painting in a private collection)

We have lots of work to do, please join us!

Tips [edit]

- Paintings in scope meet notability guidelines in Wikipedia, as they are notable for being in a notable GLAM (Q1030034) collection, are painted by notable painters, and/or generally have secondary reliable source documentation written about them.
- GLAM is an acronym and means galleries, libraries, archives and museums, but in a broader sense includes also public buildings like castles, town halls, administrative buildings or religious buildings like churches, cloisters, temples etc. that can hold paintings or wallpaintings and other pieces of art.
- Use the guidelines of item structure when adding data to each painting entry. Think of it like a collections management database of information.

WIKIMEDIA
SVERIGE

Wikidata at the National Library of Wales

Has been working the Wikimedia Commons for a long time, uploading digitized material.

A lot of structured data that couldn't be used efficiently on Commons.

A strategic project – the Welsh Landscape prints.

Over 4,000 objects, 500 artists and engravers.

Being able to work with a whole collection provided an interesting opportunity to examine how the Wikimedia projects can fit into the institution's work.

WIKIMEDIA
S V E R I G E

Dwynwen

Search
Browse
Map

Sök

Resultat 1 - 20 av 27

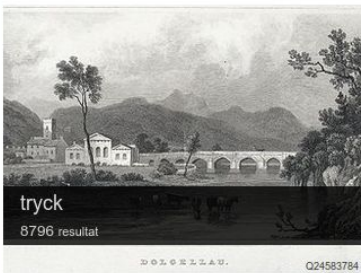Är en/ett | Skapare | Genre | Samling | Ämnesord | Motiv | Publicerad i | Minsta 3
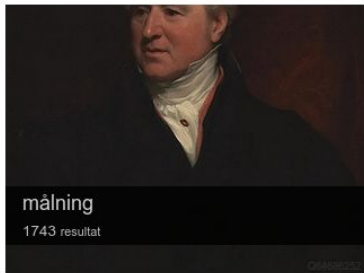
1 | 2 | Nästa | Slumpvis

tryck
8796 resultat
Q24583784

fotografi
6388 resultat

målning
1743 resultat

teckning
833 resultat
Q78093813

akvarell
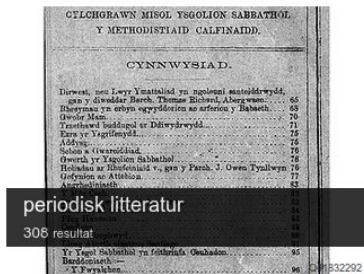606 resultat
Q77883331

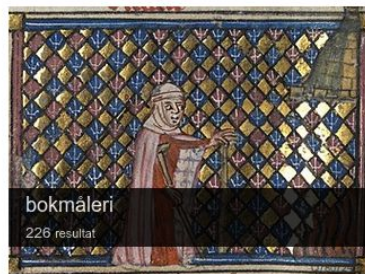publikation
435 resultat

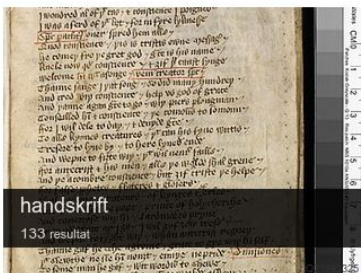periodisk litteratur
308 resultat
Q1832292

journal
247 resultat
Q144740

bokmåleri
226 resultat

karta
152 resultat
Q95562906

handskrift
133 resultat

tidning
62 resultat
Q20872266

veckotidning
57 resultat

http://zone47.com/crotos/dwynwen/
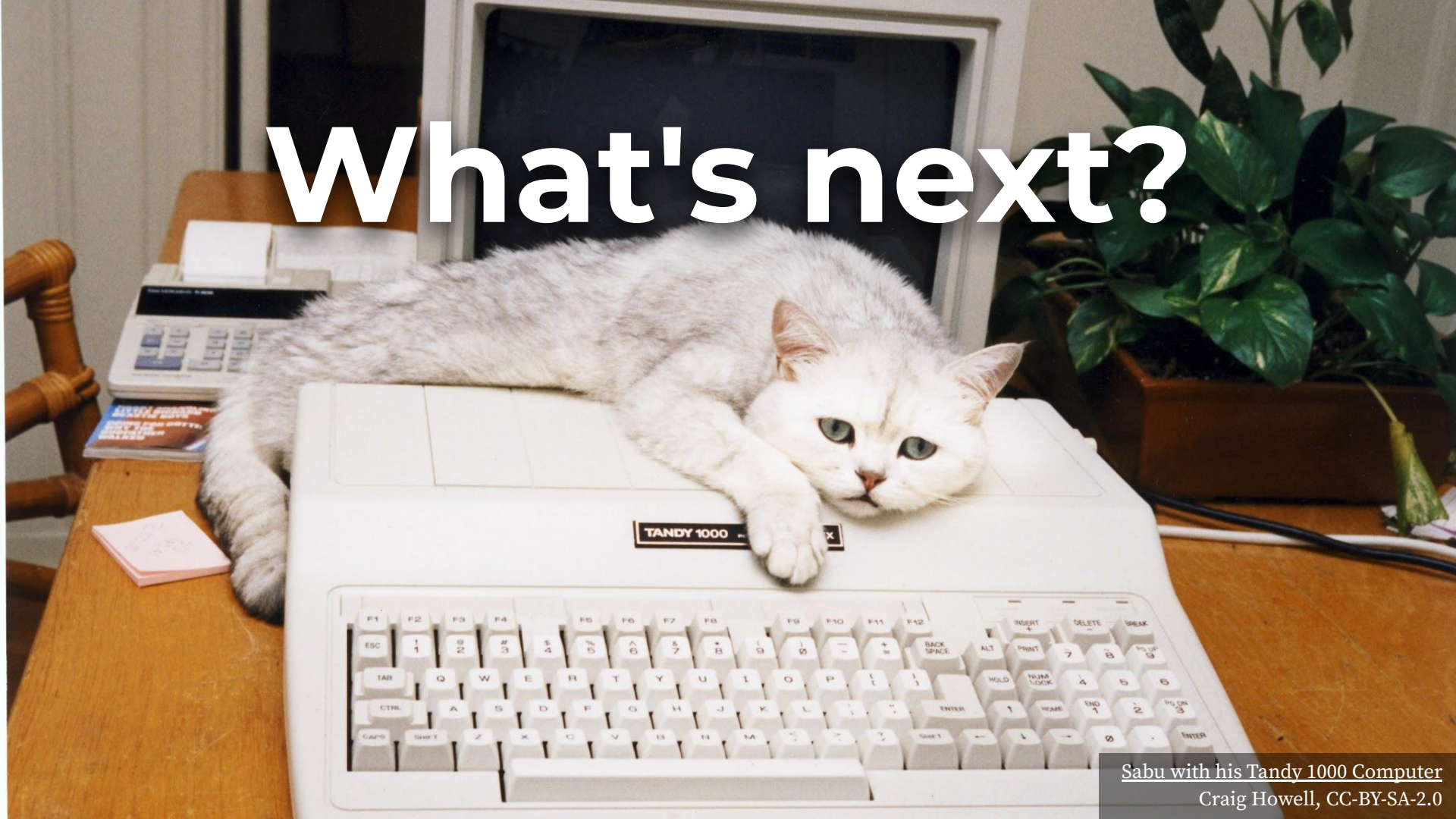
The Wikimedia projects as a channel

