

■ ENRICHING METADATA FOR A UNIVERSITY REPOSITORY BY MODELLING AND INFRASTRUCTURE: A NEW VOCABULARY SERVER FOR PHAIDRA

by Anna Bellotto and Janos Bekesi

Abstract: *This paper illustrates an initial step towards the ‘semantic enrichment’ of University of Vienna’s Phaidra repository as one of the valuable and up-to-date strategies able to enhance its role and usage. Firstly, a technical report points out the choice made in a local context, i.e. the deployment of the vocabulary server iQvoc instead of the formerly used SKOSMOS, explaining design decisions behind the current tool and additional features that the implementation required. Afterwards, some modelling characteristics of the local LOD controlled vocabulary are described according to SKOS documentation and best practices, highlighting which approaches can be pursued for rendering a LOD KOS available in the Web as well as issues that can be possibly encountered.*

Keywords: KOS; LOD KOS; iQvoc; semantic enrichment; cultural heritage object metadata

DAS ANREICHERN VON METADATEN FÜR EIN UNIVERSITÄRES REPOSITORIUM DURCH MODELLIERUNG UND INFRASTRUKTUR: EIN NEUER VOCABULARYSERVER FÜR PHAIDRA

Zusammenfassung: *Dieser Beitrag erörtert einen der ersten Schritte im Zusammenhang mit der „semantischen Anreicherung“ des Phaidra-Repositoriums der Universität Wien. Zunächst wird in einem technischen Report auf die in einem lokalen Kontext getroffene Auswahl hingewiesen, d.h. auf die Bereitstellung des Vocabularyservers iQvoc anstelle des zuvor verwendeten SKOSMOS, und erläutert die für die Implementierung erforderlichen Entscheidungen hinsichtlich des Designs des aktuellen Tools sowie zusätzliche Funktionen. Anschließend werden einige Modellierungsmerkmale des lokalen LOD-gesteuerten Vokabulars gemäß der SKOS-Dokumentation und den -Best Practices beschrieben, wobei aufgezeigt wird, welche Ansätze zur Bereitstellung eines LOD-KOS im Web verfolgt werden und welche Probleme dabei möglicherweise auftreten können.*

Schlagwörter: KOS; LOD KOS; iQvoc; semantische Anreicherung; Metadaten von Cultural Heritage-Objekten

DOI: <https://doi.org/10.31263/voebm.v72i2.3206>



Contents

1. *Introduction*
2. *Implementing the vocabulary manager iQvoc*
3. *Features and benefits of a LOD KOS vocabulary*
4. *Modelling the LOD controlled vocabulary: preliminary key choices*
5. *Editing the LOD controlled vocabulary: contextualisation in practice*
6. *Future work and conclusions*

1. Introduction

As stated by Zeng (2019, p.7), enriching metadata “has become a common initiative in LAM [Libraries, Archives, and Museums] data enhancement efforts, in order to overcome challenges relating to data quality and discoverability in the digital age, while providing more context and multilingual information for cultural heritage (CH) objects”. This overall outcome, commonly known as ‘semantic enrichment’, is one of the strategies that would enable LAM data to turn into bigger and smarter data, i.e. structured (or semi-structured), highly integrated and much more meaningful data which would support researchers and general users in widely exploring and reusing them (Zeng, 2019, p.30).

From a technical viewpoint, the set of established standards and semantic technologies collectively referred to as ‘Linked Data’ forms the needed environment to put into practice the process of semantic enrichment, whereas the key approaches for its application can be multiple and varied. Metadata records in a digital repository often contain multiple fields with a (semi-)closed/controlled set of values, which are either represented as a string-value or sourced from a local set of values. This is suited for enrichment, i.e. a procedure which consists in providing metadata of “more contextualized meanings” by expressing various types of relationships (Zeng, 2019, p.7). If pursued, this roadmap would then make LAM institutions achieve a broader conceptual shift from document-centric metadata to RDF-based data-centric metadata, shifting from primarily human-oriented consumption to rather machine processability, shareability and mashability (Alemu, et al., 2012).

