Uniform description and access to Knowledge Organization Systems with BARTOC and JSKOS

Jakob Voß*, Andreas Ledl**,
Uma Balakrishnan***

*Verbundzentrale des GBV (VZG)
jakob.voss@gbv.de / uma.balakrishnan@gbv.de
http://www.gbv.de/

**UB Basel
andreas.ledl@unibas.ch
http://www.ub.unibas.ch/

Abstract. The Basel Register of Thesauri, Ontologies & Classifications (BARTOC) provides information about a large number of Knowledge Organization Systems (KOS) such as classifications, thesauri, authority files etc. To further improve availability and usefulness of both the description and content of KOS, they are mapped to the uniform JSKOS data format being developed in project coli-conc. Specification of a corresponding JSKOS-API will allow users to directly browse and search in KOS from any place.

Résumé. Le Basel Register of Thesauri, Ontologies & Classifications (BARTOC) renseigne sur un grand nombre de systèmes d'organisation des connaissances (SOC) tels que des classifications, des thésaurus et des fichiers d'autorité. Afin d'optimiser l'accès aux informations contenues dans le registre et leur potentiel, la description et le contenu des SOC sont mappés en format uniforme de données JSKOS, développé dans le cadre du projet Coli-conc. Les spécifications d'une interface de programmation d'application JSKOS qui permettra en outre aux utilisateurs de naviguer et de chercher directement dans des SOC du monde entier.

1. Introduction

The number of Knowledge Organization Systems (KOS) which "encompass all types of schemes for organizing information and promoting knowledge management" (Hodge, 2000) is growing rapidly. Terminology registries help to identify, describe and make accessible these KOS, ideally in a human- and machine-readable way. The Basel Register of Thesauri, Ontologies & Classifications (BARTOC)[1] has quickly evolved to one of the largest of these terminology registries and fulfills both requirements. Access to information about KOS is made possible by a user interface with multiple search and access modes and by methods to download and retrieve KOS description in a data format named JSKOS. This format has been developed in

[1] http://bartoc.org/

project coli-conc[2] to provide uniform description and a common API to query content of KOS for terminology-based web applications.

2. The BARTOC registry

2.1 Overview

Since its launch in November 2013 BARTOC has grown to a collection of more than 1.900 KOS descriptions. There are 14 types of KOS being collected, as defined by NKOS: categorization schemes, classification schemes, dictionaries, gazetteers, glossaries, lists, name authority lists, ontologies, semantic networks, subject heading schemes, synonym rings, taxonomies, terminologies and thesauri. In addition BARTOC refers to 72 other portals, making it also a meta registry of terminology registries.[3] The registry is based at the Basel University Library, Switzerland, and addresses both the library & information science community, but also terminologists, taxonomists and scientists from all over the world. The scope of BARTOC is very broad as it includes any kind of KOS from any subject area in any language, any publication format, and any form of accessibility. This wide approach stands out against the majority of registries and was one of the main causes that led to the foundation of BARTOC. Among the four types of KOS registries specified by Golub et al. (2014, p.1903), BARTOC is best fitting into category basic terminology registry, meaning that it contains "only" metadata of Knowledge Organization Systems. But it is also striving to become a full terminology registry with the possibility of searching for single terms or concepts, turning it into a Swiss army knife of KOS.

But there are more unique selling points: BARTOC is the only terminology registry approved by the International Society for Knowledge Organization (ISKO); and while many of the older registries are no longer maintained, BARTOC tries to keep its content current by community building. The registry quickly reached a point where it became too big for being maintained by just one person. Consequently, a circle of supporters gathered around it so that editorial offices are currently existing in France, Germany, Greece, Italy, Norway, Portugal, Spain, and the UK. This has strengthened the crowd-sourcing aspect and will help to better keep the content up-to-date. Further important reasons to start BARTOC - shared with other platforms - were to "make traditional resources more visible", to "provide key characteristics of resources", to "encourage human assessment of these resources by applicability to semantic projects", and to "promote information exchange and knowledge sharing" (Hlava, 2011, p. 20).

