Building a Knowledge Graph for the History of Vienna with Semantic MediaWiki[⋆,⋆⋆]

Bernhard Krabina[a,*,1]

[a] *Vienna University of Economics and Business, Austria*

ARTICLE INFO

Keywords:
Semantic MediaWiki
Knowledge Graphs
Vienna History Wiki
Semantic Wikis
Linked Data
Open Government
Cultural Heritage
Digital Curation

ABSTRACT

While research on semantic wikis is declining, Semantic MediaWiki (SMW) can still play an important role in the emerging field of knowledge graph curation.

The Vienna History Wiki, a large knowledge base curated by the city government in collaboration with other institutions and the general public, provides an ideal use case for demonstrating strengths and weaknesses of SMW as well as discussing the challenges of co-curation in a cultural heritage setting. This paper describes processes like collaborative editing, interlinking unique identifiers on the web, sharing data with Wikidata, making use of Schema.org, and other ontologies. It presents insights from a user survey, access statistics, and a knowledge graph analysis.

This work contributes to the scarce research in wiki usage outside of the Wikipedia ecosystem as well as to the field of community-based knowledge graph curation. The availability of a now significantly improved RDF representation indicates future directions for research and practice.

1. Introduction

Knowledge graphs are an emerging form of knowledge representation [14]. While much of the emphasis lies on knowledge harvesting methods that have enabled the automatic construction of knowledge bases from web resources [32], the process of collaborative, manual knowledge base curation needs specialized tools. Knowledge graph construction in the Semantic Web domain is of high complexity with a very fragmented tool portfolio [23].

A shift in cultural heritage institutions (aka *GLAM*: galleries, libraries, archives, and museums) towards participatory approaches has occurred, where public audiences have been involved in media-supported co-curation processes including collaboration platforms on the web [3, 23].

The "Wien Geschichte Wiki" (Vienna History Wiki at www.geschichtewiki.wien.gv.at) can be described as a domain-specific knowledge graph powered by the open-source collaboration tool Semantic MediaWiki (SMW). SMW is potentially the only semantic wiki still in active development. While we see that general scholarly interest in semantic wikis seems to be declining and that SMW is often overlooked (see Section 2.3), SMW has the potential of serving as an interface for manual knowledge graph curation and creation from semi-structured sources, which is much needed in cultural heritage institutions [21].

The Vienna History Wiki is unique in many ways, thus making it an interesting use case for research on a community-based knowledge graph:

- it is operated by cultural heritage institutions of the city government,
- it is a scholarly wiki with an editorial team from several municipal departments and institutions outside of the city administration,
- it is open to the public, meaning that all interested parties can add or edit content,
- it has a regional focus and is the world's largest city wiki at this time,[1]
- it has a focus on historical knowledge and is currently the second largest history wiki and the largest history wiki powered by SMW,[2]
- it ranks among the ten largest SMW installations.[3]

This paper's aim is therefore twofold: (1) to demonstrate the capabilities (and shortcomings) of Semantic MediaWiki in building up and collaboratively curating a knowledge graph based on a concrete use case from the digital humanities, and (2) to give insights into the operation and maintenance of a large knowledge base curated by a city government in collaboration with citizens. Therefore we undertake the following exploratory work:

- we describe steps towards building a knowledge graph with SMW (Section 3),
- we deliver detailed insights based on access statistics and a user survey conducted in 2019 updating first empirical data from 2015 [17] (Sections 4.1 and 4.2), and
- we conduct a knowledge graph analysis as well as a description of the resulting RDF representation (Sections 4.3)

*Corresponding author
bernhard.krabina@s.wu.ac.at (B. Krabina)
https://www.linkedin.com/in/krabina/ (B. Krabina)
ORCID(s): 0000-0002-6871-3037 (B. Krabina)

[1] https://wikiapiary.com/wiki/Special:RunQuery/Website_by_Tag?Website_by_Tag%5BTag%5D=city%20wiki&wpRunQuery=true

[2] https://wikiapiary.com/wiki/Special:RunQuery/Website_by_Tag?Website_by_Tag%5BTag%5D=history&wpRunQuery=true

[3] https://wikiapiary.com/wiki/Semantic_statistics

The remainder of this paper is structured as follows: After providing background on the relevant topics (Section 2), we describe the required steps carried out during the recent development of the Vienna History Wiki towards a public knowledge graph (Section 3) and conduct a threefold analysis (Section 4). After the discussion (Section 5) we outline potential future directions for research and practice (Section 6).

2. Background and Related Work

This section presents important aspects of digital curation in cultural heritage, wikis, data, collaboration, Semantic MediaWiki, and linked open data, and it provides background information about the Vienna History Wiki, highlighting aspects that make it a special use case where comparable research hardly exists.

2.1. Digital Curation and Cultural Heritage

The digital environment has redefined the humanities, archives, and the practice of curation [25]. Archivists and librarians manage, maintain, preserve, and ensure access to information by digital curation, which is a relatively new concept that attempts to bridge boundaries among archivists, librarians, records managers, and other information professionals [10].

Historical questions can often only be answered by combining information from different sources, from different researchers and organizations, who increasingly use the internet as a medium for publication and exchange [20]. Therefore, collaborative data collection initiatives are becoming increasingly pivotal to cultural institutions and scholars [7].

While the amount of digital cultural heritage data produced is growing rapidly, many repositories publish as raw dumps in different file formats lacking structure and semantics, limiting the capabilities of users to contextualize information from distributed repositories [21].

Digital humanists are keenly interested in building scholarly editions, data sets (and data visualizations), digital thematic research collections, websites, and digital archives [25], all of which can be supported by semantic wiki technology.

2.2. Wikis, Data and Collaboration

Wikis are not only used in open communities, but also in corporations. The key to successful wiki usage is to create an environment in which people feel a strong sense of commitment to the shared knowledge repository, to design a system that requires little effort to reorganize, and to add to knowledge created by others [2]. Topic-oriented and regional wikis can also be used for open collaboration with the government or within government institutions [19].

MediaWiki – the open-source software that powers Wikipedia, Wikidata, and many other projects of the Wikimedia Foundation – is used by many individuals and organizations for various purposes outside of the Wikimedia ecosystem, which has led to the creation of the *MediaWiki Stakeholders Group*.[4]

[4] https://www.mediawiki.org/wiki/MediaWiki_Stakeholders%27_Group

Several options are available[5] for managing structured data within MediaWiki [28], the most notable one being Wikibase[6] – the MediaWiki extension that powers Wikidata. While it would be possible to build up a knowledge base with facts about the history of Vienna using Wikibase, it would not be feasible to create an online encyclopedia based on these facts in the same environment. Just like in the Wikidata/Wikipedia ecosystem, it would be necessary to run a MediaWiki installation with Wikibase as the data backbone and a MediaWiki installation accessible to the general public where articles are edited that use the structured data edited in Wikibase. For a query interface, an additional component would be needed: a triple store with a SPARQL endpoint.[7]

2.3. Semantic MediaWiki

Back in 2013, when the decision was taken to implement the Vienna History Wiki [17], Wikibase was in a much too early stage of development. Apart from that, one of the benefits of Semantic MediaWiki is the internal query language[8] that makes it possible to query for data within the wiki in every page or template. Furthermore, SMW is designed to manage text corpora alongside structured data.

Since the publication of the first scientific papers in 2006 and 2007 suggesting SMW [31, 18], a peak in research papers regarding semantic wikis and SMW occurred from 2012 to 2013, with another peak for SMW in 2018 (see Figure 1).[9]

While a study from 2012 [5] lists 25 semantic wikis, an article about authoring with semantic wikis two years later [20] focuses on Semantic MediaWiki and OntoWiki. Currently, Semantic MediaWiki is still in active development,[10] while the latest commit to OntoWiki is from 2017.[11] According to WikiApiary[12], 6% (1,642 of 26,222) of the wikis powered by MediaWiki and listed in WikiApiary use SMW to store more than 1,033,440,009 values for 1,182,404 properties.

One could argue that SMW has become the de facto standard for semantic wikis, but the problem seems to be that oftentimes it tends to be overlooked by research and practice. A recent study of semantic web-based repositories for cultural heritage [21] comparing open-source solutions like WissKi[13], Arches[14], ResearchSpace[15], Omeka S[16] did not mention SMW at all, neither do research reports of projects

[5] https://www.mediawiki.org/wiki/Manual:Managing_data_in_MediaWiki

[6] https://wikiba.se

[7] https://www.mediawiki.org/wiki/Wikibase/Suite

[8] https://www.semantic-mediawiki.org/wiki/Help:Inline_queries

[9] The numbers for the keywords have been compiled with https://app.dimensions.ai/discover/publication in mid July 2022. The numbers for 2022 have been multiplied by two, to approximate the values.