Phaidra is the repository for the long-term preservation and archiving of digital resources of the University of Vienna. The ongoing migration

of Phaidra's metadata to RDF is another step in the University Library's strategy to uncover the semantic potential of its data. Phaidra currently has two major development goals: the first one is to decouple the service core from all frontend interfaces by enforcing the use of its API (Application Programming Interface); this allows connecting all sorts of different frontends to the service core, independent of programming or scripting language or of layout and design considerations constricting some frontend solutions. The second goal is the increased usage of RDF which will be facilitated by a careful upgrade and migration of Phaidra's core component Fedora from current version 3.8 to the latest one. This requires thorough reworking of central object model definitions and therefore cannot be undergone lightly, but with considerable conceptualization effort. Since Phaidra's operation may not be interrupted for the sake of its users, and furthermore those changes and transformations should be valid and viable for future years, the migration to a new, updated core framework has to be carried out with extreme caution and circumspection.

The opportunity of interlinking with a global network of loosely standardized data (as is realized in RDF) will open a new perspective not only for enriching metadata of objects stored in Phaidra, but also by connecting its content more tightly and seamlessly to other LOD-enabled repositories all over the world, like e.g. Europeana or the British Museums Collection, and not to forget to other Phaidra instances currently in service in neighbouring provinces and countries. A first small, somewhat local step towards this goal is the introduction of a vocabulary server within the University of Vienna computing network. This server not only provides certain metadata information in a proper RDF formatted way, i.e. as consistent, durable and trustworthy endpoint for the respective data, but also operates as a source for some selection choices in the editing process of Phaidra objects' metadata, and moreover offers an easy to use workflow for the editing of the concepts it stores.

2. Implementing the vocabulary manager iQvoc

The implementation of the vocabulary server, based on the software iQvoc was preceded by the evaluation, deployment and woefully short operation of SKOSMOS¹, another vocabulary server based on PHP and MySQL. During its running time some shortcomings were detected which led to the termination of its service: though powerful and reliable in its day-to-day operation, it was soon found out that modifying and inserting

data was too cumbersome, as it had to be done by editing the exported data in a different ontology editor (like e.g. Protégé), and by manually importing it into the server's data storage afterwards. Eventually its use was brought to an end after a few such editing cycles; the effort did not justify the benefits.

Subsequently, however, the University of Vienna Computer Center evaluated, adapted and deployed a solution which is a better fit to its requirements: the vocabulary server iQvoc², developed by the German company Innoq, initially commissioned by the German Federal Environment Agency to provide an “Open Source SKOS maintenance and publishing tool”. It is licensed with EUPL, the European Union Public License³, a GPL-compatible open source license adapted for use in the European Union; the underlying technologies, Ruby on Rails and jQuery, ensure quick adaptability, since both are fairly wide-spread and not too difficult to master.

