

An International Knowledge Base for all Heritage Institutions (Part 1*)

Autoren : Giovanna Fontenelle, Beat Estermann

Datum : 4. Juli 2019



Heritage institutions are places in which works of art, historical records, and other objects of cultural or scientific interest are sheltered and made accessible to the public. The equivalent of that in the digital world, is already taking shape, through digitization and sharing of digital-born or digitized objects on online platforms. In this article we shed light on how the issue of structured data about heritage institutions is being tackled by Wikipedia, and its sister Wikidata, through their “Sum of All GLAM” project.^[1]

Access to these objects, and information about them, is provided and mediated both through platforms maintained by the heritage sector itself and through more general-purpose platforms, which often serve as a first point of entry for the wider public. These platforms include Google, Facebook, YouTube, and Wikipedia, which also happen to be among the most visited websites on the Web. In this emerging data and platform ecosystem, Wikipedia and related Wikimedia projects play a special role as they are community-driven, non-profit endeavours. Moreover, these projects are working hard to make data and information available in a free, connected and

structured manner, for anybody to re-use.

There are various layers of information about heritage institutions, ranging from descriptions of institutions themselves and descriptions of their collections, to descriptions of individual items. There may be digital representations of these items, and in some cases even searchable content within the items. Figure 1 illustrates how the top four layers of data and information are currently addressed in Wikipedia, with Wikidata and Wikimedia Commons increasingly focussing on providing structured and linked data alongside the unstructured or semi-structured encyclopaedic information contained in Wikipedia articles.