[10] https://github.com/SemanticMediaWiki/SemanticMediaWiki

[11] https://github.com/AKSW/OntoWiki

[12] https://wikiapiary.com

[13] https://wiss-ki.eu

[14] https://www.archesproject.org

[15] https://www.researchspace.com

[16] https://omeka.org/

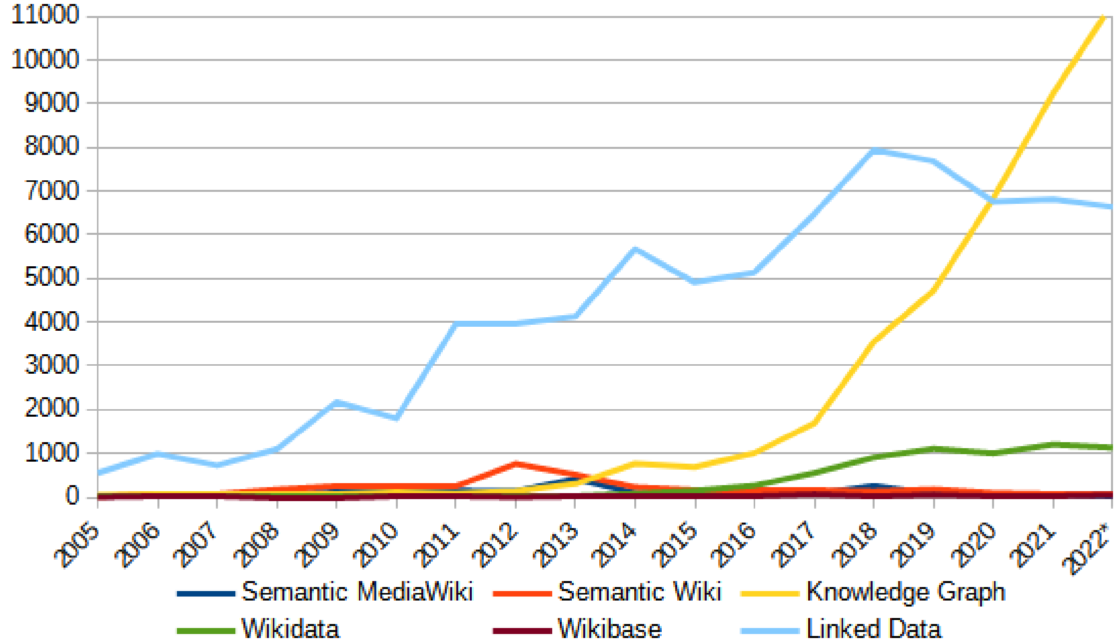

Figure 1: Google Scholar Publications

using Wikibase[17] in GLAM institutions [11, 9, 27]. Unfortunately, no current research on related work[18] based on SMW exists. Notable exceptions are [7] introducing *CLEF* as a novel linked data platform for cultural heritage, [29] providing an ontology-based approach to creating SMW instances, and [15] implementing a SMW based collaboration platform for research integrity and ethics.

2.4. Linked Open Data

The well-known linked data principles described by Tim Berners-Lee are:[19]

- Use URIs as names for things.
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provide useful information, using the relevant standards (RDF, SPARQL).
- Include links to other URIs, so that they can discover more things.

Linked open data serves as a bridge between humanities disciplines and underutilized digital collections and is an essential function of repositories, since one of the aims of data curation is to support research across multiple data sets, collections, and text corpora [24].

While SMW was developed to support linked open data, we will show in this paper how it must be set up to fulfill all the above-mentioned principles.

[17] https://wikiba.se/

[18] For example the HEALD project, see https://heald.nga.gov.

[19] https://www.w3.org/DesignIssues/LinkedData.html

2.5. The Vienna History Wiki

The initial idea of the Vienna History Wiki was to implement an online version of the six-volume encyclopedia *Historisches Lexikon Wien* edited by *Felix Czeike*, published 1992–2004. It was put together by the Municipal and Provincial Archives of Vienna and the Vienna City Library and was opened to the public on September 11, 2014. It is a geo-referenced, historical knowledge platform of the city of Vienna aiming to combine knowledge from the city administration with that of external experts.[20]

In contrast to other wikis, the Vienna History Wiki does not rely solely on a voluntary community, but is governed by an editorial team formed by several administrative departments of the Vienna city administration as well as several external project partners, such as the Wien Museum[21], the Department of Urban Archaeology of Vienna[22], the Jewish Museum Vienna[23], the Association for the History of Vienna[24], the Austrian Mediathek[25], the Austrian Institute of Historical Research[26], and the Centre for Environmental History[27]. Edits by users are not displayed immediately, they are subject to review by the editorial team before they are accepted. Not only do the partner institutions provide the editorial team to revise user-generated content, but they also provide staff to do regular edits, upload images, and write new articles [17].

[20] https://www.geschichtewiki.wien.gv.at/Vienna_History_Wiki

[21] https://www.wienmuseum.at/en/

[22] https://stadtarchaeologie.at/

[23] https://www.jmw.at/en

[24] https://www.geschichte-wien.at/

[25] https://www.mediathek.at/ueber-uns/information-in-english/

[26] https://geschichtsforschung.univie.ac.at/en/

[27] https://boku.ac.at/en/zentrum-fuer-umweltgeschichte

3. Building a Knowledge Graph

The complexity of Semantic Web technologies makes it difficult – especially for nontechnical experts – to use these technologies [4, 23]. Because of the ecosystem of extension that has evolved around SMW, it is possible to provide an environment that hides a lot of that complexity: The users of the Vienna History Wiki are not even aware of the underlying semantic technologies. Because the original content of the *Czeike* encyclopedia was not suited to the purposes of the Vienna History Wiki, users did not have to semantically annotate the original texts. Instead they had to initially copy and paste the original texts, improve them, and fill out forms in order to provide structured data alongside the text. The semantic *info boxes* (a term frequently used in Wikipedia) feature a link to the RDF representation.

The knowledge base built around historical knowledge for the City of Vienna covers several main categories:[28] 16,259 *People*, 11,018 *Topographic Objects* (roads, parks, rivers, cemeteries,...), 5,497 *Structures* (buildings, bridges, churches,...), 4,867 *Organizations*, 944 *Events*, 2,821 *Memorials* (statues, plaques, stones, signs), 266 *Maps*, 1,448 *Terms*, 207 *Border Stones*, and 2,564 *Other* entries. The wiki currently contains 45,891 encyclopedic entries and 12,960 images.

The content has been expanded far beyond the original scope of the six-volume *Czeike* encyclopedia. The category *Czeike* still shows the 26,235 entries of the encyclopedia (without redirects), which make up roughly 57% of the 45,891 entries. Not only have new entries continually been added since the release of the last print volume in 2004, but other resources from the Vienna archive and library have also been added as well e.g. content from other books, most notably 1,180 entries from a book about buildings in Vienna[29] and 945 entries from a book about Viennese buildings named "Hof".[30]

In our attempt to build a knowledge graph from this historic knowledge base, we refer to a definition commonly used for knowledge graphs [22]. Based on this definition, a knowledge graph

- mainly describes real world entities and their interrelations, organized in a graph,
- defines possible classes and relations of entities in a schema,
- allows for potentially interrelating arbitrary entities,
- covers various topical domains.

In order to satisfy the first condition, we can argue that even if SMW still stores its values internally in the MySQL database, the fact that RDF representations are available qualifies it as a graph-based data structure, as SMW essentially transforms pages and links to concepts and relations with attributes [31]. Generally speaking, SMW can be connected to RDF databases,[31] but the infrastructure of the city administrations's IT department cannot currently provide this.

What still needs to be done is to define classes and relations of entities in a schema. In order to define relations, SMW features the new *Property:* namespace where attributes are defined in wikitext. For example, introducing the property *Date from* can simply be done by annotating wikitext as follows: `[[Date from::1740]]`. By default, properties in SMW are of the *page* type, which in this example could make sense if a page *1740* in the wiki is wanted. Otherwise, a property can be explicitly declared by adding the wiki page *Property:Date from* with the wikitext annotation `[[Has type::Date]]`, which then transforms all already annotated values from page to date.[32] For classes, SMW uses the already existing category mechanism of MediaWiki.[33]

The two remaining aspects of the knowledge graph definition – interrelating arbitrary entities with each other and covering various topical domains – are also met.

3.1. Collaborative Editing

The editorial process of writing wiki articles collaboratively is supported by many features of MediaWiki that are well known from Wikipedia (e.g. version history of each page, user rights, namespace restrictions, special pages for editorial purposes). In the case of the Vienna History Wiki, the editorial process is further enhanced by the extension *Approved Revs*,[34] which provides mechanisms for *approving* an edit by an editorial team and displaying the approved version of a page instead of the most recent version of a page containing unapproved edits.

Also, SMW supports the process of *semantic gardening*,[35] an activity that allows the monitoring of the health of value statements and property declarations as part of data curation activities.

3.2. Form-Based Data Entry

Entering structured data in SMW is often supported by a forms extension, the most notable being *Page Forms*.[36]

In conjunction with the extension *External Data*,[37] a form can be provided that queries Wikidata for the page name to be created or edited and suggests Wikidata and GND identifiers (see Figure 2). The GND (Integrated Authority File) managed by the German National Library is an important identifier, especially in the GLAM domain.[38]

If the page name to be created or edited is found in Wikidata, it will provide the Wikidata description as well

[28] These values change daily, see https://www.geschichtewiki.wien.gv.at/Statistik

[29] https://www.geschichtewiki.wien.gv.at/Paul_Harrer:_Wien,_seine_H%C3%A4user

[30] https://www.geschichtewiki.wien.gv.at/Wolfgang_Wirsig:_Wiener_Hofnamen

[31] https://www.semantic-mediawiki.org/wiki/Help:Using_SPARQL_and_RDF_stores

[32] https://www.semantic-mediawiki.org/wiki/Property:Has_type

[33] https://www.semantic-mediawiki.org/wiki/Help:Classification

[34] https://www.mediawiki.org/wiki/Extension:Approved_Revs

[35] https://www.semantic-mediawiki.org/wiki/Semantic_gardening

[36] https://www.mediawiki.org/wiki/Extension:Page_Forms

[37] https://www.mediawiki.org/wiki/Extension:External_Data

[38] https://d-nb.info/standards/elementset/gnd

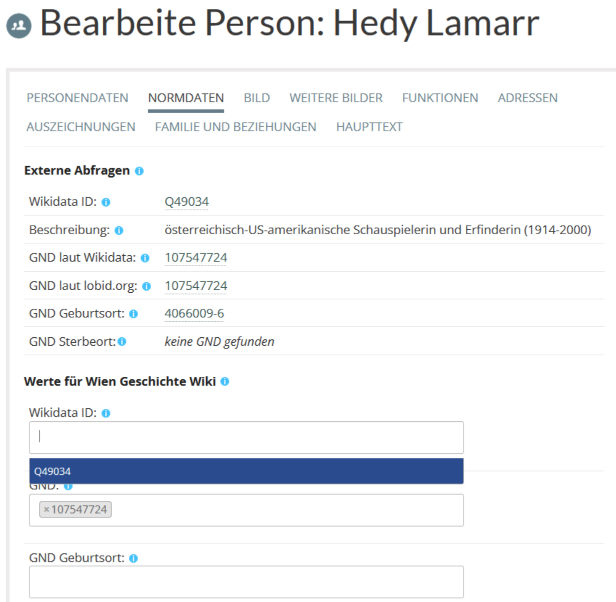

Figure 2: Form suggesting Wikidata IDs

as the GND found in Wikidata but also from Lobid's GND service[39], which provides linked open data services for libraries [6]. As the query based on the name may result in an incorrect match, the form only suggests the Wikidata and GND identifiers, and the user has to click on the suggested entry manually to confirm it, or enter a different GND.

