EasyChair Preprint

№ 1853

Digital Curation with Semantic MediaWiki and Wikidata in numismatic research

Bernhard Krabina

November 5, 2019

Digital Curation with Semantic MediaWiki and Wikidata in numismatic research

Bernhard Krabina[1][0000-0002-6871-3037]

[1] KDZ – Centre for Public Administration Research, Guglgasse 13, 1110 Vienna, Austria
krabina@kdz.or.at

Abstract. The FINA wiki is a Semantic MediaWiki (SMW) supporting numismatic research showing how digital curation can be facilitated. It focuses on curation of non-published material such as letters or manuscript sources scattered over archives and libraries across the world and previously neglected by scholars. A vast number of knowledge visualisations like maps, timelines, charts, word clouds or flow charts provide researchers with better support for handling large amount of content. With a simple mechanism, SMW can query Wikidata and users adding content can decide to use the suggested Wikidata IDs as point of reference. By leveraging Semantic Web standards, SMW instances can ensure long-term viability of valuable digital content.

Keywords: Digital Curation, Semantic MediaWiki, Wikis, Wikidata, Semantic Web, Knowledge Visualisation, Collaboration, Metadata, Controlled Vocabularies, Numismatics.

1 Introduction

1.1 Digital curation

Librarians and archivists collect valuable information, organize it, keep it in usable condition and provide access to users. Though the forms of content have changed throughout the years from handwritten items and published books now to digital content, the purpose of librarianship and archiving has always been to collect and provide effective access to curated information. [4]

Digital curation is defined as “the active management and enhancement of digital information assets for current and future use. [11]

1.2 Wikis

Traditional wikis have the capability to support collaborative authoring of content, mostly by linking wiki pages that follow a document-oriented structure by using titles, headlines, paragraphs, lists, etc. However, they lack the possibility of presenting structures and relationships in a partly automated fashion. Voigt et al. [14] describe semantic, structured and hybrid wikis that enable users to add structured information to the

wiki. Also the wealth of numerical data is only available as plain text and thus cannot be processed by its actual meaning. [8] Bry et al. [2] give an overview and comparison of semantic wikis.

1.3 Semantic MediaWiki and Wikidata for digital curation

Krötzsch et al. [8] introduced Semantic MediaWiki (SMW) in 2006 which has since gone on to establish a vital community[1]. SMW is an extension of MediaWiki, the software used for Wikipedia and many other projects.

SMW is not only a semantic wiki, but with the extension PageForms[2] it can be extended to allow users to implicitly provide semantics by filling in data in particular fields of a form, as described for hybrid wikis. [10] In the case of the FINA wiki (see below), the structure of the wiki is fixed by the wiki configuration and users are not aware of the "semantics" applied in the wiki, they simply enter semantic annotations by filling in data in online forms (see Fig. 3).

Since a precondition for ensuring long-term viability of valuable digital content is to adhere to standards that are supported by the digital curation community, standardisation is key to effective management and future access. [4]

Semantic MediaWiki exposes wiki content to the Semantic Web [1] and allows referencing to other resources via Semantic Web standards. DeRidder [4] advises content creators to use open-source software when creating their materials which allow the content to be exported in open formats for long-term storage in simple CSV format or complex XML with associated schemas. All of these requirements are met by the open-source software SMW by providing various "result formats"[3].

2 FINA Wiki – a Semantic MediaWiki supporting numismatic research

2.1 Unpublished manuscripts and letters as sources of research

A growing and general interest in antiquarianism and historiography, of which numismatics forms an important part, together with the developing awareness that studies on antiquarian numismatics have been based mainly on printed books, which – as numerous as they were – are just the tip of the iceberg, led to the creation of the project Fontes Inediti Numismaticae Antiquae (FINA). Beyond printed books, which fundamentally shaped the development of knowledge, there is a wealth of other evidence, generally ignored even by the few who have bravely attempted to write the history of ancient numismatics. Archives are full of manuscripts that were never published,