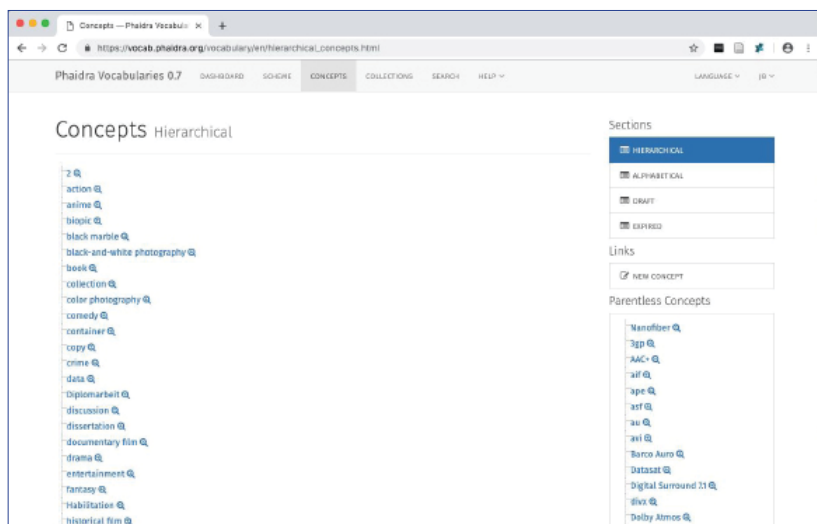


Fig. 1: The list of concepts in the administrator's view

One live example of iQvoc that can be easily inspected is UMTHESES, the thesaurus of the German Federal Environment Agency, containing about 14.000 concepts and 66.000 synonyms and English translations⁴. The scope of used terms and its huge number certainly exceeds Phaidras projected uses, but it can illustrate the range iQvoc is able to cope with. Core features of iQvoc (as per requirements of its initial employer) are:

SKOS(XL)-compliance; web interface for navigation and browsing; multilingual capabilities; editing with validation; editorial team and workflow support; as well as Linked Data support.

Here it might be appropriate to highlight some design decisions concerning the support of Linked Data: in accordance with the requirement for sustainable references (“Cool URIs don’t change”), deprecated concepts stored in the vocabulary are only expired, but not deleted, so any concept created stays there to be requested as long as the service is running. Secondly, Linked Data support is realized by rendering content for concept endpoints following the “303 URI forwarding specification”, thereby providing RDF-notations as requested. Additionally the vocabulary server provides a synchronization feature for submitting all of its data to a triple store proxy (Sesame or Virtuoso out-of-the-box), thus supplementing some missing features like JSON-LD content delivery or a SPARQL endpoint.

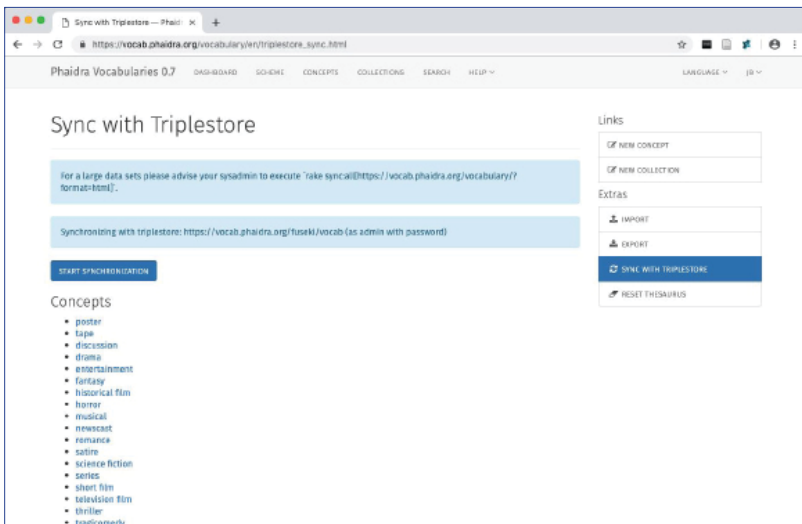


Fig. 2: Triple store synchronization page

After evaluation of iQvoc as promising solution for a vocabulary server, the University of Vienna Computer Center made some minor modifications and amendments to the source code, in particular a more suitable way for concepts to be represented as URIs as well as the addition of the Jena/Fuseki triple store engine as a synchronization target. The former consists in the replacement of the last URI path element (underscore followed by a

random number) with a “base32” notation, which has a good balance between compactness, error resistance and readability/citability by humans.⁵ The latter had to be implemented due to the fact that Jena/Fuseki is used as triple store in the University’s network. While doing so, the capability to deal with secured connections (https://) to synchronization target was also added.

Finally, some installation quirks were tackled during deployment of the Ruby on Rails framework as an Apache Passenger application. All those modifications and improvements will be available as Open Source as soon as the last minor flaws are removed.⁶

3. Features and benefits of a LOD KOS vocabulary

Alongside the conventional umbrella term ‘KOS’ widely used to refer to the existing types of Knowledge Organization Systems which accurately and consistently “organize information and provide terminology to catalog and retrieve information” (Harpring, 2010, p.12), as thesauri, controlled vocabularies and classification schemes, the label ‘LOD KOS’ is currently used to designate those same KOS in a Semantic Web framework (Zeng and Mayr, 2018). As such, a LOD KOS vocabulary follows the principles of Linked Open Data: it uses unique HTTP URIs for distinctively denoting its entities; it expresses its data in an RDF syntax, such as JSON-LD, and it models them according to an established standard, such as SKOS; it allows its data to be accessed through a SPARQL endpoint; and finally, it enriches its data with inbound and outbound links to concepts within and outside the vocabulary (Zeng and Mayr, 2018).