3.3. Linking Data: Persistent Identifiers

In 2018 a first request was made by the Bavarian Academy of Sciences and Humanities[40] to deliver a *BEACON* file [30] that is commonly used to interlink portals that support the *GND*. This is a rather simple text file, so SMW was able to deliver it out of the box[41]:

```
#FORMAT: BEACON
#PREFIX: http://d-nb.info/gnd/
#TARGET: https://www.geschichtewiki.wien.gv.at/
Special:URIResolver/?curid=
#VERSION: 0.3
#CREATOR: Bernhard Krabina
#CONTACT: office@krabina.com
#HOMEPAGE: https://www.geschichtewiki.wien.gv.at/
index.php?title=Wien_Geschichte_Wiki/BEACON
#FEED: https://www.geschichtewiki.wien.gv.at/
index.php?title=Wien_Geschichte_Wiki/BEACON/
Personen-Bearbeitet&action=render
#INSTITUTION: Wien Geschichte Wiki
#TIMESTAMP: 20220720102536
#UPDATE: always
7512885-8||7462
1068930500||105
...
```

[39] https://lobid.org/gnd
[40] https://badw.de/en/
[41] https://www.geschichtewiki.wien.gv.at/Wien_Geschichte_Wiki/BEACON

The above file indicates that the GND 7512885-8 equals the identifier 7462 that can be addressed by https://www.geschichtewiki.wien.gv.at/Special:URIResolver/?curid=7462 which will resolve to the entry for "10er Marie", a still existing building established in 1740. While MediaWiki has the capability of delivering pages based on the identifier with https://www.geschichtewiki.wien.gv.at/?curid=7462, SMW adds the special page *Special:URIResolver* which adds the capability of content negotiation[42]: In our example, a browser pointing to https://www.geschichtewiki.wien.gv.at/Special:URIResolver/?curid=7462 will receive the wiki page (HTML), while a request for RDF will deliver the RDF/XML representation.

In order for a knowledge base to provide mechanisms to interlink with other knowledge bases, they need to be able to deliver persistent identifiers, which are essential for getting access and referring to library, archive and museum collection objects in a sustainable and unambiguous way [16] which is an integral part of the aforementioned linked open data principles.

This is especially easy for SMW-powered knowledge bases, because MediaWiki provides a page ID that remains persistent on page edits and renaming (aka *moving*) pages. Even deleted pages try to reclaim their original page ID once restored.[43] With SMW, the page ID can be assigned to a special property.[44] Once the page ID is known, it can be used in an URL, as described above, to link to the page regardless of any potential changes to the page name. Furthermore, MediaWiki also offers a revision ID that changes with every edit of a page. This way, all old revisions can still be retrieved, which is especially useful in case of the Vienna History Wiki: As many of the initial entries stem from the printed encyclopedia, the first version of the article that was taken from the original entry can still be shown. For example: the first entry about Hedy Lamarr is from November 2013 and can be retrieved by its revision ID 58367 via the URL https://www.geschichtewiki.wien.gv.at/index.php?title=Hedy_Lamarr&oldid=58367.

3.4. Exporting to Wikidata

Soon after the initial provision of the page ID property, the Wikidata community suggested the creation of a Wikidata property *Vienna History Wiki ID* which is now available in Wikidata as property *P7842*: https://www.wikidata.org/wiki/Property:P7842.

With the help of the already established GND, it becomes easier to uniquely define matching elements. This is helpful in the next required steps: adding the Vienna History Wiki ID to Wikidata and adding the Wikidata ID to the Vienna History Wiki.

[42] https://www.rfc-editor.org/rfc/rfc9110.html#name-content-negotiation
[43] https://www.mediawiki.org/wiki/Manual:Page_table#page_id
[44] https://www.geschichtewiki.wien.gv.at/Attribut:PageID

To query pages with their Vienna History Wiki ID, SMW provides several *result formats* and a dedicated search interface called *Semantic Search*.[45] On Figure 3 is an example of a query for all pages in the category *People* that have a property *WikidataID* and return the name of the pages, as well as the properties *PageID* and *WikidataID*:

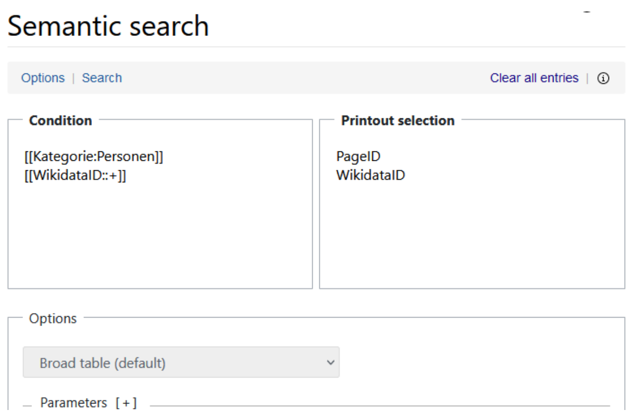

Figure 3: Semantic Search Interface

The conditions can be interpreted as "look for all pages of the category *People* that have a *WikidataID*". In the box on the right (printout selection), the information that should be returned in the result is given: "and give me the *PageID* and *WikidataID*". The result can be viewed in a table first and then exported to JSON, CSV, RSS and RDF formats.

SMW also provides API modules where queries like this can be submitted.[46]

A simple template[47] can be defined via the *templatefile* format of SMW,[48] that delivers the required format for the Wikidata *Quickstatement tool*.[49] Here is an example of the output:

```
Q7259 P7842 "32795" P1810 "Ada Lovelace"
/* Export from Vienna History Wiki 20211023143634 */
```

This command sequence, separated by tabs, can be explained as follows:

- Q7259: item (Ada Lovelace)
- P7842: property (Vienna History Wiki ID)
- 32795: value for the ID
- P1810: property (named as)
- Ada Lovelace: Name in the Vienna History Wiki
- comment (between /* */)

[45] https://www.geschichtewiki.wien.gv.at/index.php?title=Spezial:Semantische_Suche
[46] https://www.semantic-mediawiki.org/wiki/Help:API
[47] https://www.geschichtewiki.wien.gv.at/index.php?title=Vorlage:QuickstatementV1
[48] https://www.semantic-mediawiki.org/wiki/Help:Templatefile_format
[49] https://quickstatements.toolforge.org/#/

This structure results in a statements added to Wikidata with the Vienna History Wiki ID and a qualifier with the additional property *named as*, which can be important, because names of historical people can often be spelled quite differently: "Aegidius Aquila" (Q55123358) is known as "Ägidius Adler" in the Vienna History Wiki.

For the Quickstatement import a batch size of 10,000 was tried multiple times, but did result in timeouts. A reduced size of 5,000 at a time worked well. The error rate shown below is the initial error rate. There was no indication regarding the source of the errors shown, only *"No success flag set in API result"* was displayed on mouse-over of the error status indicator. After running *"Try to reset errors"* (sometimes multiple times, without changing anything else), the remaining post-error number is shown in brackets.

The final successful Quickstatement import done in October 2021 delivered the following results:[50]

- Batch #68684 with 1,104 People: 2,208 statements, 15 errors, 0.68% error rate, (0%)
- Batch #68779 with 5,000 People: 10,000 statements, 440 errors, 4.4% error rate (0.01%)
- Batch #68784 with 5,000 People: 10,000 statements, 755 errors, 7.55% error rate (0.03%)
- Batch #68791 with 180 items (23 Structures, 3 Events, 145 Organizations, 3 Memorials, 6 Topographic Objects): 360 statements, 155 errors, 43.01% error rate (0%)

3.5. Importing External IDs

Adding a WikidataID property in SMW is easy. The harder task is to align datasets that do not share common identifiers. For this reason a *reconciliation service* for Wikidata was implemented [8].

There is a W3C Community Group Draft Report that describes a reconciliation service API as implemented in OpenRefine 2.8 to 3.2.[51] OpenRefine was used successfully by the Vienna History Wiki editorial team to identify matching items.

With the extension *Data Transfer*[52] it was possible to import the Wikidata IDs from the reconciliation process in OpenRefine in CSV format to SMW. The simple structure of the UTF8 formatted CSV file is:

```
Titel,Person[WikidataID]
Abraham Myron Schein,Q61198351
Abraham a Sancta Clara,Q61137
Ada Christen,Q85002
Ada Lovelace,Q7259
...
```

The first column is the title of the wiki page into which the data will be imported, the second column indicates the

[50] The batches can be retrieved using the batch number at https://quickstatements.toolforge.org/#/batch/68684
[51] https://reconciliation-api.github.io/specs/0.1/
[52] https://www.mediawiki.org/wiki/Extension:Data_Transfer

Table 1
Vocabularies and Ontologies

Ontology	Year	Comment
Generic		
Schema.org	2011	preferred by Google and Vienna map project
DBPedia	2007	based on Wikipedia categories
Wikidata	2012	collaborative and messy, but with future potential, see [13]
GND	2012	used for GND property
OWL	2009	used by SMW
FOAF	2000	used by SMW
SKOS	1997	used for English category labels
Dublin Core	1995	widely used vocabularies
PROV	2013	for provenance information
Archival domain		
OAI-PMH	2001	OAI protocol, Google prefers standard XML sitemaps
OAI-ORE	2006	Resource Maps
Bibliographic domain		
METS/MODS	2001, 2002	data models for bibliographic description
FRBR	2005	foundation for RDA
BIBO	2009	for citations and bibliographic references
RDA	2010	expressed in RDF
BIBFRAME	2012	expressed in RDF
VIAF	2003	international authority file
Other domains		
AAT thesaurus	1990	controlled vocabulary for arts & architecture
CIDOC CRM	1999	for cultural heritage
Iconclass	1994	for images, used by Wien Museum
Europeana data model	2010	for institutions participating in Europeana
DCAT	2014	for data catalogs

template *Person* and the field *WikidataID* in the template into which this information will be placed. The import can be configured to leave any other information already on the page unchanged.[53]

3.6. Vocabularies and Ontologies

In order to further develop the knowledge base, the re-use of existing ontologies and controlled vocabularies is key. Table 1 gives an overview of potential ontologies to be considered.

