

Case Report Sum of all Swiss GLAMs

Project Team: Maya Bangerter, Daniel Burkhard, Valérie Hashimoto, Lisa Oberli

Case report: Daniel Burkhard & Valérie Hashimoto, 6.1.2020, CC-BY 4.0

Table of content

1. Introduction
2. Current state of the Wikidata items on Swiss GLAMs and problematic modelling issues
3. Automated Wikipedia infobox templates extracting information from Wikidata
 - 3.1 Italian speaking Wikipedia
 - 3.2 French speaking Wikipedia
 - 3.3 German speaking Wikipedia
 - 3.4 Challenges
4. Integration of the ISIL Database
 - 4.1 Goals
 - 4.2 Challenges
 - 4.3 Outcome
5. How to pursue the project: Sensitization of Swiss GLAMs to Wikidata
6. Conclusion

1. Introduction

The project Sum of all Swiss GLAMs was initiated during the Wikidata Hackday Series of the Swiss National Library in collaboration with the Zentralbibliothek Zürich and the ETH Library Zurich. The Hackday Series took place from 27 June to 19 November 2019.

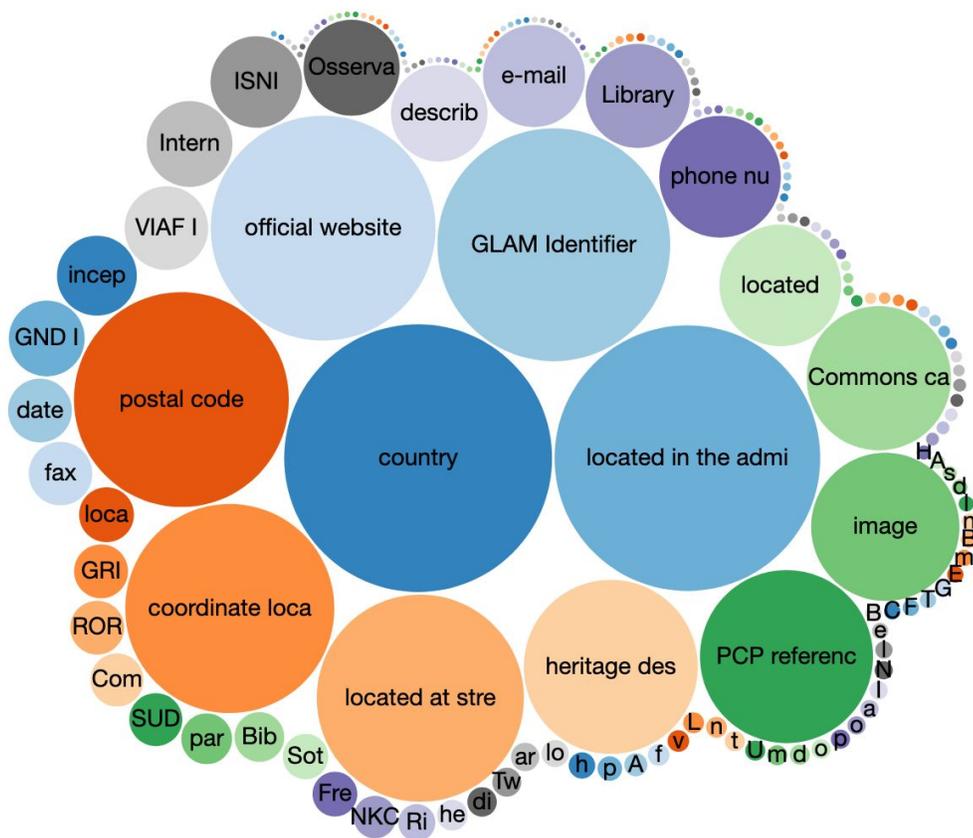
Sum of all Swiss GLAMs is a pilot project aiming at the improvement of Wikidata items related to Swiss heritage institutions (GLAMs standing for Galleries, Libraries, Archives, Museums). It is part of the more global [Wikiproject Heritage institutions](#), which intends to create a complete and high-quality database of worldwide heritage institutions through Wikidata.