[2] https://coli-conc.gbv.de/

[3] http://bartoc.org/en/terminology-registries

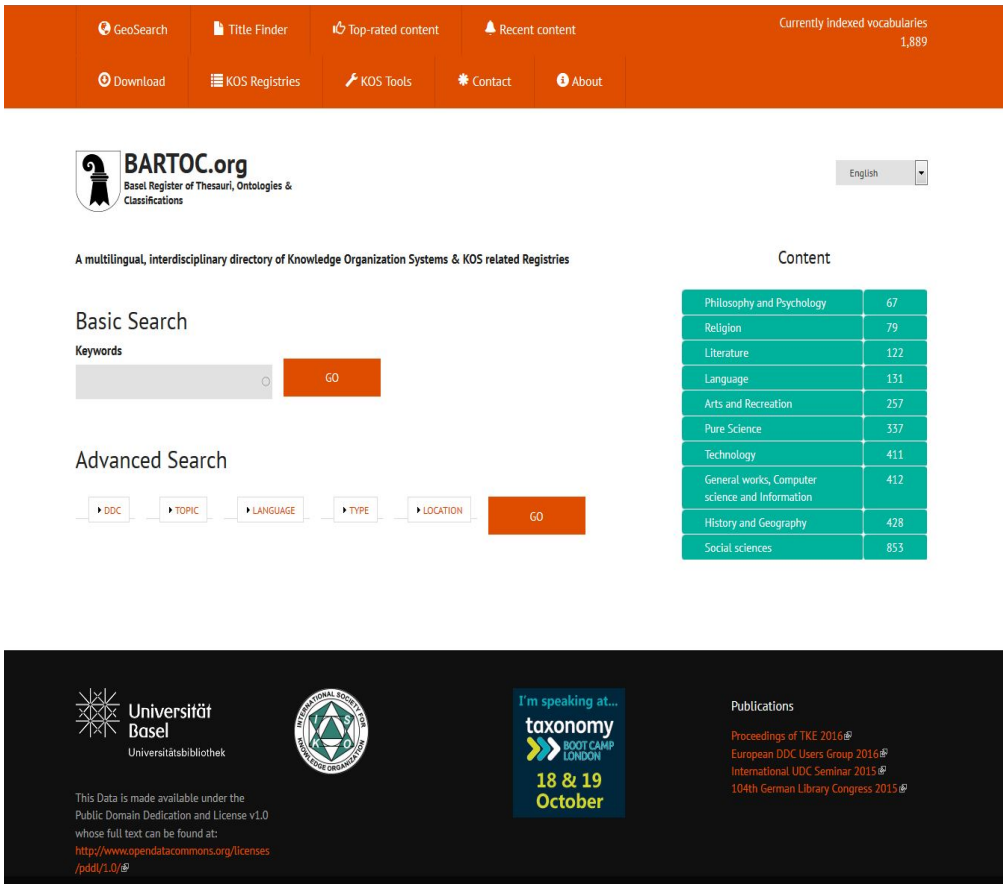

Fig. 1: BARTOC Homepage

2.2 Key features and collaborations

BARTOC's search interface is available in 20 European languages and offers four different ways to search for content (Fig. 1): The Basic Search is suitable for queries based on multilingual keywords entered in the search box. There is an autocomplete function which also calculates the number of results that you get by choosing the keyword; the Advanced Search provides drop-down menus for predefined search criteria like DDC, Topic, Language, Type or Location. Regardless what type of search was conducted, there will always be facets on the right side of the result list to further refine search results. And there are two more options in the navigation bar: Each record in BARTOC is georeferenced, so that a special "GeoSearch" can be offered as a third variant; and finally there is the "Title Finder" with autocomplete function for known item searches.