In 2021, the editorial team was approached by a project team in Vienna's city administration responsible for piloting a new mapping solution for the city. As integrating data from many different sources is key in this project, a discussion about relevant categories from the Vienna History Wiki started. Since the preferred choice for the mapping solution was Schema.org, because it is the most commonly used ontology on the web [13], the decision was made to use Schema.org.

3.7. Schema.org

Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the internet.[54]

Re-using external vocabularies like Schema.org in SMW is quite easy. There are two steps involved, illustrated here:

1. Creating a wiki page called *MediaWiki:Smw import schema* that holds all the terms to be used with their SMW property datatype.[55]
2. Adding the annotation `[[Imported from::schema:alternateName]]` to the property page (in this case: *Attribut:Andere Bezeichnung*.[56]) The same can be done for categories: Adding the annotation `[[Imported from::schema:AdministrativeArea]]` to the category page *Kategorie:Bezirk* (category:district). For SMW users not aware of linked open data, this is an important step often omitted, because property declarations are a common step in SMW setup, while adding annotations to categories is completely optional.

In case the *imported from* annotation uses a term that is not specified in the *MediaWiki:Smw import schema* page, this is indicated as an annotation problem by SMW. Because the namespace *MediaWiki:* is used for the vocabulary definition, only users with administrative rights can change the vocabulary.

[53]Example of such an edit: `https://www.geschichtewiki.wien.gv.at/index.php?title=Ada_Lovelace&type=revision&diff=450368&oldid=305768`
[54]`https://schema.org`
[55]`https://www.geschichtewiki.wien.gv.at/MediaWiki:Smw_import_schema`
[56]`https://www.geschichtewiki.wien.gv.at/Attribut:Andere_Bezeichnung`

Table 2
Schema.org Usage on Categories

Category	translation	schema:	# cat	coverage
Adressen	Adresses	PostalAddress	3	33%
Bauwerke	Structures	LandmarksOrHistoricalBuildings	12	58%
Begriffe	Terms	DefinedTerm	7	14%
Bilder	Images	ImageObject	1	100%
Ereignisse	Events	Event	1	100%
Erinnern	Memorials		10	0%
Grenzsteine	Border stones		1	0%
Karten	Maps	Map	9	11%
Organisationen	Organizations	Organization	65	48%
Personen	People	Person	1	100%
Sonstiges	Other		1	0%
Topografische Obj.	Topographic Objects	Place	20	30%
Verweise	Redirects		9	11%
			140	36%

With a coverage of 36%, we were able to find representations for roughly one third of the classes used in the historical context in Schema.org (see Table 2). For very special terms like *border stone*, *park sign*, *nazi institution* or *passage* this is not surprising, but surprisingly, some very general terms like *forest*, *foundation*, *industry*, *market* or *public bath* are missing.

A 30% coverage can also be found in schema.org properties: for 33 out of 105 custom properties, a representation could be used. Important ones being *alternateName*, *award*, *hasOccupation*, *license*, *caption*, *startDate*, *endDate*, *birthPlace*, *deathPlace*, *geo*, *isRelatedTo* and *sameAs*.

The vocabularies used by default in SMW is described in the SWIVT ontology.[57] Further vocabularies that have been introduces in the Vienna History Wiki are:

- *gndo:gndIdentifier* for the GND (Ingegrated Authority File)
- *foaf:depiction* for a property describing places depicted in images
- *skos:prefLabel* and *skos:scopeNote* for a property to describe the main categories in several languages. For multilingual description of properties, SMW offers the built-in properties *Preferred property label* and *Property description*.[58]

Other vocabularies have not been investigated yet, due to the overwhelming amount of options described in Table 1 and the lack of ontology preferences from Vienna's cultural heritage institutions. However, plans are to re-use more domain-specific vocabularies for terms not covered by Schema.org and to further investigate the suitability of Wikidata properties, as suggested in [13].

[57] https://www.semantic-mediawiki.org/o/swivt/
[58] https://www.semantic-mediawiki.org/wiki/Help:Special_properties

3.8. RDF Dump

The result of the process described above is not only a better RDF representation on the individual pages. An RDF dump that is generated every Sunday is now available. Due to the huge amount of address data that was imported in the *Adresse:* namespace, the dump is now around 5 GB in size (compressed to a 200MB file), because it includes almost 280,000 entries from the data import.

The current master version of SMW features the option of specifying namespaces to be included in the RDF dump, which will be used once it is available in the Vienna History Wiki.[59] A description and download option is provided at https://www.geschichtewiki.wien.gv.at/RDF.

The RDF representation of Hedy Lamarr now changed because of the re-used vocabularies[60]. The RDF result format has the option of output the RDF not only in RDF/XML, but also in Turtle syntax[61] (see next page).

The GND identifier is now represented by *gndo: gndIdentifier* and the former property "Beruf" (occupation) by *schema:hasOccupation*, without needing to rename the property, which is still referred to as "Beruf" in the wiki. The property "Abweichende Namensform" remains indicated as *property:*, because *schema:additionalName* did not seem a good match (as it is used for middle names instead of alternate names)[62] and the better matching *schema:altnernateName*[63] is a property related to things, not people. Due to the usage of the datatype *external identifier*, SMW automatically adds the *skos:exactMatch* annotation.

[59] https://github.com/SemanticMediaWiki/SemanticMediaWiki/issues/5031
[60] In SMW there are also RDF representations of individual pages. See e.g. the output of the page for Hedy Lamarr at https://www.geschichtewiki.wien.gv.at/Spezial:RDF_exportieren/Hedy_Lamarr
[61] https://www.geschichtewiki.wien.gv.at/Spezial:Semantische_Suche/-5B-5B:Hedy-20Lamarr-5D-5D/-3FPersonenname/-3FAbweichende-20Namensform/-3FGeschlecht/-3FPageID/-3FGND/-3FWikidataID/-3FGeburtsdatum/-3FGeburtsort/-3FSterbedatum/-3FSterbeort/-3FBeruf/format%3Drdf/syntax%3Dturtle
[62] https://schema.org/additionalName
[63] https://schema.org/alternateName

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix swivt: <http://semantic-mediawiki.org/swivt/1.0#> .
@prefix wiki: <http://www.geschichtewiki.wien.gv.at/Special:URIResolver/> .
@prefix category: <http://www.geschichtewiki.wien.gv.at/Special:URIResolver/Category-3A> .
@prefix property: <http://www.geschichtewiki.wien.gv.at/Special:URIResolver/Property-3A> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix wikiurl: <https://www.geschichtewiki.wien.gv.at/> .
@prefix schema: <https://schema.org/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix gndo: <https://d-nb.info/standards/elementset/gnd> .

[ rdf:type  owl:Ontology ;
  swivt:creationDate  "2022-12-04T17:56:57+01:00"^^xsd:dateTime ;
  owl:imports  <http://semantic-mediawiki.org/swivt/1.0> ]

wiki:Hedy_Lamarr
  rdf:type  swivt:Subject ;
  rdfs:label  "Hedy Lamarr" ;
  rdfs:isDefinedBy  <https://www.geschichtewiki.wien.gv.at/Special:ExportRDF/Hedy_Lamarr> ;
  swivt:page  <https://www.geschichtewiki.wien.gv.at/Hedy_Lamarr> ;
  swivt:wikiNamespace  "0"^^xsd:integer ;
  swivt:wikiPageContentLanguage  "de-formal" ;
  property:Personenname  "Lamarr-Kiesler, Hedy" ;
  property:Abweichende_Namensform  "Kiesler, Hedwig Eva Maria" ,  "Kiesler, Hedy" ;
  schema:gender  "weiblich" ;
  schema:identifier  "30218" ;
  skos:exactMatch  <https://www.geschichtewiki.wien.gv.at/Special:URIResolver/?curid=30218> ,
  <http://d-nb.info/gnd/107547724> ,  <http://www.wikidata.org/entity/Q49034> ;
  gndo:gndIdentifier  "107547724" ;
  schema:sameAs  "Q49034" ;
  schema:birthDate  "1914-11-09Z"^^xsd:date ,  "2420445.5"^^xsd:double ;
  schema:birthPlace  "Wien" ;
  schema:deathDate  "2000-01-19Z"^^xsd:date ,  "2451562.5"^^xsd:double ;
  schema:deathPlace  "Altamonte Springs (Florida, Vereinigte Staaten von Amerika)" ;
  schema:hasOccupation  "Filmschauspielerin" ,  "Erfinderin" .