The following case report summarizes the goals, procedures and main challenges of the project. Our initial goals were the following:

1. Analyse the current state of the Wikidata items on Swiss GLAMs
2. Implement automated Wikipedia infobox templates extracting information from Wikidata
3. Integration of the ISIL Database
4. Sensitize Swiss GLAMs to Wikidata

2. Current state of the Wikidata items on Swiss GLAMs and problematic modelling issues

In 2016, a project was dedicated to [Ingesting Swiss heritage institutions into Wikidata](#). As a result, most Swiss heritage institutions already possess a WD-item. The most used properties for these items, without distinguishing between libraries, archives and museums, are as follows ([query](#)):



Thus, the most used properties refer to the geographic localisation of the institutions, the GLAM Identifier (a new property introduced during the ingesting process of 2016 and attributing a unique identifier to each institution) and the official website, followed by the Commons category, the image and external identifiers. Important information such as the logo image, the director/manager or the number of visitors per year are respectively used for less than 30 items out of a total of about 2250. One of the long-term goals of the Sum of all Swiss GLAMs is to change this by involving the Swiss heritage institutions.

During our analysis of the data, we observed that the structure of the information was rather heterogeneous. For example the property "[located in the administrative territorial entity](#)" (P131) sometimes gives a canton and / or a city / and or a more specific area of a city. While this is not really problematic, we were also confronted with more profound issues related to data modelling. Indeed, we noticed that some of the items referred both to the institution as an organisation and as a building. This reflects different ways of handling data by the users of Wikidata, which results from the lack of a generally accepted consensus, from differences between the languages as well as inconsistencies in hierarchical class trees. To tackle these issues, a subpage specially dedicated to [data modelling issues](#) was created on the Wikiproject Heritage institutions. One of the challenges here is to define best practices which will be accepted and applied by the different language communities.

3. Automated Wikipedia infobox templates extracting information from Wikidata

On Wikipedia, more and more infobox models are linked to Wikidata, which permits to extract some information automatically from the corresponding WD-item. This is useful when creating a new infobox, because the user has to enter less information manually. Furthermore, the information which changes regularly, such as the number of visitors per year or the director of the institution, are updated automatically as soon as it is updated on Wikidata. Lastly, information which are no longer up to date are more likely to be detected on Wikipedia due to its greater visibility compared to Wikidata. This allows a better maintenance of the data on Wikidata, from which all language communities profit.

At the moment, "classical" infobox models for each libraries, archives and museums exist in the main Swiss national languages (French, Italian and German). One of our goals is to adapt these existing infobox models so that they are systematically linked to Wikidata. This has to be done for each language separately, as each Wikipedia linguistic community has their own models and ways of functioning. The community needs to be involved in the process to guarantee the long term acceptance and use of the new templates.

3.1 Italian speaking Wikipedia

In the Italian speaking Wikipedia, the interlinking of infoboxes to Wikidata is already widespread. Since 2016, the infobox models for heritage institutions have been progressively modified to extract information from Wikidata:

- [template for archives](#)
- [template for libraries](#)
- [templates for museums](#)

The functioning of the templates and the parameters are well explained on the template page. It is very clear which parameters fetch the information from Wikidata and which do not. The parameters which are left empty by the user and have a corresponding value in Wikidata are completed automatically. The user can always edit parameters manually (which automatically blocks the link to Wikidata for that specific parameter) or deactivate the extraction of information from Wikidata altogether.

3.2 French speaking Wikipedia

As of November 2019, only the [infobox template for museums](#) extracts information from Wikidata in the French speaking Wikipedia. Similarly to the template in Italian, the parameters can either be inserted manually or left empty so that they are fetched automatically from Wikidata. A [discussion](#) has been launched in 2018 to adapt the archives infobox following the museum model. The code is now being tested and should soon be ready to use. For the library infobox, the user who adapted the museum infobox has been asked if he could do the same for libraries.

3.3 German speaking Wikipedia

In the German speaking Wikipedia, the community is rather reluctant to fetch information from Wikidata. In this specific community, all changes in Wikipedia have to be “viewed and approved” by a member in order to avoid vandalism and the propagation of wrong or not neutral information. The automatic update of information is considered as a loss of control over the contents, as the editing of Wikidata is not submitted to any restrictions. As of November 2019, only very few information is automatically added to the infobox templates for heritage institutions (ISIL number for archives and museums, ISIL number, director and collections for libraries). A

[discussion](#) has been launched within the community to progressively link more parameters to Wikidata .