When an artwork is on Wikipedia, the readers are one click away from learning more about its context and about the collection it belongs to.

The Association of Swedish museums reports that in 2020, Wikipedia was an important channel for several institutions.

14 museums received in total 174 million views to "their" files on Wikipedia.

https://www.sverigesmuseer.se/wp/wp-content/uploads/2021/01/Rapporten-Museibes%C3%B6ken-2020.pdf

WIKIMEDIA
SVERIGE

What's next?

Controlled vocabularies and Wikidata

There are lots of terminologies, authorities and other types of controlled vocabularies in use across GLAM institutions, which is a barrier to discovery and analysis of cultural heritage data.
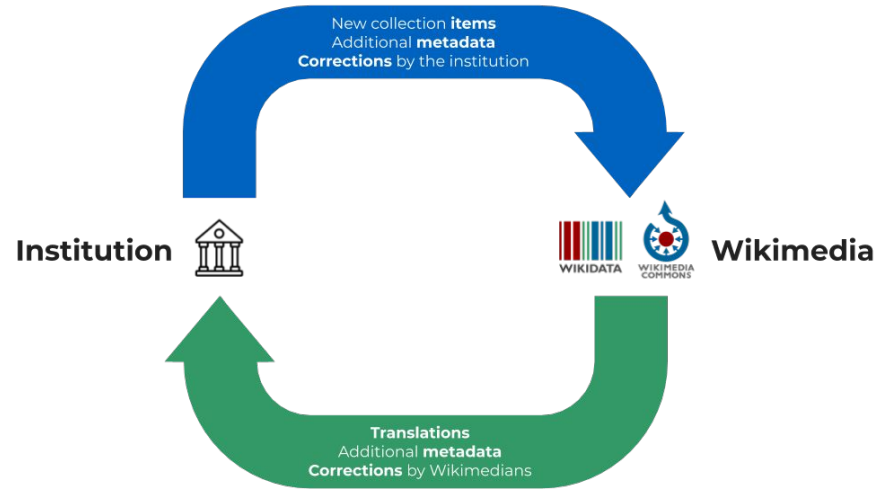
Wikidata has a lot of potential as a hub. By linking an item in an institution's vocabulary to its corresponding Wikidata item, you indirectly create links to corresponding entries in other vocabularies.

KulturNav (kulturnav.org) allows its users to define equivalent terms in other vocabularies, and it encourages editors to link to corresponding Wikidata items.

WIKIMEDIA
S V E R I G E

Data Roundtripping

A process to ingest the changes people have made on the Wikimedia projects back into the institutional database that data originated from.

There's a growing interest in recognizing these contributions.

Institution

New collection **items**
Additional **metadata**
Corrections by the institution

WIKIDATA | WIKIMEDIA COMMONS | Wikimedia

Translations
Additional **metadata**
Corrections by Wikimedians

WIKIMEDIA
S V E R I G E

Data Roundtripping

The Swedish National Heritage Board has investigated the potential for data roundtripping among some institutions, including Musikverket, Nationalmuseum and Nordiska museet.

[Wikimedia Commons Data Roundtripping - Final report (2019).](#)

Can we trust crowdsourced data?

Wikidata vs. local repositories

The larger Wikidata becomes, the more important is to reflect on its role in the cultural heritage knowledge ecosystem.

Can it contain everything? Can it replace local databases?

https://www.wikidata.org/wiki/Wikidata:Notability

Wikidata as a hub

Wikidata enables synergies between different data sources.

But it's still far from being complete, so when reusing the data, one should think about the biases and what conclusions we can make from the data.

90 million items is not that much!

Thank you!

wikimedia.se/engagera-dig

Alicia Fagerving
alicia.fagerving@wikimedia.se

WIKIMEDIA
SVERIGE

April 1, 2021
Linnæus University