1 https://www.semantic-mediawiki.org, last accessed 2019/11/01
2 https://www.mediawiki.org/wiki/Extension:Page_Forms, last accessed 2019/11/01
3 https://www.semantic-mediawiki.org/wiki/Help:Result_formats, last accessed 2019/11/01

especially those too technical or too expensive to be printed. Libraries today still possess many copies of books richly annotated by famous scholars. Most of all, numismatic letters form a fascinating field, enormous in size and rich in colourful stories which imply a much more extensive and profound network of people beyond the few who were knowledgeable or powerful enough to be quoted in books. [3]

The Belgian historian François de Callataÿ [3] has compiled the "Grand Document" as a working document that brings together evidence primarily about numismatic correspondence exchanged before 1800. He started this task at the beginning of 2011, generating an electronic document in Microsoft Word format which has grown over the years to 3,500,000 characters on over 900 pages.

2.2 The FINA wiki

In 2019, the KDZ – Centre for Public Administration Research implemented a FINA wiki for the Austrian Academy of Sciences as a Semantic MediaWiki installation with the aim to convert the "Grand Document" to a wiki [12]. The different numismatic resources addressed in the document are annotated printed books, copies or translations of printed books, catalogues of numismatic collections, manuscripts of unpublished works and – by far the largest section – correspondence (see Fig. 1).

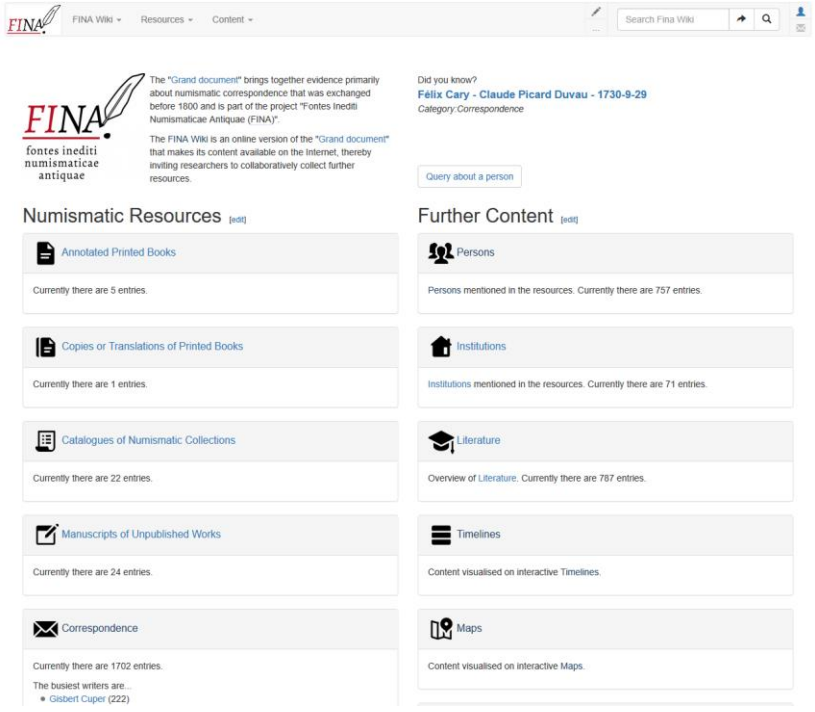

Fig. 1. FINA wiki at https://fina.oeaw.ac.at/wiki, last accessed 2019/11/01

By default, the properties used in Semantic MediaWiki are of the "page" type (meaning that they link to a page within the wiki). This can be changed by simply adding property definition attributes in wikitext, e. g.

```
[[has type::Text]]
```

on the "Property:Author" wiki page would change the property from the "page" type to the "text" type. For standardization purposes the property "Author" in FINA wiki was defined as follows (see Table 1 and Fig. 2):

Table 1. Property definition attributes

has type / imported from	Defines the data type, e. g. page, text, number or, date. In this example, it is of the page type, imported from the FOAF vocabulary foaf:name[4].
has preferred property label	Can be used to define the label for the property for different user language settings.
has property description	A description of the property (in different languages) that is displayed when a link to the property is mouse-overed.