Looking at their usage, LOD KOS do not just offer great potential to “open the doors of the silos” (Bizer, et al., 2008), as not machine-readable datasets are. Rather, being “primary sources which enable datasets to become 4-star and 5-star Linked Open Data”⁷, they can be seen as “invaluable engines” (Zeng and Mayr, 2018). When concepts from LOD vocabularies populate the allowed values for an element in RDF-based metadata records, metadata descriptions become connected with heterogeneous sources, facilitating datasets to be visible and accessible in a more enriched way. As a result, the resources can be cited more broadly. Additionally, by implementing LOD KOS vocabularies, metadata sharing and reuse is also augmented producing a decentralized and more efficient workflow. Indeed, metadata providers can reuse existing data already formed by others or collaborate with other LAM institutions, while concen-

trating their effort on creating descriptions of their local expertise (Alemu, et al., 2012, p.10; Open Metadata Handbook, 2012). AGROVOC⁸, the LOD Thesaurus functioning as the backbone of the bibliographic database for agricultural science named AGRIS (Subirat and Zeng, 2014), can be suggested as a representative example.

The use of the Simple Knowledge Organization System (SKOS) data model, a standard recommended by the W3C community to represent the Knowledge Organization Systems in Semantic Web applications, pivotally contributes to make all of these benefits happen. Expressing structural and content features commonly shared by controlled vocabularies and other KOS types, the aim of SKOS is to turn these stand-alone entities of organized information into a global machine-readable network of highly integrated conceptual schemes (W3C, 2009a), publishable in the Web, readable and automatically discoverable by applications (W3C, 2009a; W3C, 2009b; Bellotto and Bettella, 2019).

4. Modelling the LOD controlled vocabulary: preliminary key choices

When constructing a vocabulary, technical decisions about the desired structure, the construction methods, as well as the vocabulary relationships with the repository data model, should first be established. According to this, one of the first issues that needed to be considered was “which data needs to have controlled terminology” (Harpring, 2010, p.136), i.e. to make a distinction between which fields of the Phaidra data model should contain data values drawn from controlled terms (controlled fields) and which fields should be left without controlled sources (free-text fields). At the first stage of the vocabulary implementation, this choice was made in order to enhance the normalization and validation of metadata properties as much as possible. By focusing on this target, fields relating to the type of resource being catalogued, the material of which it is composed, or its genre, for instance, were recognized as metadata elements demanding a standardized format.

Once this preliminary decision (constantly open, considering the possibility of RDF graphs to be extended with new nodes and new relationship types effortlessly) was made, the following step consisted in choosing how to logically and consistently divide the terminology forming the local controlled vocabulary. The required efforts to put into this stage were extensive. Not only it is at this level that, according to Zeng and Mayr (2018), it is possible to identify the identity of KOS vocabularies as “semantic road

maps”, modelling “the underlying semantic structures of domains” rather than being merely “sources of values” in metadata records. Additionally, this challenge was further augmented by the fact that no specific guidelines about how to structure a SKOS-formatted vocabulary exist (this being one of the advantages of SKOS).

When arranging a large vocabulary, several levels exist for structuring its hierarchy: ‘facets’, ‘subfacets’ and ‘node labels’ are the technical terms which label them. Facets can be considered the major divisions of a controlled vocabulary, directly descending from the highest level of the hierarchical structure (‘root’) and grouping together concepts that share similar characteristics. Each facet may then have additional subdivisions, called ‘subfacets’. Finally, node labels or guide terms, which are usually represented by angled brackets, provide a further possibility to logically distinguish groups of sibling concepts sharing a common parent concept (Harpring, 2010, pp.142–144). Overall, the existence of these concept groups in structured KOS is most often meant to support the cataloguing of resources and provide useful features for navigating a conceptual network (Baker, et al., 2016, p.16). The illustration of the Art & Architecture Thesaurus (AAT) Object Facets and its following hierarchical levels (see Figure 3) can visually help to understand this modelling.