<http://semantic-mediawiki.org/swivt/1.0#creationDate> rdf:type owl:DatatypeProperty .
<http://semantic-mediawiki.org/swivt/1.0#page> rdf:type owl:ObjectProperty .
<http://semantic-mediawiki.org/swivt/1.0#wikiNamespace> rdf:type owl:DatatypeProperty .
<http://semantic-mediawiki.org/swivt/1.0#wikiPageContentLanguage> rdf:type owl:DatatypeProperty .
<http://www.geschichtewiki.wien.gv.at/Special:URIResolver/Property-3APersonenname> rdf:type owl:DatatypeProperty .
<http://www.geschichtewiki.wien.gv.at/Special:URIResolver/Property-3AAbweichende_Namensform> rdf:type owl:DatatypeProperty .
<https://schema.org/gender> rdf:type owl:DatatypeProperty .
<https://schema.org/identifier> rdf:type owl:DatatypeProperty .
<http://www.w3.org/2004/02/skos/core#exactMatch> rdf:type owl:ObjectProperty .
<https://d-nb.info/standards/elementset/gndgndIdentifier> rdf:type owl:DatatypeProperty .
<https://schema.org/sameAs> rdf:type owl:DatatypeProperty .
<https://schema.org/birthDate> rdf:type owl:DatatypeProperty .
<https://schema.org/birthPlace> rdf:type owl:DatatypeProperty .
<https://schema.org/deathDate> rdf:type owl:DatatypeProperty .
<https://schema.org/deathPlace> rdf:type owl:DatatypeProperty .
<https://schema.org/hasOccupation> rdf:type owl:DatatypeProperty .

# Created by Semantic MediaWiki, https://www.semantic-mediawiki.org/
```

4. Analysis

The following sections summarize the approaches and results of three different analysis steps delivering empirical results: an online survey, access statistics, as well as an analysis of the knowledge graph that is now available for re-use in the form of an RDF dump.

4.1. Online Survey

An online survey was conducted, that was available to visitors of the Vienna History Wiki in April 2019. Users were invited to participate in the short survey via a site notice[64] that was displayed on every wiki page.

454 users completed the survey — which is double the number of users who completed a comparable survey in 2015 [17]. The population of users who visited the wiki during the online survey can only be estimated: 154,993 visits in April 2019 corresponds to a response rate of 0.3 percent. The methodological implications of this survey approach and the results of the 2015 survey are described in more detail in [17]. The online survey data is available online.[65]

Compared to the first survey in 2015, users have become younger but are still dominantly male (59%) and relatively old, with the largest age group being 55-64 years old (compared to 65-74 year olds in 2015, see Figure 4). This corresponds with the (subjective) impression of archivists and librarians of the regular users visiting Vienna's archive and library physically.

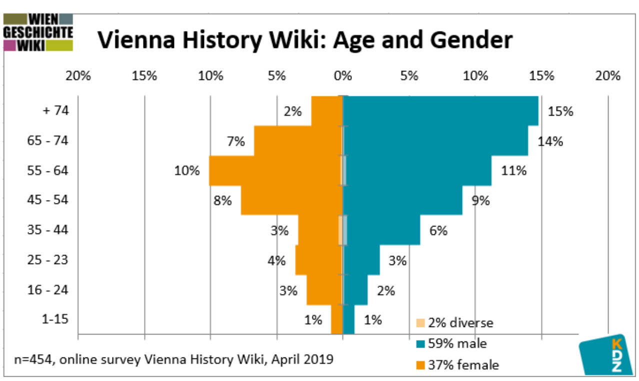

Figure 4: User Demography

Two thirds of the users visit as a result of an internet search. One third are first-time users (50% in 2015, see Figure 5) and 70% use the wiki out of private interest (see Figure 6).

The general satisfaction of the users is very high: 94% indicate that they will use the Vienna History Wiki again (see Figure 7).

Nevertheless, the majority of the users are not aware that they could also edit content. The number of those who have engaged in editing (8%), are quite satisfied with the help

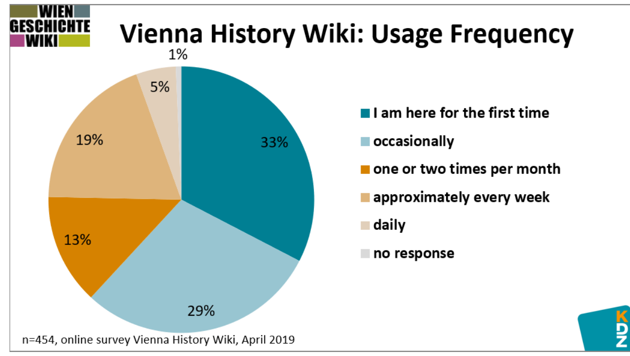

Figure 5: Frequency of Use

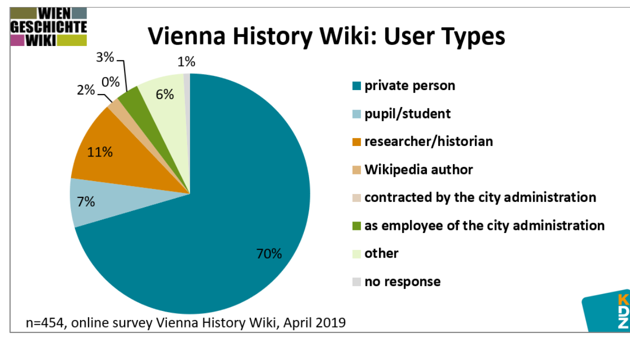

Figure 6: Types of Users

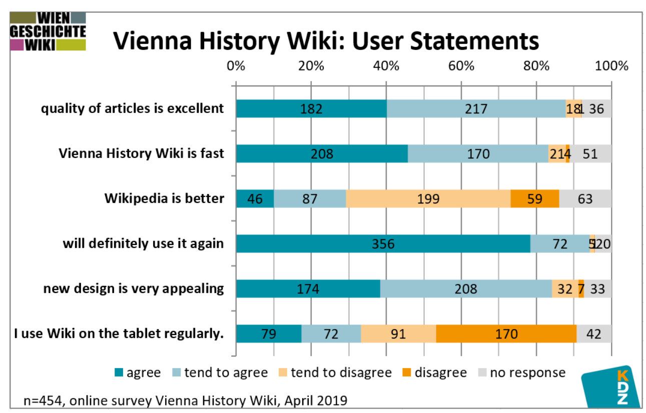

Figure 7: Statements regarding User Satisfaction

provided (74.3% strongly agree or agree), the editing itself (74.3%) and the review process (80%) (see Figure 8).

4.2. Access Statistics

A user account is only needed for people intending to add or edit content. Figure 9 shows the development of user account creation since the beginning, giving a good indication of the interest in participation.

Each year, between 100 and 300 new user accounts are created by citizens (*DYN* prefix), which can be monitored on the *Special:Userlist*[66] page. The *WL* users are mainly teachers, who have been logged in automatically whenever they

[64] https://www.mediawiki.org/wiki/Manual:Interface/Sitenotice

[65] http://data.opendataportal.at/dataset/online-umfrage-wien-geschichte-wiki-april-2019

[66] https://www.geschichtewiki.wien.gv.at/Spezial:Benutzer

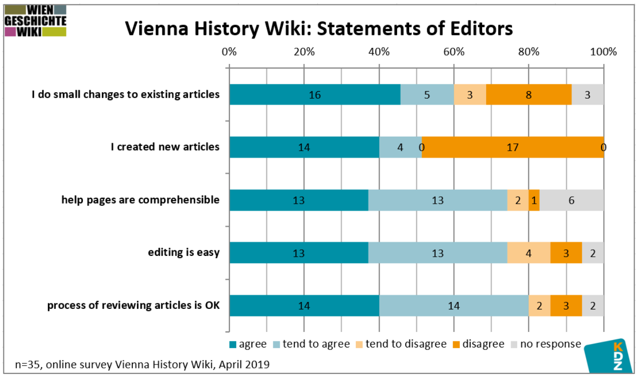

Figure 8: Statements from the Editors

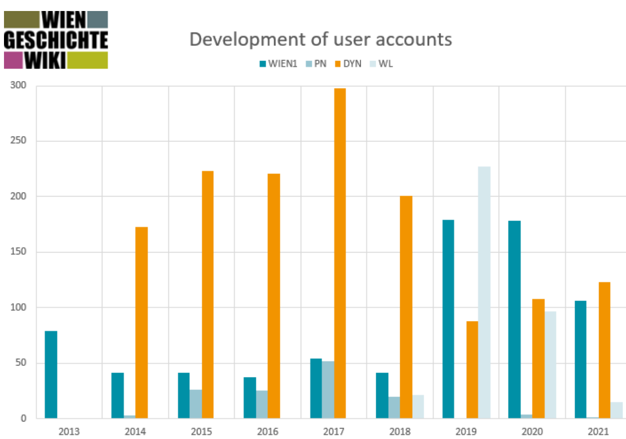

Figure 9: User Account creation each year.

use school equipment since 2018. What is also interesting is a look at the list of active users[67] (with contributions in the last 30 days): This regularly shows that around half of all active users are based in the city's administration.

An analysis of web server statistics is available on the Vienna City Administration's intranet. The analysis is based on a configuration of the log file analysis software Webalizer V. 2.23 as well an improved analysis mechanism provided by the commercial service Siteimprove.com.

On average, around 250,000 visits per month are counted (see Figure 10). The three COVID-19 related lockdowns in Austria are visible in the usage statistics as a result of people increasingly working from home, with peaks of around 300,000 from March to May 2020 (1st lockdown), in November 2020 (2nd lockdown) and in January 2021 (3rd lockdown) which is in line with the observation that the pandemic has accelerated digital content production and interaction of the European National Libraries [26].

4.3. Knowledge Graph Analysis

Metadata available in SMW was exported for the main content categories for the purpose of a knowledge graph analysis. Other categories were left out, because they are used for purposes other than organizing the content and because the Address category has a size of more than 240,000 instances which would dominate the result.

A typical SMW query looks as follows, querying all pages of a specific category with requested properties provided by SMW as printouts (indicated by question mark) and further querying options, e.g. specifying the CSV format for data download:

```
{{#ask: [[Category:Bauwerke]]
  |?Category
  |?Creation date
  |?Is a new page
  |?Last editor is
  |?Modification date
  |format=csv
}}
```

The resulting CSV files were further processed with spreadsheet programs. Necessary processing steps were to remove duplicate category information, because MediaWiki provides specific *tracking categories*,[68] and some pages were erroneously entered into more than one category. The modification date was used to calculate the number of months since the last edit. Some information that seemed interesting at first, was discarded in the process. For example, the data allows identification of the number of edits authored by the city administration, because their usernames begin with *WIEN1*, while the usernames of external users begin with *DYN*. However, due to the scholarly process the wiki is based on, it is obvious that edits of external users often get adjusted by the editorial team, leaving a *WIEN1* user as the last editor, even though a *DYN* user may have contributed considerable editing.

The results of the knowledge graph analysis are summarized in Table 3. The number of total pages in each category is shown (redirecting pages were not counted), along with the number of entries that were originally in the *Czeike* encyclopedia. The *Extensions* column indicates the percentage change of new entries related to the original *Czeike* entries. *Recency* is the average number of months from the date a page was last edited to the time of the analysis (October 28, 2021). The *Subcat* column shows the number of subcategories of the respective main category.

The knowledge graph analysis shows that, on average, edits in the Vienna History Wiki are 17 months old. Generally speaking the most recently edited pages are in the categories Memorials and Structures, the least recently edited entries are in the category Other, followed by Events, Terms and People. Once the content of the *Czeike* encyclopedia had been added and expanded, a new Maps category was established, and most of the content in the categories Events, Memorials and Organizations does not stem from the *Czeike* files. The category with the fewest new entries since *Czeike* is Topographic Objects, with only 13.9% new articles. The number of subcategories varies, with several still having no

[67] https://www.geschichtewiki.wien.gv.at/Spezial:Aktive_Benutzer

[68] https://www.mediawiki.org/wiki/Help:Tracking_categories

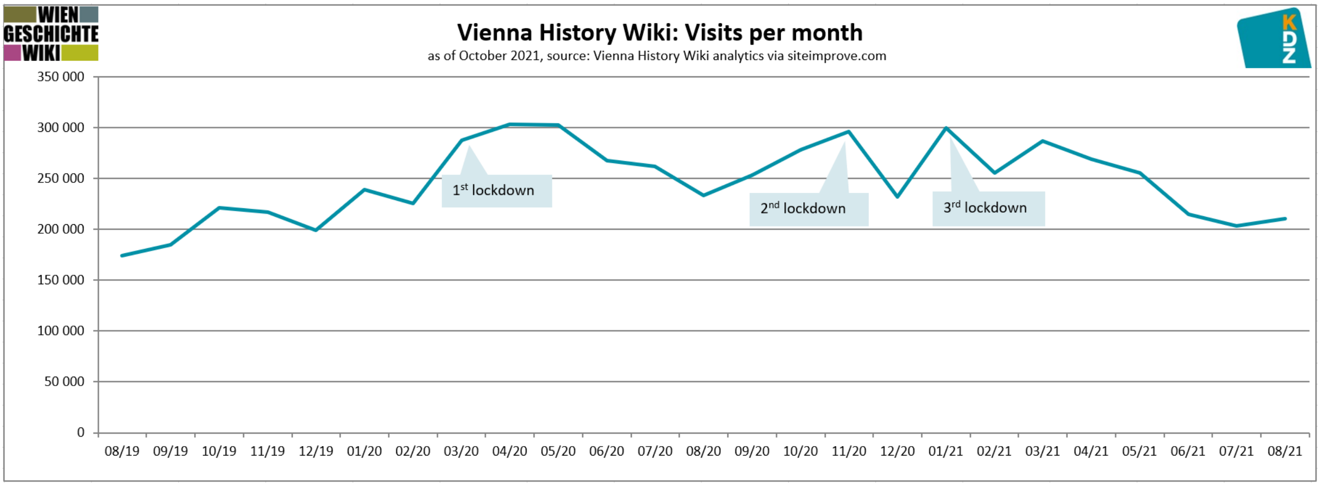

Figure 10: Visits per Month and Lockdowns in Austria

Table 3
Knowledge Graph Analysis

Category	Pages	*Czeike*	Extension	Recency	Subcat
People	16,259	8,383	93.95%	22.2	0
Topographic Obj.	11,018	9,672	13.9%	12,4	19
Structures	5497	3,730	47.4%	9.2	11
Organizations	4,867	1212	301.6%	20.7	64
Events	944	159	493.7%	25.9	0
Memorials	2,821	487	479.3%	5.3	9
Maps	266	0		17.6	8
Terms	1,448	657	120,4%	24.1	6
Other	2,564	1,935	32.5%	47.3	0
	45684	26235	74.13%	17.2	13

Table 4
Knowledge Graph Metrics

Metric	Value	Comment
# Instances	326,006	pages
# Assertions	27,203,391	property values
# Classes	180	categories
# Relations	461	properties
Avg. depth of class tree	1.31	only main cat.
Avg. branching factor of class tree	1.11	only main cat.

subcategories at all and the deepest Organization ontology boasting an impressive 64 subcategories.

Table 4 shows some knowledge graph metrics as described in a comparison of publicly available knowledge graphs [14].

The most straightforward way to assess the content focus of a knowledge graph is to look at the size of the extension of its classes [14]. A visual representation can be seen in Figure 11.

5. Discussion

Motivational aspects play a significant role in the willingness to participate in media-supported co-curation activities. While digital media can support co-curation, they have to be carefully designed to overcome challenges of authority and motivation inherent to participatory processes in cultural heritage institutions [3]. The online survey from 2019 showed a high degree of satisfaction among both readers and editors: 80% of editors are satisfied with the process of reviewing articles, even though this process is governed by the editorial team formed by the cultural heritage institutions and differs from the processes in Wikipedia.

User account development has changed over time, peaking at almost 300 new citizen-held accounts in 2017. While this number has decreased to around 100 additional accounts each year over the last three years, this should not be considered a sign of declining interest: User accounts created years ago will continue to work for users editing years later. Also, the list of active users (with edits in the last 30 days) is split approximately down the middle between citizens and the city administration's editorial team. Furthermore, the number of accounts attributed to the city of Vienna (including teachers) has increased noticeably in the last three years, evidencing the importance of the Vienna History Wiki for Vienna's cultural heritage institutions and municipal departments.

In spring 2021, several noteworthy restructuring activities were carried out in the Vienna History Wiki: (1) use of subcategories in all main categories (e.g. selecting *Bridge* in the field *Type of Object* now adds the page to the category *Bridge*, which is a subcategory of *Topographic Objects*), (2) harmonization of property usage (previously, some categories had used *Date from* while others had used *Year from*) and finally (3) implementation of Schema.org as base vocabulary. The choice of using schema.org as much as possible was not a decision based on the best matching vocabulary or the best coverage (terms in the ontology matching those used in the Vienna History Wiki), but rather the result of a request by the municipal department in charge of integrating

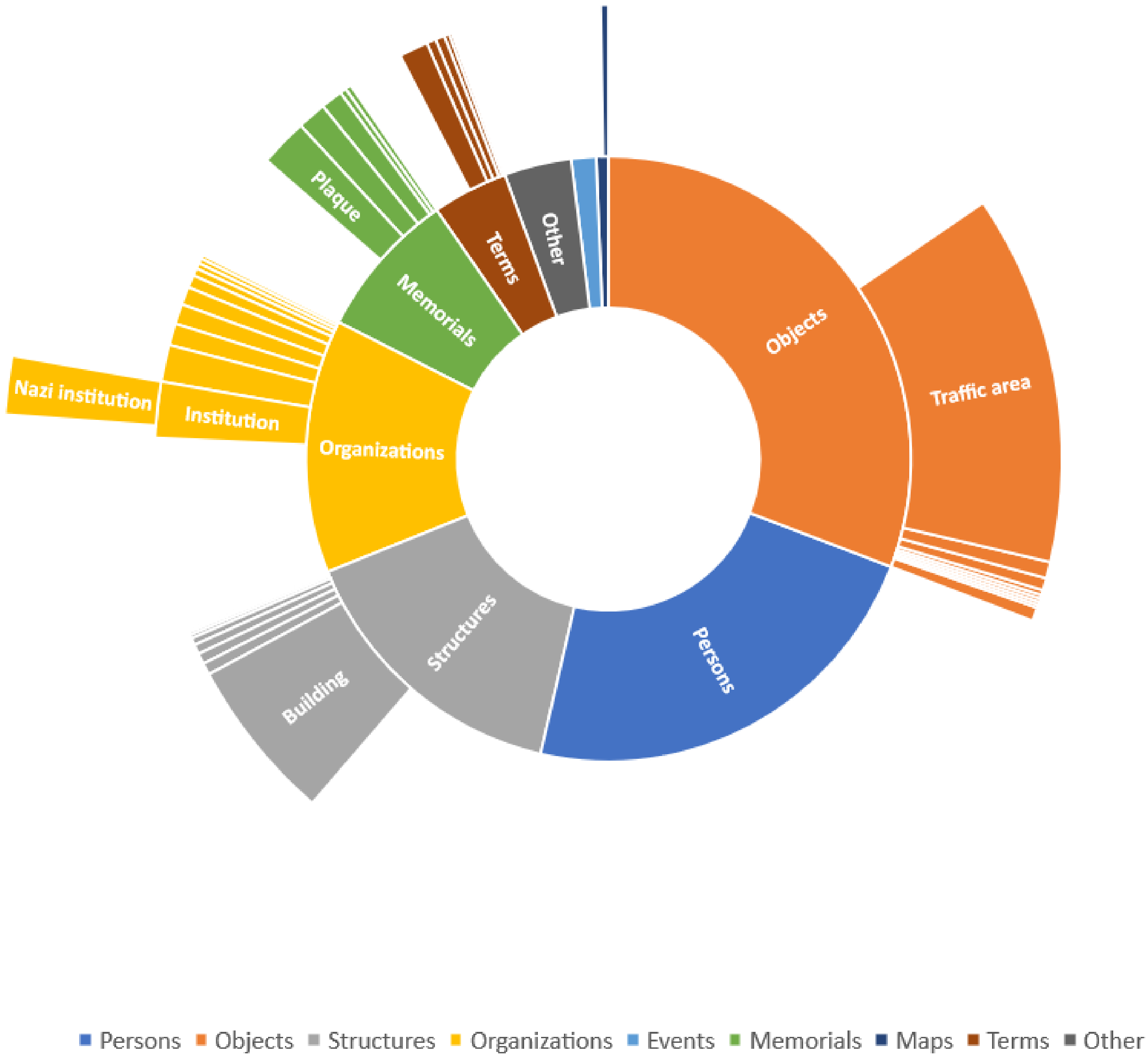

Figure 11: Instances in Vienna History Wiki

data from the wiki into the city map. Additionally, almost 280,000 pages of the open data address dataset, containing all the known addresses in Vienna, were imported into the wiki *Adresse:* namespace in May 2021.

Since its launch, the Vienna History Wiki has become the central historical platform for cultural heritage institutions in Vienna that use the knowledge base for their work. The commemoration day index [17], a product delivered on CD-ROM and paper to Vienna's municipal departments, is now replaced by a query interface and pre-defined pages[69] querying the knowledge graph.

The knowledge graph was expanded several times, partly by research from outside of the cultural heritage institutions. A specific property[70] documents the main source of content that has been transferred to the wiki, partly manually, partly by imports. Some examples are:

- Content from several books apart from the original *Czeike*. E. g. *Paul Harrer: Wien, seine Häuser*, *Wolfgang Wirsig: Wiener Hofnamen*, *Historischer Atlas von Wien*, *Murray G. Hall: Österreichische Verlagsgeschichte*.
- Data from internal databases like *POLAR*[71] (database of the biographies of Viennese politicians) or the commemoration day index application.
- Data from external research projects like *POREM* (Politics of Remembrance and the Transition of Public Spaces. A Political and Social Analysis of Vienna by

[69] https://www.geschichtewiki.wien.gv.at/Gedenktage

[70] https://www.geschichtewiki.wien.gv.at/index.php?title=Attribut:Quelle

[71] https://www.wien.gv.at/advuew/internet/AdvPrSrv.asp?Layout=histpolauswahl&Type=S

the Vienna University)[72] and *Was haben diese Plätze schon gesehen?*, a project about demonstrations in Vienna by the Institute for Jewish History in Austria.[73]

The improved RDF representation and the availability of a complete RDF dump are now the source that the municipal IT department relies on to incorporate data from the knowledge graph into the future version of the city map.

5.1. Lessons Learned

Creating unique identifiers and re-using existing vocabularies in SMW is quite straightforward. The simple wiki pages naming external vocabularies and their data types can be reused easily. As adding properties to category pages is usually not required, this is an extra step to consider when attempting to provide proper class definitions.