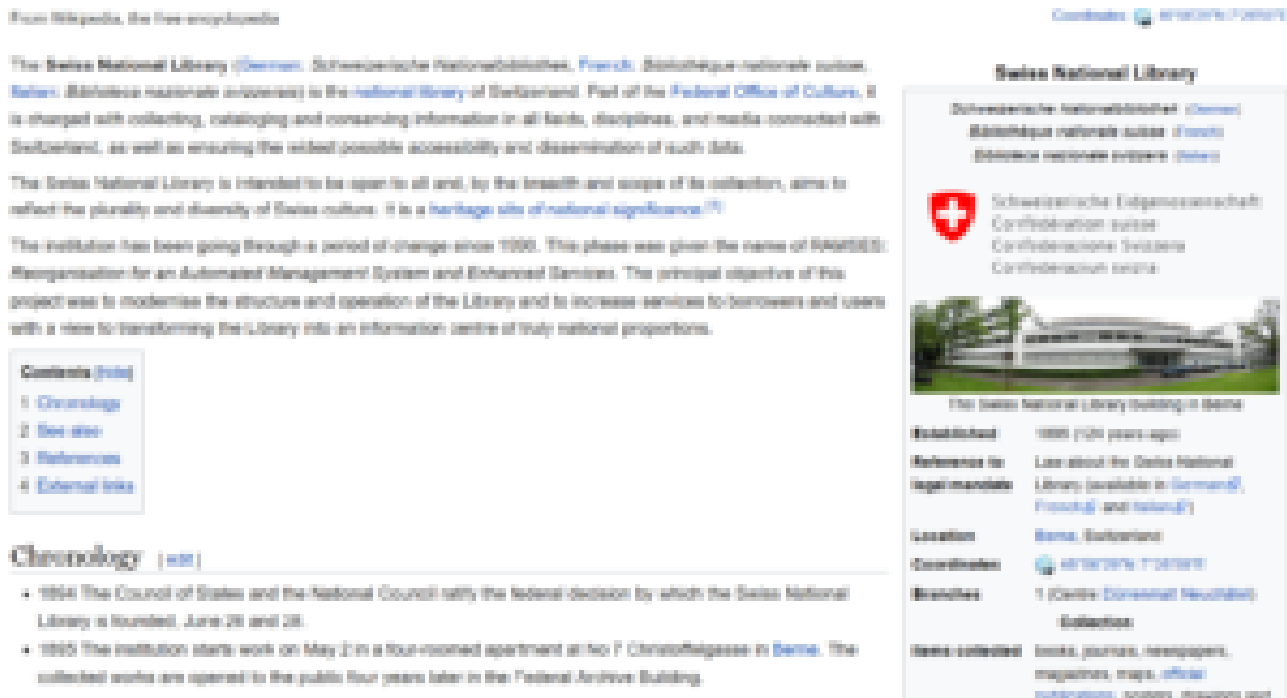


Figure 1: Heritage data and content in the context of Wikipedia and its sister projects

Structured data about institutions and collections, as well as some item-level data are maintained on Wikidata, which serves as Wikipedia's repository for structured data. Wikimedia Commons serves as Wikipedia's repository for media files and is currently being prepared for the linked data era through the "[Structured Data on Wikimedia Commons](#)" project. This project accompanies the transition of Wikimedia Commons to linked open data, foresees the provision of item level metadata as linked data and monitors the implementation of the IIIF standard, to allow easier cross-platform manipulation and media file sharing. While similar efforts are on their way at all the different levels of information, we will focus the the remainder of the article on a project that is dedicated to improving data quality and completeness of the top layer, i.e. the data about the heritage institutions themselves. This project lays the foundations for an [International Knowledge Base for Heritage Institutions](#). The project is currently managed by Wiki Movement Brazil in cooperation with OpenGLAM Switzerland and will be expanded to further countries in the near future; it is also being coordinated with "[FindingGLAMs](#)", a project run by Wikimedia Sweden, UNESCO, and the Wikimedia Foundation, which pursues similar goals, but addresses different layers of information, including aspects related to structured data on

Wikimedia Commons.

To succeed, the International Knowledge Base for Heritage Institutions needs to address all stages of the linked data value chain, from data provision to data use (figure 2):

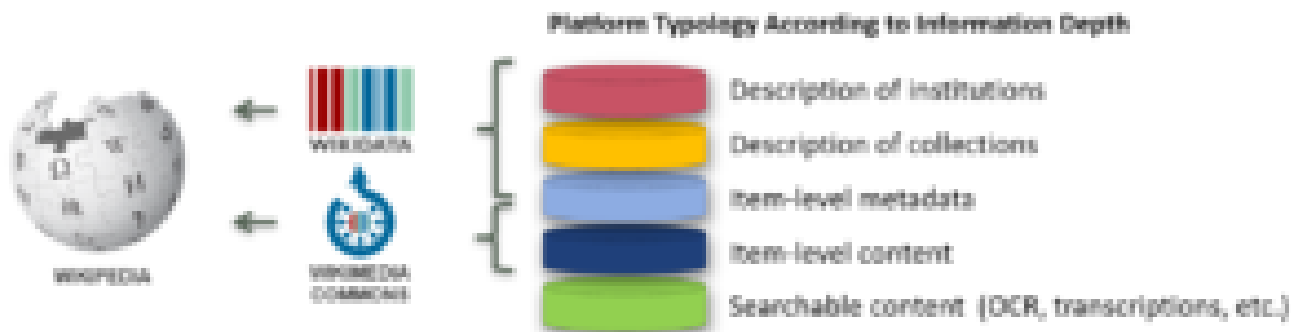


Figure 2: Core processes of linked data publication (source: [eCH-0205 – Linked Open Data](#))

Parts of the data have already been ingested into Wikidata and the relevant elements of the ontology have already been implemented so currently most of the effort is going into data maintenance. The goal is to provide the data in a coherent way that makes it fit for its use in Wikidata infoboxes (see figure 3 for an example). However, before it is ready to go there are various issues to be addressed, such as data quality, correct data modelling, and data completeness.