Fig. 3: Partial display of ‘photographs’ in the AAT Objects Facet (The J. Paul Getty Trust, 2004)

Although the SKOS format was conceived to support the migration of different types of KOS in RDF language while reflecting best thesaurus construction principles, an exact compatibility with ISO 25964, the latest standard on thesauri, is missing, especially with respect to concept groups. And no clear guidance about how to encode non-flat vocabularies has been officially supplied, forcing data modelers to opt for ad-hoc solutions (Baker, et al., 2016, p.16). `<skos:ConceptScheme>` and `<skos:Collection>` are the only two elements defined by SKOS standard to express grouping of concepts. `<skos:ConceptScheme>` represents “the notion of an individual thesaurus, classification scheme, subject heading system or other knowledge organization system” (W3C, 2009b), while `<skos:Collection>` refers to group of concepts that “share something in common”, conveniently grouped “under a common label” (W3C, 2009b). SKOS documentation explicitly highlights that the modelling of node labels as instances of `<skos:Collection>` is the best practice towards “semantic accuracy”, instead of expressing them as mere concepts (W3C, 2009a). However, it does not point out whether (and when) microthesauri⁹ and narrower subdivisions that structured KOS usually feature should be preferably represented either as `<skos:ConceptScheme>` or `<skos:Collection>`.

In the context of Phaidra Vocabulary Server, display-related considerations motivated the need of distinguishing homogeneous classes of concepts, referring, for example, to categories of objects (Object type) or movie genres (Genre by motion pictures). Facing the issue outlined above, these subsets of concepts were encoded as `<skos:Collection>` according to the proposed correspondence between the ISO 25964 semantics and SKOS data model (NISO, 2013). Indeed, in this document both thesaurus arrays (`<iso-thes:ThesaurusArray>`) and concept groups (`<iso-thes:ConceptGroup>`) are formally defined as subclasses of `<skos:Collection>`. Nevertheless, as argued by Alexiev and Cobb (2017), due to additional constructs not covered by the SKOS standard, some limitations still occur, including the fact that “you can’t say explicitly which are Top Collections in a scheme”. Albeit in the Getty Vocabulary Program (GVP) ontology this constraint was ontologically coped with defining an additional class (Alexiev and Cobb, 2017), in our case study we assessed the implementation of nested collections as the most straightforward solution.

5. Editing the LOD controlled vocabulary: contextualisation in practice

Along with technical decisions, “appropriate editorial rules for building the vocabulary” require to be identified and adopted in order to assure consistent, accurate and trustworthy data (Harpring, 2010, p.138). However, in the frame of a LOD KOS, these rules must involve additional approaches towards an enhanced integration and interoperability: the establishment of mapping relationships with external controlled vocabularies.

Central to the SKOS data model is the notion of a ‘concept’ (<skos:Concept>): it is an abstract unit of thought, i.e. an idea, a meaning, a class of objects or events, uniquely identified by an URI and independent from the ‘terms’ (<skos:prefLabel>, <skos:altLabel>), i.e. multilingual expressions used to label that concept in natural language (W3C, 2009a; W3C, 2009b). This emphasis of SKOS data model on semantics rather than on terminology does not only disclose great benefits among conceptual schemes facing comprehension issues in a multilingual framework. It also enables concepts coming from different contexts and possibly following dissimilar modelling principles to be connected, compared and matched according to their meaning only.