Besides the editorial offices, BARTOC is trying to set up partnerships with experts, associations, institutions and companies in order to support further development. With regard to the content, the registry has recently been able to import ~14'000 glossaries from Glossarissimo!,[4] which will be enriched with metadata and published over time. Moreover, thesaurus managers, taxonomy developers and ontology engineers from international organizations, universities or public administrations indicate changes in their vocabularies out of their own. BARTOC is also used as a test area by researchers, e.g. to examine how a phenomenon-based classification performs compared to a discipline-based one. This all shows that BARTOC has become a well-established service in the international field of KOS, used by ~600'000 visitors up to now (August 2016). Since there is overlap and complementarity between the data and services of BARTOC and Linked Open Vocabularies (LOV)[5] there are already comprehensive talks about future cooperation as well. The most advanced collaborative effort is actually existing with project coliconc, which is not only an editorial office but also a partner in data curation, access and interoperability.

2.3 Description of KOS in BARTOC

The metadata scheme used to describe KOS in BARTOC originates from the early days when BARTOC was just a blog called "Thesaurusportal". With migration of the database to Drupal CMS the scheme was extended with a mapping to RDFa and the database was labeled with an Open Data Commons Public Domain Dedication and License (PDDL),[6] so KOS description in BARTOC can be used as Linked Open Data (LOD). The current metadata scheme contains "a relatively sufficient amount of metadata" (Bratková et al., 2014) with fields for Title, Abstract, Alternative Title, Size, Year of Creation, Author, VIAF, Link to the resource, Topic, DDC, ILC, Access, Format, Type, Wikidata/Wikipedia (links to corresponding records), Language, Location, Geopoint (which is populated automatically from the address field) and license. The fields have been mapped to SKOS, Dublin Core, schema.org, FOAF and SIOC (Ledl and Voß, 2016, table 1). As described in section 4, this RDF representation is further mapped to JSKOS format to be downloaded or retrieved via JSKOS-API.

Bratková et al. note BARTOC's "advantage that it specializes in supplementing Dewey's decimal classification terms (up to the third hierarchic level) ... , as well as providing the multilingual EUROVOC thesaurus descriptors" (Bratková et al., 2014). KOS are further indexed with the former mentioned KOS Types Vocabulary developed by the DCMI NKOS Task Group. Recently we have begun to tag BARTOC's content also with notations and captions from ISKO's Integrative Levels

[4] http://glossarissimo.wordpress.com/

[5] http://lov.okfn.org/dataset/lov/

[6] http://opendatacommons.org/licenses/pddl/

Classification (ILC). Controlled vocabularies are also used for format, access, languages, and licenses of KOS but only languages and licenses are published as terminologies as part of the mapping to JSKOS format.

EuroVoc, the Multilingual Thesaurus of the European Union, was chosen, although developed especially for the European parliamentary activities, because it is maintained by a trusted authority, it is open data, its domains are multidisciplinary and its terms are available in 25 languages, which is essential for BARTOC's multilingual search. The Dewey Decimal Classification is the most widely used library classification system in the world and BARTOC is addressing an international audience. This qualifies DDC to make the search interface more easily accessible to wide-ranging groups of users, especially because of the captions available in various languages. Also DDC gives a good overview of BARTOC's content.

3. JSKOS data format and API

3.1 Motivation

The coli-conc project at Verbundzentrale des GBV (VZG) is funded by German Research Foundation (DFG) to facilitate management and exchange of concordances between KOS. This requires collection of information about both KOS and KOS concepts in a common format. To some degree such format is given with the Simple Knowledge Organization System (SKOS) ontology. SKOS allows the exchange of KOS as Linked Data but it comes with the complexity of RDF and it requires extensions with other ontologies to cover more than basic use cases. Several alternative standards exist (MARC21, MADS, ISO 25964-1 XML etc.) and many KOS are not made available in machine readable format at all. Last but not least none of the existing formats is optimized for use in web applications. These shortcomings motivated the creation of a dedicated format for all aspects of KOS data.