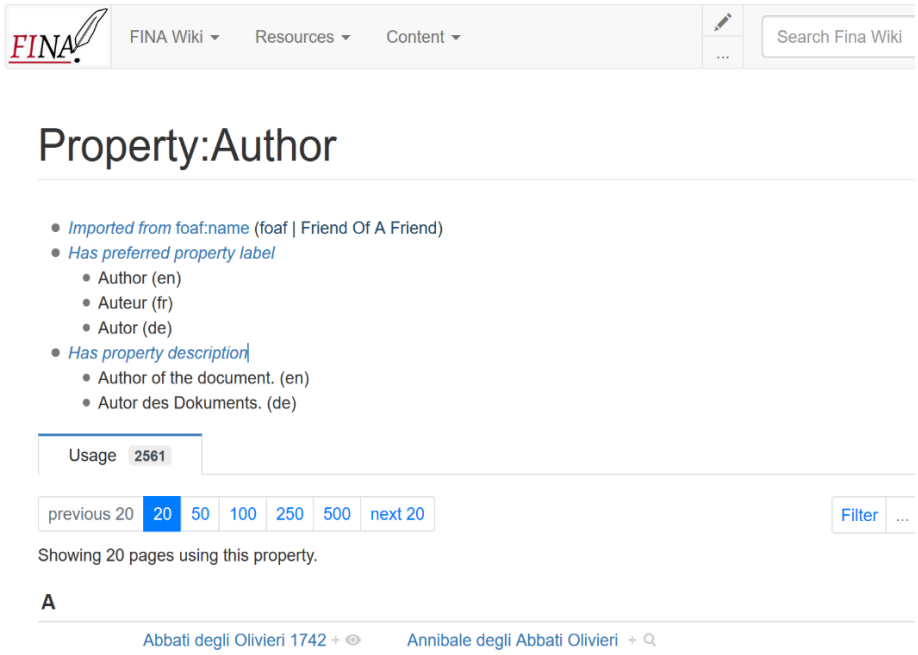

Fig. 2. Flexible property definition in Semantic MediaWiki[5]

[4] http://www.foaf-project.org, last accessed 2019/11/01
[5] https://fina.oeaw.ac.at/wiki/index.php/Property:Author, last accessed 2019/11/01

Further possibilities of property management in SMW are constraint schemas[6] (e. g. allowing only certain values) or data curation[7] (as process and role to cater for the correctness of structured data within a wiki).

2.3 Semantic MediaWiki referencing Wikidata

Wikidata, the free knowledge base that everyone can edit, uses unique IDs to identify items. These item IDs can be used as language-independent identifiers to facilitate data exchange and integration across application boundaries. By referring to Wikidata items, applications can provide unambiguous definitions for the terms they use, which at the same time are the entry point to a wealth of related information. [9]

Using the SMW "external identifier"[8] data type allows the property in FINA wiki to reference an external source. In this case, Wikidata IDs are used within the wiki and link to Wikidata. For example, the page about the Austrian numismatist "Joseph Eckhel" has the property WikidataID "Q78749". Since the external identifier data type is used, the display of the value links to the URI http://www.wikidata.org/entity/Q78749, which is the entry for Joseph Eckhel in Wikidata.

Because the FINA wiki uses form-based data entry as default, it is possible to query Wikidata for the name of a person at the time a user wants to create (or edit) the page. If an entry is found, the description of the person is displayed, and the Q-number is suggested in the entry field. The person about to enter (or edit) the person page can check in Wikidata whether the match is correct and either accept the suggested ID, enter a different ID or no ID (see Fig. 3).

---

6 https://www.semantic-mediawiki.org/wiki/Help:Constraint_schema, last accessed 2019/11/01

7 https://www.semantic-mediawiki.org/wiki/Data_curator, last accessed 2019/11/01

8 https://www.semantic-mediawiki.org/wiki/Help:Type_External_identifier, last accessed 2019/11/01

Fig. 3. Wikidata query when entering persons in FINA wiki[9]

Technically, the Wikidata lookup is done using the External Data extension[10] querying the Wikidata API[11].