In the Phaidra thesaurus, once the newly created concepts were assigned preferred lexical labels, the alignment of most of them with external controlled vocabularies was pursued through relationships of equivalence or similarity (<skos:exactMatch> and <skos:closeMatch> respectively) in order to represent their underlying semantics explicitly in a machine-processable manner. This semantic enrichment was applied as a result of a specific internal workflow. Authoritative and well-established reference resources, as for instance the Art & Architecture Thesaurus (AAT) by The Getty Research Institute and the Controlled Vocabulary for Resource Type Genres by COAR, were thoroughly compared observing the definitions or scope notes they provided for the matching concepts. The external concepts which then best fitted the intended meaning of the internal concepts were targetted for the linking. Nevertheless, with the intent of considering an amount of matchings as extensive as possible while assuring at the same time the high quality of these links (Bizer, et al., 2008), the whole task is still in progress and pending for further review.

The implications of such a method, recommended in a LOD framework, would not simply embrace the concepts of the thesaurus: despite the fact that the migration of Phaidra platform to a Linked Data environment is still under way, more broadly the repository metadata records would be potentially connected to a vast array of external datasets in the Web.

6. Future work and conclusions

The limited extent of this case study has been pointed out. On the one hand, it is limited on account of the Phaidra Vocabulary Server being a living, growing tool, available to be extended by users outside the reviewing editorial team. On the other, by reason of the Phaidra RDF data model being a new foundational data profile, flexible and extensible to any future needs (Bellotto and Bettella, 2019). Consequently, different structural decisions may take place in the near future. Additionally, other tasks will require forthcoming considerations. We aim to develop a plan for the maintenance and workflow of the Vocabulary Server, taking into account the different types of its future users, the vocabulary editors, regulating the addition of new concepts, the editing of their terms, relationships and notes, in order to assure data consistency and the trustworthiness of the resource. Concurrently, a supporting documentation of the tool regarding how the thesaurus has been and should be structured, what kinds of relationships should be included, and which information is assessed as particularly relevant to ensure a correct interpretation of the concepts (e.g. <skos:definition>), may be a valuable aid for training the new editors.

Regardless of the aforementioned aspects that still need to be addressed, the paper has outlined initial steps towards the ‘semantic enrichment’ of Phaidra repository as one of the valuable and up-to-date strategies able to enhance its role and usage. A first technical report has pointed out the choice made in a local context, i.e. the deployment of the vocabulary server iQvoc instead of the formerly used SKOSMOS, explaining design decisions behind the current tool and additional features that the implementation required. Afterwards, some modelling characteristics of the local LOD controlled vocabulary have been described according to SKOS documentation and best common practices, highlighting which approaches can be pursued for rendering a LOD KOS available in the Web as well as issues that can be possibly encountered.

Anna Bellotto, MA

ORCID: <https://orcid.org/0000-0003-1148-5456>

University of Vienna, Vienna University Computer Center

E-Mail: anna.bellotto@univie.ac.at

Dr. Janos Bekesi

University of Vienna, Vienna University Computer Center

E-Mail: janos.bekesi@univie.ac.at

References

- Alemu, G., Stevens, B., Ross, P., and Chandler, J. (2012). Linked data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models. [online pdf] Available at: <http://conference.ifla.org/sites/default/files/files/papers/wlic2012/92-alemu-en.pdf> [Accessed 29 April 2019].
- Alexiev, V., and Cobb, J. (2017). Getty Vocabularies: Linked Open Data – Semantic Representation. Version 3.4. [HTML version] Available at: <http://vocab.getty.edu/doc> [Accessed 29 April 2019].
- Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., and Summers, E. (2016). Key Choices in the Design of Simple Knowledge Organization System (SKOS). [online pdf] Available at: <https://arxiv.org/pdf/1302.1224.pdf> [Accessed 29 April 2019].
- Bellotto, A., and Bettella, C. (2019). Metadata as Semantic Palimpsests: The Case of PHAIDRA@unipd. Manghi P., Candela L., Silvello G. (2019). Digital Libraries: Supporting Open Science. IRCDL 2019. Communications in Computer and Information Science 988, 167–184. https://doi.org/10.1007/978-3-030-11226-4_14
- Berners-Lee, T. (2006). Linked Data. [web page] Available at: <https://www.w3.org/DesignIssues/LinkedData.html> [Accessed 29 April 2019].
- Bizer, C., Heath, T., and Berners-Lee, T. (2008). Linked Data: Principles and State of the Art. [online slides] Available at: <https://www.w3.org/2008/Talks/WWW2008-W3CTrack-LOD.pdf> [Accessed 29 April 2019].
- Harpring, P. (2010). Introduction to Controlled Vocabularies. Terminology for Art, Architecture, and Other Cultural Works. [online] Available at: https://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/ [Accessed 29 April 2019].
- NISO (2013). Correspondence between ISO 25964 and SKOS/SKOS XL Models. [online pdf] Available at: https://groups.niso.org/apps/group_public/download.php/12351/Correspondence%20ISO25964-SKOSXL-MADS-2013-12-11.pdf [Accessed 29 April 2019].
- Open Metadata Handbook/Recommendations (2012). [wikibook] Available at: https://en.wikibooks.org/wiki/Open_Metadata_Handbook/Recommendations [Accessed 29 April 2019].
- Subirat, I., and Zeng, M. L. (2014). Using KOS as the Connectors of Linked Datasets. The 77th Annual Meeting of the Association for Information Science and Technology (Seattle, WA, USA). [online slides] Available at: <http://nkos.slis.kent.edu/ASIST2014/ASIST-AGROVOC-AAT-SubiratsZeng.pdf> [Accessed 29 April 2019].