3.2 JSKOS data format

The JSKOS data format for Knowledge Organization Systems (Voß, 2016c) combines the benefit of RDF for data aggregation and JSON for easy access and storage. JSKOS defines a set of object types such as concepts, concept schemes, mappings, concordances and registries, and fields for description of these objects. JSKOS object types and fields extend RDF classes and properties found in SKOS to support the most used information found in KOS descriptions. Making use of a JSON-LD mapping, JSKOS can also be converted to and from RDF. In contrast to simple JSON-LD, the format strictly defines how to encode repeatable and non-repeatable fields and removes other possibly ambiguities. An additional feature of JSKOS is the optional support of closed-world statements for instance to express that a concept has more labels than included in a record or to state that no narrower concepts exist.

Figure 2 gives an example of an abbreviated JSKOS record of a record from GeoNames. The application of JSKOS to describe KOS from BARTOC is illustrated in another paper (Ledl and Voß, 2016).

```
{
  "uri": "http://sws.geonames.org/614540/",
  "inScheme": [
    { "uri": "http://bartoc.org/en/node/15" }
  ],
  "prefLabel": {
    "en": "Georgia",
    "fr": "Géorgie",
    "de": "Georgien",
    "-": "..."
  },
  "broader": [
    { "uri": "http://sws.geonames.org/6255147/" }
  ],
  "narrower": [ ]
}
```

Fig. 2: Abbreviated example of a JSKOS record

The advantages of JSKOS compared to plain SKOS/RDF include ease of use, a uniform description of mappings, concordances, and registries in addition to concept schemes and concepts, and optional closed world-statements.

3.3 JSKOS-API

The specification of JSKOS data format is accompanied by a HTTP based API to query KOS data for terminology-based (web) applications. Some terminology services already allow open or limited access to their KOS via custom APIs[7] and several application for KOS management include web services (Voß, 2016d; Cox, 2014), but no common standard has emerged yet. One possible candidate would be SPARQL to access KOS data expressed in RDF but SPARQL and RDF are often too complicated to directly provide and make use of. For this reason an API to access KOS data is being developed in project coli-conc based on JSKOS format. Final specification and implementation of this JSKOS-API requires further evaluation of existing APIs and use cases. A subset has already been published as Entity Lookup Microservice API (ELMA). ELMA covers the most required use cases of terminology services (Voß, 2016b) which are:

- **Entity Lookup** to get JSKOS data of a concept with known URI

[7] For instance the ZBW library with econ-ws at http://zbw.eu/labs/de/project/econ-ws.

- **Entity Search** to get a list of concepts matching a query string with relevance ranking. This access method is based OpenSearch Suggestions API to be used for typeahead when selecting a concept of unknown URI.

Other parts of JSKOS-API to query concept mappings have been implemented[8] but not fully specified yet. Access to several KOS (VIAF, GND, ORCID, GeoNames…) via ELMA is provided by wrappers to existing terminology services.[9] The wrappers include a transformation of BARTOC RDF to JSKOS (Ledl and Voß, 2016) which is also used to generate daily database dumps of BARTOC in JSKOS format[10]. The service is used to build a terminology service at VZG (Figure 3) among other applications.

Fig. 3: Terminology service of VZG [11]

4. Uniform access to KOS information

The growing number of Knowledge Organization System requires better ways to find and make use of these systems. BARTOC terminology registry gives a good overview of KOS by making them searchable and comparable under one interface. The content of BARTOC is further made available for download without restric-

[8] See coli-conc mapping database at http://coli-conc.gbv.de/concordances/.

[9] See soure code and live demo at https://github.com/gbv/jskos-php-examples.