Museo d'arte della Svizzera italiana (MASILugano)



Ubicazione

Stato Svizzera
Località Lugano Arte e Cultura e Palazzo Reali
Indirizzo Piazza Bernardino Luini 6, 6901 Lugano

Caratteristiche

Tipo arte moderna e arte contemporanea
Apertura 12 settembre 2015
Direttore Tobia Bezzola
Visitatori 62 876 (2017)



Sito web [↗](#)

Modifica dati su Wikidata · Manuale

Museo d'arte della Svizzera italiana (MASILugano)



façade du siège "Lugano Arte e Cultura" du MASILugano

Informations générales

Type musée d'art moderne et contemporain
Ouverture 12 septembre 2015 [↗](#)
Dirigeant Tobia Bezzola (d) (depuis 2018) [↗](#)
Visiteurs par an 62 876 (2017) [↗](#)
Site web www.masilugano.ch [↗](#)

Localisation

Adresse Piazza Bernardino Luini 6
Lugano, canton du Tessin
 Suisse
Coordonnées 45° 59′ 58″ N, 8° 56′ 53″ E﻿ / ﻿45.99944° N, 8.94806° E﻿ / 45.99944; 8.94806 [↗](#)



modifier · modifier le code · modifier Wikidata

Archiv für Bauernhausforschung



Gebäude an der Hofstrasse 15 in Zug

Archivtyp Spezialarchiv
Koordinaten 47° 9′ 36,4″ N, 8° 30′ 57,1″ O﻿ / ﻿47.15983° N, 8.51558° O﻿ / 47.15983; 8.51558
Ort Zug
Besucheradresse Hofstrasse 15, 6300 Zug
Umfang 200'000 Fotonegative, 24'000 Diapositive, 10'000 Planzeichnungen
ISIL [CH-001401-1](https://nbn-resolving.org/urn:nbn:ch:swiss-ethz-isil-001401-1) [↗](#)
Träger Schweizerische Gesellschaft für Volkskunde
Website zg.ch/.../bauernhausforschung [↗](#)

Infobox models in Italian, French and German.

3.4 Challenges

As mentioned above, one of the challenges regarding the adoption of automated infoboxes is the fact that this change has to come from within the given linguistic community. This requires efforts of persuasion, discussions and time.

Furthermore, the automatic retrieval of information from Wikidata sometimes reveals some of the data modelling issues mentioned above. In the Italian template, for example, the location (ubicazione) is composed by the country ([P17](#)), then the "location" (località) and finally the

address ([P6375](#)). The property used for “località” in this case is [location](#) (P276), which in some Wikidata items indicates the building(s) where the institution is located (and not the city or town). In these specific cases, the term “località” is not adequate and should be replaced by “sede” (headquarters). Unfortunately, the parameter title cannot be modified in the template. Similarly, the modelling of the localisation in the French template for museums is not always easy to handle.

4. Integration of the ISIL Database

4.1. Goal

The Swiss National Library provides a central database for Swiss heritage institutions: The ISIL-database. It was the idea of the project to reconcile the data within Wikidata with the data provided by the ISIL-database and to enrich missing information in Wikidata.

4.2. Challenges

In the first step, the data of the ISIL-database, which was created in a ScopeArchives software, was exported to a csv-file and imported into OpenRefine. In OpenRefine we mapped the fields of the ISIL data with a minimum of needed properties from Wikidata. Other fields that showed great inconsistencies, such as the descriptions of the collection, were left out. We focused on the properties “Name” (Item Name), “instance of” (P31), “part of” (P361), “has part” (P527), “inception” (P571), “country” (P17), “location” (P276), “phone number” (P1329), “e-mail address” (P968), “official website” (P856) and “International Standard Identifier for Libraries” [ISIL-Number] (P791).

Unfortunately, it was not possible with the given data structure to verify the hierarchy and relation within the institutions regarding the property “is part of” (P361) and “has part” (P527). We hoped to find an inherent structure of the organisation in the ISIL-Code itself, but there wasn’t any. The identification code “ISIL” as it is used in Switzerland lacks an inherent structure. There is therefore no way to draw conclusions from the ISIL-number to the institutional hierarchies or relations.