The screenshot shows a Wikipedia article with an infobox containing structured data for various countries. The data is organized into columns: Name, Capital, Coordinates, Population, Area, Official language, National anthem, and Collection of statistics. The countries listed include Italy, United States of America, Germany, Brazil, France, Japan, United Kingdom, Sweden, Poland, Switzerland, Spain, Russia, Netherlands, Czech Republic, Belgium, Republic of China, Canada, and Mexico.

Name	Capital	Coordinates	Population	Area	Official language	National anthem	Collection of statistics
Italy	Rome	41°56′N 12°29′E	60,725,201	301,330 km²	Italian	Il Canto degli Italiani	en:it
United States of America	Washington, D.C.	38°53′N 97°07′W	325,719,171	9,826,675 km²	English	The Star-Spangled Banner	en:us
Germany	Berlin	52°31′N 13°24′E	82,705,534	357,021 km²	German	Deutschlandlied	en:de
Brazil	Brasília	15°47′S 47°51′W	215,625,361	8,511,965 km²	Portuguese	Imno Nacional Brasileiro	en:br
France	Paris	48°51′N 2°21′E	67,395,515	643,801 km²	French	Marseillaise	en:fr
Japan	Tokyo	35°41′N 139°47′E	125,810,000	377,915 km²	Japanese	Kimigayo	en:jp
United Kingdom	London	51°30′N 0°07′W	67,899,000	244,820 km²	English	God Save the Queen	en:gb
Sweden	Stockholm	59°18′N 18°03′E	10,335,000	450,295 km²	Swedish	Då det började dån	en:se
Poland	Warsaw	52°10′N 21°03′E	38,080,000	312,685 km²	Polish	Mazurek Dąbrowski	en:pl
Switzerland	Bern	46°56′N 7°25′E	8,658,000	41,284 km²	German, French, Italian, Romansh	Gott segne unser Vaterland	en:ch
Spain	Madrid	40°41′N 3°41′W	45,929,000	505,992 km²	Spanish	Marcha Real	en:es
Russia	Moscow	55°45′N 37°37′E	145,934,000	17,095,200 km²	Russian	State Anthem of the Russian Federation	en:ru
Netherlands	Amsterdam	52°18′N 4°46′E	17,170,000	41,826 km²	Dutch	Wilhelmus	en:nl
Czech Republic	Prague	50°05′N 14°31′E	10,708,000	78,861 km²	Czech	Kde seš, ty můj kochánku	en:cz
Belgium	Brussels	50°51′N 4°23′E	11,589,000	30,528 km²	Dutch, French, German	Die Vlaamse Leeuw	en:be
Republic of China	Taipei	25°03′N 121°30′E	23,500,000	361,928 km²	Mandarin	义勇军进行曲	en:tw
Canada	Ottawa	45°28′N 75°40′W	38,232,000	9,984,670 km²	English, French	O Canada	en:ca
Mexico	Mexico City	19°28′N 99°08′W	128,129,000	1,964,375 km²	Spanish	El Himno Nacional Mexicano	en:mx

Figure 3: Wikipedia article with infobox containing structured data

The goal of the “Sum of All GLAM” project is to complete entries for all the heritage institutions of a given country, with all the data that is required for the infoboxes. To monitor progress in achieving this goal, we are currently putting in place various instruments that can be used by community members to focus their efforts in improving existing data entries (see figure 4 for an example). While the issues related to data modelling will be addressed by members of the Wikidata community at an international level, the project team is planning to involve members of the heritage community in the various countries to help improve completeness of the data and to make sure that all the data entries are properly sourced. While existing Wikidata community members are expected to work on this hand-in-hand with members of the heritage community, the project will heavily rely on the heritage sector to help keep the information about their institutions up-to-date in the longer run. In fact, in the case of medium-sized and larger institutions, regularly updating their existing Wikidata entry should eventually become part of the tasks carried out by an institution’s communication department. For smaller institutions, other solutions need to be found – possibly via the intermediation of umbrella organizations or specialized institutions which take care of coordination at a national level.