- The J. Paul Getty Trust (2004). Art & Architecture Thesaurus Online Hierarchy Display. [web page] Available at: <http://www.getty.edu/vow/AATHierarchy?find=photograph&logic=AND¬e=&subjectid=300046300> [Accessed 29 April 2019].
- World Wide Web Consortium (W3C) (2009a). SKOS Simple Knowledge Organization System Primer. Working Group Note. Available at: <https://www.w3.org/TR/skos-primer/> [Accessed 29 April 2019].
- World Wide Web Consortium (W3C) (2009b). SKOS Simple Knowledge Organization System Reference. W3C Recommendation. Available at: <https://www.w3.org/TR/skos-reference/> [Accessed 29 April 2019].
- Zeng, M. L. (2017). Create Microthesauri and other datasets from the Getty LOD Vocabularies. Museums and the Web 2017 Conference (Cleveland, Ohio, April 2017). [online slides] Available at: http://getty.edu/research/tools/vocabularies/zeng_microthesauri_getty_lod.pdf [Accessed 29 April 2019].
- Zeng, M. L. (2019). Semantic enrichment for enhancing LAM data and supporting digital humanities. [online] Available at: <https://recyt.fecyt.es/index.php/EPI/article/view/epi.2019.ene.03/42134> [Accessed 29 April 2019].
- Zeng, M. L., and Mayr, P. (2018). Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review. *International Journal of Digital Libraries* 20(3), 1–22. <https://doi.org/10.1007/s00799-018-0241-2>
- 1 For the analysis of an initial attempt of SKOSMOS in Phaidra, see Belotto and Bettella (2019).
 - 2 For detailed information on iQvoc, see the documentation in <https://github.com/innoq/iqvoc/wiki>.
 - 3 https://joinup.ec.europa.eu/community/eupl/og_page/eupl
 - 4 <https://sns.uba.de/umthes/de.html>
 - 5 For example, “_e037ef2e.html” is now replaced with “7E4S-MA30.html”.
 - 6 Adventurous users can give it a try by cloning the repository from <https://gitlab.phaidra.org/bekesij9/iqvoc>.
 - 7 In 2006, Tim Berners-Lee developed a star rating system for encouraging and guiding people to create good Linked Open Data. Specifically, data would obtain 4-star and 5-star if, additionally to be released under an open license and machine-readably structured in a non-proprietary format, things were identified using open W3C standards (RDF and SPARQL), and linked to external data in order to provide context (Berners-Lee, 2006).

- 8 <http://aims.fao.org/standards/agrovoc/functionalities/search>.
- 9 “A designated subset of a thesaurus that is capable of functioning as a complete thesaurus”: definition of the ISO Standard 25964-2:2013 provided in Subirat and Zeng (2014) and Zeng (2017).