Editing properties (and categories) as well as importing data can result in a considerable increase of jobs in the *job queue*.[74] Figure 12 illustrates the monitoring of the job queue via WikiApiary[75]

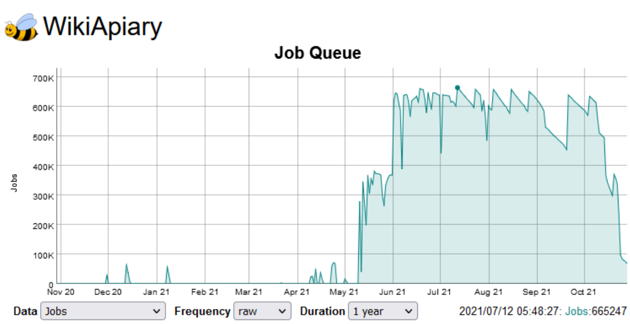

Figure 12: Vienna History Wiki Job Queue

It is recommended to consider re-using vocabularies early in the process of setting up a knowledge base. While this is a general good practice in ontology development, in SMW it is better to care of the property definitions before mass-importing content, as doing it the other way around would result in a lot of workload for the job queue.

5.2. Shortcomings of SMW

SMW provides an RDF/XML representation that is indicated by `<link rel="alternate" type="application/rdf+xml">` in the HTML source of each page. In the RDF, SMW can use the Schema.org vocabulary, but since Schema.org forces websites to use a Microdata, RDFa, or JSON-LD representation, SMW cannot deliver Schema.org-compliant HTML: The Schema.org validator does not find any elements.[76]

While from a Semantic Web perspective, ignoring an RDF/XML representation and forcing other formats can be considered rather a shortcoming of Schema.org than of SMW, SMW should still provide the required formats. Conceptually, Microdata and RDFa do not go well with SMW, as there are two ways of annotating content on a wiki page: *in-text annotation*[77] in the form of `[[Birth Place::Vienna]]` that just renders "Vienna" on the wiki page and stores the triple *Pagename -> Birth Place -> Vienna* in the database when the page is saved. The second option is not to annotate in the text, but to provide form fields for metadata instead and to use the *set:* parser function in the resulting template to declare property values, e. g. `{{#set:Birth Place=Vienna}}`, which does not display anything on the wiki page, and is thus referred to as *silent annotation*[78].

While `<span>` is one of the few allowed HTML tags in MediaWiki's wikitext syntax[79], using `<span itemprop="schema:birthDate">Vienna</span>` in Microdata or `<span property="schema:birthDate">Vienna</span>` in RDFa is not supported by wikitext editors or SMW. Thus, a JSON-LD representation is the most straightforward implementation, as it could potentially be integrated into the *Semantic Meta Tags* extension.[80]

Due to the lack of a connected RDF store, SMW cannot directly provide a *SPARQL endpoint*. The extension *RDFio* is currently not compatible with the latest versions of MediaWiki and SMW.[81] It would be necessary to set up SMW to store its data in an external triple store.[82] However, as it is possible to connect SMW to RDF stores (also at any later point in time), this is more a shortcoming of the Vienna History Wiki, rather than of SMW itself.

Performance is an issue, especially when large amounts of data being displayed on interactive maps can result in page loading times of more than 30 seconds.[83] Also, exporting large amounts of data (more than 5,000 items at a time) might result in the web server timing out before SMW can deliver the results. Therefore, rethinking to allow the connection of SMW to Elasticsearch[84] or an RDF database[85] should be considered.

6. Conclusions and Future Work

The Vienna History Wiki is still growing, and demonstrates a *satisfied user base*. From a collaboration perspective, the awareness among users of their ability to add and edit content should be improved. The editorial process involving approval of a page status has proven suitable for a scholarly wiki, without the risk that historians and curators may lose leverage over historical content, knowledge, and their handle on digital historiography [25]. Manual editing

[72] https://www.porem.wien
[73] https://plaetze.at
[74] https://www.mediawiki.org/wiki/Manual:Job_queue
[75] https://wikiapiary.com/wiki/Wien_Geschichte_Wiki
[76] https://validator.schema.org/#url=https://www.geschichtewiki.wien.gv.at
[77] https://www.semantic-mediawiki.org/wiki/Help:In-text_annotation
[78] https://www.semantic-mediawiki.org/wiki/Help:Setting_values
[79] https://www.mediawiki.org/wiki/Help:Formatting#HTML_tags
[80] https://github.com/SemanticMediaWiki/SemanticMetaTags/issues/69
[81] https://www.mediawiki.org/wiki/Extension:RDFIO
[82] https://www.semantic-mediawiki.org/wiki/Help:Using_SPARQL_and_RDF_stores
[83] https://www.geschichtewiki.wien.gv.at/Karte_der_Erinnerung
[84] https://www.semantic-mediawiki.org/wiki/Help:Using_SPARQL_and_RDF_stores
[85] https://www.semantic-mediawiki.org/wiki/Help:Using_SPARQL_and_RDF_stores

via forms in conjunction with the possibility of exporting and importing structured data supports the data curation process.

To date, there has been little work on the routines that organically emerge within peer-production with scarce empirical evidence [1]. Therefore, a logical next research goal is to investigate whether editing in the Vienna History Wiki follows similar emerging routines as those investigated by [1] for Wikipedia articles.

The growing number of users from the city of Vienna indicates that a qualitative study could give deeper insights into how archivists, librarians, historians, teachers, and other users from the city as well as from outside the editorial team see the development of the knowledge graph.

Adding *unique identifiers* was an important step in providing better linked open data. Using Schema.org as vocabulary provided a good starting point towards better linked knowledge. With a coverage of 36% for classes and 30% for properties, this is in line with the findings of [13] arguing a lack of completeness and incentives to annotate non-commercial knowledge with Schema.org.

However, Schema.org still serves as a suitable *base ontology* for our use case that can be potentially extended by future versions of Schema.org and also by using more specialized ontologies in the future.

Publishing content to *Wikidata* using a customized export from SMW as input for the Wikidata Quickstatements tool is a first step. Identifying matching items that do not yet share a common identifier is the next step, especially for categories other than the *People* category, for which this has already been done. Despite *OpenRefine* being a useful available tool that can import data directly from SMW and deliver data back via CSV export, implementation of a *reconciliation process* directly in SMW could be of value and is being discussed in the SMW community.[86]

Aside from being in line with the *open government strategy* of the city's administration, publishing and interlinking content from a special knowledge graph with Wikidata also provides benefits regarding *content quality*: 11 of the 11,000 statements we published to Wikidata were manually reverted by the Wikidata community because of errors in the identification of matching items. These corrections were taken into account by the editorial team, thus improving the overall quality of the articles.

A more thorough investigation of the Vienna History Wiki knowledge graph according to [12] as well as more emphasis on knowledge graph quality metrics such as *precision* and *recall* [32] is needed and should be tackled, once the interlinking and exchanging of content between Wikidata and the Vienna History Wiki has been improved. Provision of a *SPARQL endpoint* via one of the SMW triple store connectors would be a significant improvement not only for knowledge graph researchers but also for historians.

Future research comparing current systems for collaboratively maintaining knowledge graphs is needed with a focus not only on features or performance, but especially also on the sustainability aspect, since specifically in the open source world, it remains to be seen if new solutions like *CLEF* [7] will survive the period of first funded research and development.

Despite some shortcomings, we were able to demonstrate that Semantic MediaWiki is a very powerful and easy-to-use open source tool for setting up and maintaining special-interest knowledge graphs. As a result of the work described in this paper, the now available and regularly updated RDF dump can be re-used by researchers.