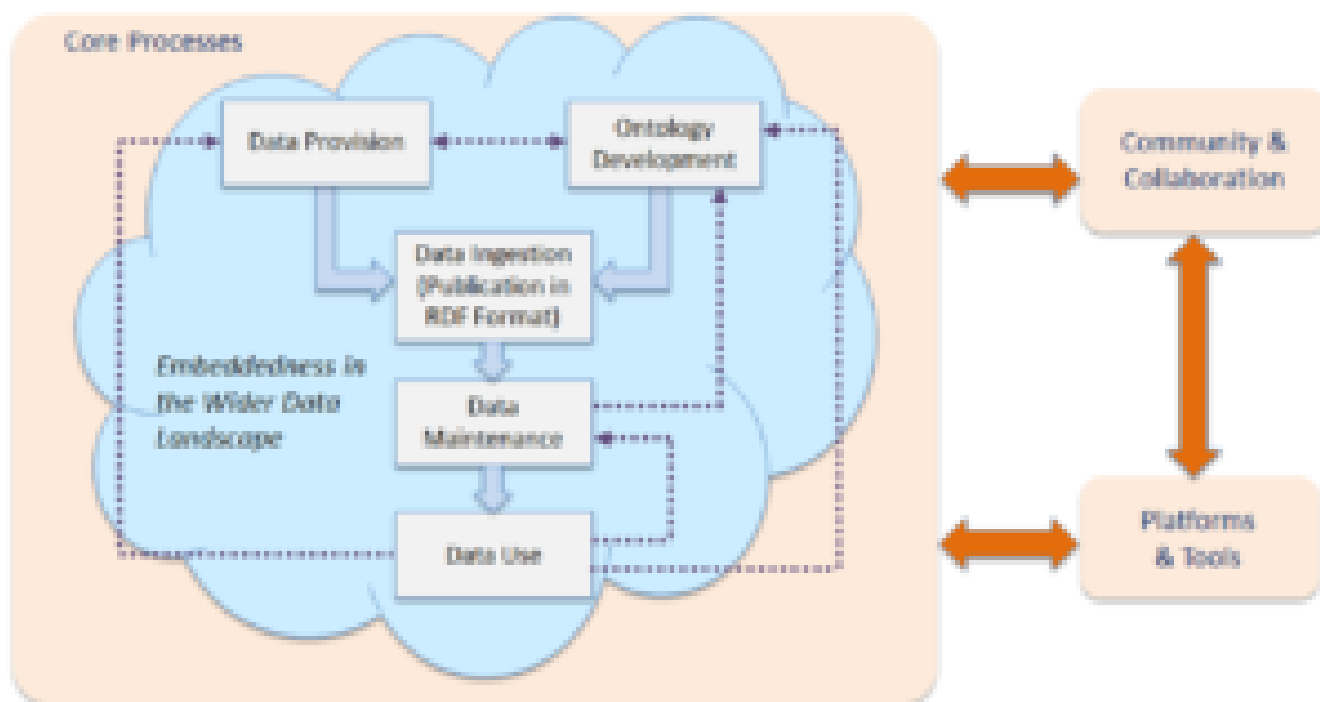


Figure 4: Table indicating the completeness of data about museums for different countries

To involve the members of the heritage sector in various countries, internationalization is being pursued early in the project: To do so, a well-documented model project is currently implemented in Brazil, which can in turn be implemented in other countries. To make sure that the project documentation is fit for its purpose, international partners ready to implement parts of the project in their country are currently being recruited. And to facilitate the tailoring of the project to local needs, the model project will be broken down into various modules that can be implemented separately or in combination with other modules, as the local partners see fit. In

the second part of this article we will describe some of these modules.

Part 2 of this article is published [here](#).

References

^[1]The working title, GLAM stands for “Galleries, Libraries, Archives, Museums”; the acronym is commonly used to refer to heritage institutions.