2.4 Exporting data from SMW to Wikidata or other systems

Semantic MediaWiki offers various export formats ranging from simple CSV files or the popular JSON format to the sophisticated RDF[12]. As the license of the FINA wiki is CC0[13], it is both legally and technically easy to add facts from the FINA wiki to Wikidata (or any other system). Here are some parts of the RDF output as an example:

```
<skos:exactMatch rdf:resource="http://www.wikidata.org/entity/Q78749"/>
<dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Joseph
Eckhel</dc:title>
<property:WikidataID
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Q78749</prop-
erty:WikidataID>
```

[9] https://fina.oeaw.ac.at/wiki/index.php?title=Angelo_Poliziano&action=formedit, last accessed 2019/11/01
[10] https://www.mediawiki.org/wiki/Extension:External_Data, last accessed 2019/11/01
[11] https://www.wikidata.org/w/api.php, last accessed 2019/11/01
[12] https://www.semantic-mediawiki.org/wiki/Help:RDF_format, last accessed 2019/11/01
[13] https://creativecommons.org/share-your-work/public-domain/cc0/, last accessed 2019/11/01

2.5 Visualisation

Knowledge visualisations help to compress large amounts of information, absorb complexity and render it accessible. This can be a vital prerequisite for transfer, creation and communication of knowledge. [5]

Visualisation of the knowledge managed within FINA wiki plays an important role. There are several geographic maps, charts, word clouds and timelines visualising the content of the wiki. An example is displaying the correspondence with a flowchart using the mermaid scripting language[14] (see Fig. 4).

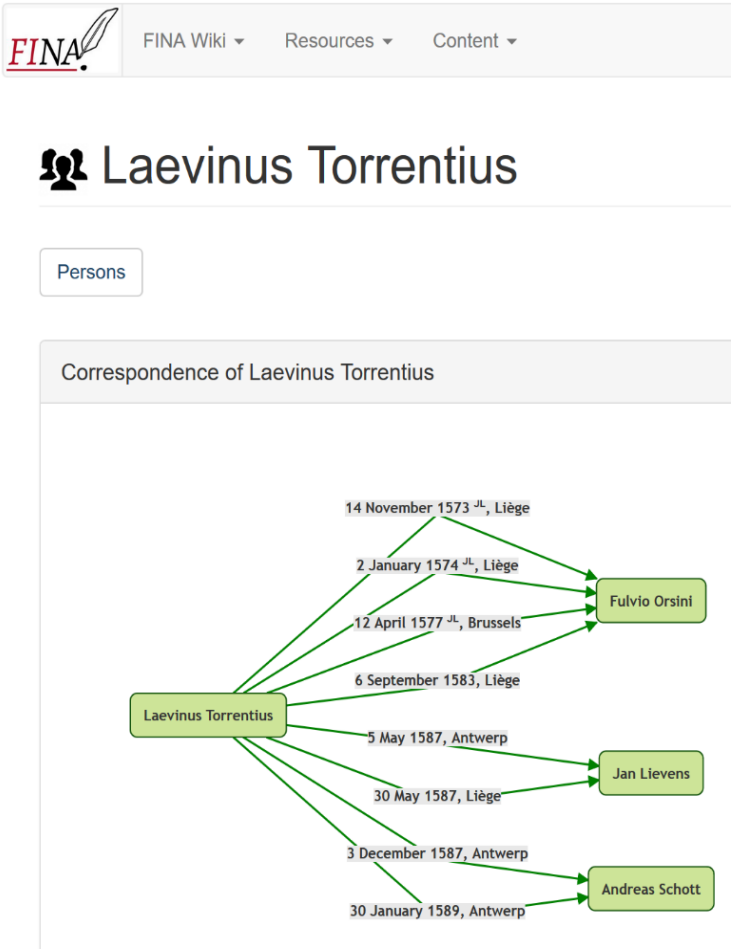

Fig. 4. Flowchart visualisation of correspondence[15]

[14] http://knsv.github.io/mermaid/, last accessed 2019-11-01
[15] https://fina.oeaw.ac.at/wiki/index.php/Laevinus_Torrentius, last accessed 2019-11-01

3 Conclusion

The concept of Linked Data has made its entrance to the cultural heritage sector due to its potential use for the integration of heterogeneous collections and deriving additional value out of existing metadata. [6]

As has been shown with the "Vienna History Wiki", a geo-referenced, historical knowledge platform of the city of Vienna aiming to combine knowledge from the city administration with those of external experts [7], Semantic MediaWiki is an excellent choice for digital curation. Referencing Wikidata (or other external identifiers like the Integrated Authority File GND[16]) is especially beneficial in special interest wikis in historical contexts.