[86] https://github.com/SemanticMediaWiki/SemanticMediaWiki/issues/4254

CRediT authorship contribution statement

Bernhard Krabina: Study conception and design, Data collection, Analysis and interpretation of results, Manuscript preparation

Declaration of competing interest

The author was employed at KDZ – Centre for Public Administration Research in Vienna, Austria. KDZ is a non-profit association which was contracted by the City of Vienna to maintain and improve the Vienna History Wiki.

Data availability

Data will be made available on request.

Acknowledgements

Part of this research was funded by the City of Vienna. Thanks go to the Vienna City Archive, the Vienna City Library and the KDZ - Centre for Public Administration Research.

References

[1] Arazy, O., Lindberg, A., Lev, S., Wu, K., Yarovoy, A., 2020. Emergent Routines in Peer-Production: Examining the Temporal Evolution of Wikipedia's Work Sequences. ACM Transactions on Social Computing 3, 5:1–5:24. URL: https://doi.org/10.1145/3366711, doi:10.1145/3366711.

[2] Beck, R., Rai, A., Fischbach, K., Keil, M., 2015. Untangling knowledge creation and knowledge integration in enterprise wikis. Journal of Business Economics 85, 389–420. URL: http://link.springer.com/10.1007/s11573-014-0760-2, doi:10.1007/s11573-014-0760-2.

[3] Blaschitz, E., Mayr, E., Oppl, S., 2022. Too Low Motivation, Too High Authority? Digital Media Support for Co-Curation in Local Cultural Heritage Communities. Multimodal Technologies and Interaction 6, 33. URL: https://www.mdpi.com/2414-4088/6/5/33, doi:10.3390/mti6050033.

[4] Breitenfeld, A., Mackeprang, M., Hong, M.T., Müller-Birn, C., 2017. Enabling Structured Data Generation by Nontechnical Experts URL: https://dl.gi.de/handle/20.500.12116/3264, doi:10.18420/MUC2017-MCI-0231. publisher: Gesellschaft für Informatik e.V.

[5] Bry, F., Schaffert, S., Vrandečić, D., Weiand, K., 2012. Semantic Wikis: Approaches, Applications, and Perspectives, in: Eiter, T., Krennwallner, T. (Eds.), Reasoning Web. Semantic Technologies for Advanced Query Answering. Springer Berlin Heidelberg, Berlin, Heidelberg. volume 7487, pp. 329–369. URL: http://link.springer.com/10.1007/978-3-642-33158-9_9, doi:10.1007/978-3-642-33158-9_9. series Title: Lecture Notes in Computer Science.

[6] Christoph, Pascal, F.S., 2017. Lobid API 2.0: Why and how. URL: https://hbz.github.io/lobid-blog/2017/06/08/lobid-api-why-how.html.

[7] Daquino, M., Wigham, M., Daga, E., Giagnolini, L., Tomasi, F., 2022. CLEF. A Linked Open Data native system for Crowdsourcing. Technical Report arXiv:2206.08259. arXiv. URL: http://arxiv.org/abs/2206.08259. arXiv:2206.08259 [cs] type: article.

[8] Delpeuch, A., 2020. Running a reconciliation service for Wikidata, in: Kaffee, L.A., Tifrea-Marciuska, O., Simperl, E., Vrandečić, D. (Eds.), Proceedings of the 1st Wikidata Workshop (Wikidata 2020) co-located with 19th International Semantic Web Conference (ISWC 2020), p. 8. URL: http://ceur-ws.org/Vol-2773/.

[9] Dittrich, J., 2019. Wikidata Use in Cultural Institutions. URL: https://upload.wikimedia.org/wikipedia/commons/e/e1/Research_Report_%E2%80%93_Use_of_Wikidata_in_GLAM_institutions_%282019-11%29.pdf.

[10] Feng, Y., Richards, L., 2018. A review of digital curation professional competencies: theory and current practices. Records Management Journal 28, 62–78. URL: https://www.emerald.com/insight/content/doi/10.1108/RMJ-09-2016-0022/full/html, doi:10.1108/RMJ-09-2016-0022.

[11] Godby, J., Smith-Yoshimura, K., Washburn, B., Davis, K.K., Detling, K., Eslao, C.F., Folsom, S., Li, X., McGee, M., Miller, K., 2019. Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage , 86doi:10.25333/FAQ3-AX08.

[12] Haller, A., Fernández, J.D., Kamdar, M.R., Polleres, A., 2020. What Are Links in Linked Open Data? A Characterization and Evaluation of Links between Knowledge Graphs on the Web. Journal of Data and Information Quality 12, 1–34. URL: https://dl.acm.org/doi/10.1145/3369875, doi:10.1145/3369875.

[13] Haller, A., Polleres, A., 2020. Are we better off with just one ontology on the Web? Semantic Web 11, 87–99. URL: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/SW-190379, doi:10.3233/SW-190379.

[14] Heist, N., Hertling, S., Ringler, D., Paulheim, H., 2020. Knowledge Graphs on the Web – an Overview. arXiv:2003.00719 [cs] URL: http://arxiv.org/abs/2003.00719. arXiv: 2003.00719.

[15] van Hoof, M., Evans, N., Inguaggiato, G., Marušić, A., Gordijn, B., Dierickx, K., van Zeggeren, D., Dunnik, H., Gesinn, A., Bouter, L., Widdershoven, G., EnTIRE and VIRT2UE consortia, 2022. The Embassy of Good Science – a community driven initiative to promote ethics and integrity in research. Open Research Europe 2, 27. URL: https://open-research-europe.ec.europa.eu/articles/2-27/v1, doi:10.12688/openreseurope.14422.1.

[16] Koster, L., 2020. Persistent identifiers for heritage objects. The Code4Lib Journal URL: https://journal.code4lib.org/articles/14978.

[17] Krabina, B., 2015. The Vienna history Wiki: a collaborative knowledge platform for the city of Vienna, in: Proceedings of the 11th International Symposium on Open Collaboration - OpenSym '15, ACM Press, San Francisco, California. pp. 1–8. URL: http://dl.acm.org/citation.cfm?doid=2788993.2789835, doi:10.1145/2788993.2789835.

[18] Krötzsch, M., Vrandečić, D., Völkel, M., Haller, H., Studer, R., 2007. Semantic Wikipedia. Journal of Web Semantics 5, 251–261.

[19] von Lucke, J., Große, K., 2014. Open Government Collaboration, in: Gascó-Hernández, M. (Ed.), Open Government: Opportunities and Challenges for Public Governance. Springer, New York, NY. Public Administration and Information Technology, pp. 189–204. URL: https://doi.org/10.1007/978-1-4614-9563-5_12, doi:10.1007/978-1-4614-9563-5_12.

[20] Ngomo, A.C.N., Auer, S., Lehmann, J., Zaveri, A., 2014. Introduction to Linked Data and Its Lifecycle on the Web, in: Reasoning Web. Reasoning on the Web in the Big Data Era. Reasoning Web: Reasoning Web International Summer School, p. 99.

[21] Nishanbaev, I., Champion, E., McMeekin, D.A., 2019. A Survey of Geospatial Semantic Web for Cultural Heritage , 28.

[22] Paulheim, H., 2016. Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web 8, 489–508. URL: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/SW-160218, doi:10.3233/SW-160218.

[23] Ryen, V., Soylu, A., Roman, D., 2022. Building Semantic Knowledge Graphs from (Semi-)Structured Data: A Review. Future Internet 14, 129. URL: https://www.mdpi.com/1999-5903/14/5/129, doi:10.3390/fi14050129.

[24] Sabharwal, A., 2015. Digital Curation in the Digital Humanities: Preserving and Promoting Archival and Special Collections. Chandos Publishing. Google-Books-ID: GpiKBAAAQBAJ.

[25] Sabharwal, A., 2017. Digital humanities and the emerging framework for digital curation. College & Undergraduate Libraries 24, 238–256. URL: https://www.tandfonline.com/doi/full/10.1080/10691316.2017.1336953, doi:10.1080/10691316.2017.1336953.

[26] Tartari, M., Giorgio, S.D., Prandoni, C., 2022. European Libraries and Their Virtual Users: How the Pandemic Affected Digital Production and Participation URL: http://ceur-ws.org/Vol-3160/short9.pdf.

[27] Tharani, K., 2021. Much more than a mere technology: A systematic review of Wikidata in libraries. The Journal of Academic Librarianship doi:10.1016/j.acalib.2021.102326.

[28] Tramullas, J., Sánchez-Casabón, A.I., Garrido-Picazo, P., 2021. Organización y descripción de información en wikis semánticos: Wikibase, Semantic Mediawiki y Cargo, Centro de Estudos Clássicos, Colibri. Universidade de Lisboa. pp. 229–238. URL: http://eprints.rclis.org/42661/.

[29] Troncoso, A.R.O., 2022. Ontology-Based Approach to Creating Semantic Wikis 15. doi:10.1145/3479012.

[30] Voß, J., Schindler, M., Thiele, C., 2011. Link server aggregation with BEACON URL: http://eprints.rclis.org/15407/.

[31] Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., Studer, R., 2006. Semantic Wikipedia, in: Proceedings of the 15th international conference on World Wide Web (WWW-06), ACM. doi:10.1145/1135777.1135863.

[32] Weikum, G., Dong, L., Razniewski, S., Suchanek, F., 2021. Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases. arXiv:2009.11564 [cs] URL: http://arxiv.org/abs/2009.11564. arXiv: 2009.11564.