Due to this problem, we decided to ingest only data where there are tracks of the institutional hierarchy in the data e.g. in the field “Name”. This was the case for data regarding the University Library of Bern. All ISIL-items regarding the University Library of Bern had the Prefix “UB Bern” attached to the Name of the institution. Thanks to this information it was possible to model the

relation and hierarchy of the institution within the data via OpenRefine. The ingest of this data (35 datasets) was done by Tuesday, 23.09.2019. For an example see [Q68330281](#).

4.3. Outcome

The ISIL-database contains interesting data regarding Swiss heritage institutions such as descriptions of the collections or information about the institutional history, but the given data structure does not allow a grouped integration in Wikidata, since the data is too heterogeneous. Cleaning the data would require not only manual up-dates on each item, but also intense research to verify the given data, since sometimes there is conflicting information contained in the data (e.g. official url leads to a deadlink or there are conflicting information about the date of inception (P571)) .

To work with the ISIL-database adaptations within the ISIL-database are necessary. For example, it would be desirable to adapt the transmission form so that the data could be better checked as soon as it is entered.

Furthermore the ISIL-database is updated by the institutions on a voluntary basis. As holder of the database, the Swiss National Library regularly asks the ISIL members to update their respective data, but does not have a controlling tool. As a result, some datasets are simply not up-to-date. Furthermore it must be addressed that the ISIL-database aims to complete information on all Swiss libraries, archives and museums, but there are also associations of each branch that collect their data as well. Thus there exists a database of all archives within the society of swiss archives (VSA), the same is true for the Swiss Museums Association (VMS), resulting in duplication of work when it comes to maintaining the data relating to Swiss heritage institutions. The use of Wikidata as a central database might be a solution to avoid this duplication of work. The use-case of the data regarding the University Library Bern could help as a model.

5. How to pursue the project: Sensitization of Swiss GLAMs to Wikidata

The last goal of the Sum of all Swiss GLAMs could not yet be achieved. It will be pursued during 2020 with a campaign targeted at Swiss heritage institutions. In order to reach out to the institutions, we hope to win the professional “umbrella” associations VSA, Bibliosuisse and VMS as intermediaries. Thus, the campaign will be run in parallel for the different institution types and in the various national languages.

In a first step, we will have to analyse the data maintained by the above mentioned “umbrella” associations. Together we will have to address the challenges and possible solutions with regard to data quality and completeness, evaluate ways of synchronizing and harmonizing the modelling of the various databases. As we see several advantages in using Wikidata as a main database, we shall discuss in which ways the various databases can complement each other and how the individual institutions could contribute directly.

The main scope of the campaign will be to inform the institutions about Wikidata and evaluate how they respond to our outreach. The use of automated infoboxes can be used as an argument to illustrate the utility of keeping the data up-to-date on a regular basis (at least in Italian and French due to the already functioning automated models). To involve Swiss heritage institutions, we will prepare basic tutorials explaining how to edit two properties which are most subject to change, namely the number of visitors per year and the manager/director. In order to observe the impact of the campaign, we will set up Wikidata queries to visualise the completeness of these two parameters (with distinction of institution type and linguistic area).

6. Conclusion

The project Sum of all Swiss GLAMs allowed us to get an overview of the current situation of Wikidata items dedicated to Swiss heritage institutions and at the same time, to analyse the adequacy and problematic aspects of the existing ISIL-database as inventory of Swiss GLAMs. The ingestion of all Swiss GLAMs carried through in 2016 provides an optimal starting point to work towards a complete and dynamic inventory of Swiss heritage institutions through Wikidata. As with the already established inventories such as the ISIL-database, the main challenge is to engage the institutions to keep their information up-to-date.

Compared to the established inventories, Wikidata has the main advantage that its data is multilingual and available to the whole world. It can be used for infoboxes on Wikipedia, which gives the information greater visibility and thus encourages a better maintenance of the data. As linked open data, the information can be retrieved through queries - one can, for example, get a list of all institutions which are managed by a woman (or by a man). However, such queries only make sense if this data is available for the majority of the institutions. Another advantage of Wikidata is the fact that the institutions can edit their information directly. Of course, this might result in a certain heterogeneity in the data modelling, but these issues can be tackled with time and by providing guidelines to the institutions.

In order to achieve a complete and high-quality inventory of Swiss heritage institutions, the involvement of the umbrella organisations (VSA, VMS, bibliosuisse) and the institutions

themselves is inevitable. This remains one of the main challenges of the project, which will be further pursued in 2020 with an outreach to the GLAMs.