The method of querying Wikidata items in FINA wiki does not (yet) use the more sophisticated approach of "reconciliation" via the OpenRefine reconciliation API [13], but it is a first and simple method of referencing Wikidata knowledge and therefore supporting curation.

Efficient mechanisms to more tightly integrate digitally curated content with controlled vocabularies without the need for external tools like OpenRefine will be a research and development agenda for the coming years.

References

1. Berners-Lee, T., Hendler, J. and Lassila, O. The Semantic Web. Scientific American, pages 96–101, May 2001.
2. Bry, F., Schaffert, S., Vrandečić, D. and Weiand, K. 2012. Semantic Wikis: Approaches, Applications, and Perspectives. In: Reasoning Web: Semantic Technologies for Advanced Query Answering. Lecture Notes in Computer Science, Volume 7487 (2012). DOI= 10.1007/978-3-642-33158-9_9
3. Callataÿ, F.: Fontes Inediti Numismaticae Antiquae (FINA) a short presentation, In M. Caccamo Caltabiano et al. (eds.), XVth International Numismatic Congress. Taormina. Proceedings, I, Messina, 2017, pp. 95-99. (2017).
4. DeRider, J. L.: Digital Curation Fundamentals. Rowman & Littlefield, Maryland (2018).
5. Eppler, M. J. & Burkhard, R. A.: Knowledge Visualization. Towards a New Discipline and its Fields of Application (2004).
6. Hooland, S., Verborgh, R., De Wilde, M., Hercher, J., Mannens, E. and Van de Walle, R.: Evaluating the success of vocabulary reconciliation for cultural heritage collections. J Am Soc Inf Sci Tec, 64: 464-479 (2013).
7. Krabina, B. The Vienna History Wiki - a Collaborative Knowledge Platform for the City of Vienna. Proceedings of the 11th International Symposium on Open Collaboration (OpenSym 2015). ACM, 2015. Available from: http://www.opensym.org/os2015/proceedings-files/p500-krabina.pdf, last accessed 2019/11/01

[16] https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd.html, last accessed 2019/11/01

8. Krötzsch, M., Vrandecic, D., Völkel, M., Haller, H., Studer, R. 2006. Semantic Wikipedia. Proceedings of the 15th international conference on World Wide Web, WWW 2006. Available from: http://www.aifb.kit.edu/web/Inproceedings1055, last accessed 2019/11/01
9. Krötzsch, M. & Vrandecic, D.: Wikidata: A Free Collaborative Knowledge Base. Communications of the ACM vol. 57, pp 78-85 (2014). Available from: http://cacm.acm.org/magazines/2014/10/178785-wikidata/fulltext, last accessed 2019/11/01
10. Matthes F.; Neubert C.; Steinhoff A. 2011. Hybrid Wikis: Empowering Users to Collaboratively Structure Information. In: 6th International Conference on Software and Data Technologies (ICSOFT), Seville (2011).
11. National Research Council, Policy and Global Affairs, Board on Research Data and Information, Committee on Future Career Opportunities and Educational Requirements for Digital Curation. Preparing the Workforce for Digital Curation. Washington (DC): National Academies Press (US); (2015). Available from: https://www.ncbi.nlm.nih.gov/books/NBK293661/, last accessed 2019/11/01
12. OeAW (Austrian Academy of Sciences): Fontes Inediti Numismaticae Antiquae (FINA) https://www.oeaw.ac.at/en/ancient/research/documenta-antiqua/numismatik/fontes-inediti-numismaticae-antiquae-fina/, last accessed 2019/11/01
13. OpenRefine: Reconciliation Service API (2019) https://github.com/OpenRefine/OpenRefine/wiki/Reconciliation-Service-API, last accessed 2019/11/01
14. Voigt, S., Fuchs-Kittowski, F., Gohr, A. 2014. Structured Wikis – Application-Oriented Use Cases. Proceedings of the 10th International Symposium on Open Collaboration (OpenSym 2014). ACM (2014). DOI=2641580.2641610