[10] https://coli-conc.gbv.de/publications/bartoc/

[11] https://taxonomy.gbv.de/

tions.[12] To extend BARTOC from a basic terminology registry to a full terminology registry (Golub et al., 2014) terminology services must be collected or created and unified. Given the JSKOS format to express KOS information (descriptions, content, and mappings), JSKOS-API or ELMA and its implementation to access KOS information, and BARTOC to collect KOS descriptions, all building blocks for uniform description and access to Knowledge Organization Systems are available. The toolset is complemented by Open Source libraries such as ng-jskos[13] and jskos-php[14] to make use of JSKOS(-API) in other applications.[15] The general workflow for uniform access to KOS information is this:

1. terminology publishers create and publish KOS
2. KOS are described in BARTOC (and other terminology registries)
3. KOS description is made available in JSKOS format
4. terminology services provide access to KOS (in many forms)
5. access to KOS is mapped/wrapped to JSKOS(-API)
6. all KOS can be browsed and searched uniformly via JSKOS(-API)

An analogous workflow is being implemented for concept mappings and concordances in project coli-conc. Although BARTOC plays a central role in this scenario, it does not have to be the only terminology registry. Collaboration with other projects and institutions is highly appreciated. Examples include Glossarissimo, as mentioned above, and the open knowledge base Wikidata.[16] The latter already contains entries for many KOS types and instances (Voß 2016a). Wikidata property BARTOC ID[17] has been used to match around 10% of all KOS collected in BARTOC. This connection allows for a distribution of tasks in KOS description between experts with a fixed set of data fields in BARTOC and volunteers with more open metadata schema in Wikidata. As Wikidata is also being mapped to JSKOS, both registries can directly be compared to support data quality analysis.

While availability of KOS description has been improved a lot since the creation of BARTOC, more work needs to be done in mapping or wrapping KOS content to JSKOS format. An important strategy to reach the goal of uniform access to KOS is to encourage terminology providers and terminology services to directly support

[12] BARTOC uses the Public Domain Dedication and Licence (PDDL).

[13] https://github.com/gbv/ng-skos

[14] https://github.com/gbv/jskos-php

[15] See https://coli-conc.gbv.de/publications/ for a list of software and programming libraries.

[16] https://www.wikidata.org/

[17] http://www.wikidata.org/entity/P2689

JSKOS and JSKOS-API by contributions to existing KOS management software (2016d). When more KOS are being made available via JSKOS-API, their API endpoints will be listed in BARTOC for direct access. The publication of more and more KOS via JSKOS-API, will allow users to directly browse and search in KOS from BARTOC and other web applications from any place.

References

Bratková, E., Kučerová, H. (2014): *Knowledge organization systems and their typology.* In: The Library. Revue of Librarianship 25(2), p. 1–25. ISSN 1801-3252. http://full.nkp.cz/nkkr/knihovna142_suppl/1402sup01.htm

Cox, S., Yu, J., Rankine, T. (2014): SISSVoc: A Linked Data API for access to SKOS vocabularies. In: Semantic Web 7(1), p. 9-24. http://doi.org/10.3233/SW-140166

Dublin Core Metadata Initiative (2015). *NKOS Task Group: KOS Types Vocabulary.* http://wiki.dublincore.org/index.php/NKOS_Vocabularies

Golub, K., Tudhope, D., Zeng, M.L., Žumer, M. (2013): *Terminology registries for knowledge organization systems: Functionality, use, and attributes.* In: Journal of the Association for Information Science and Technology 65(9), p. 1901–1916

Hlava, M. (2011): *Developing an eclectic terminology registry.* In: Bulletin of the American Society for Information Science and Technology 37(4), p. 19–22

Ledl, A., Voß, J. (2016): *Describing Knowledge Organization Systems in BARTOC and JSKOS.* In: Proceedings of International Conference on Terminology and Knowledge Engineering, p. 168-178. http://hdl.handle.net/10760/29366

Hodge, G (2000): *Systems of Knowledge Organization for Digital Libraries.* The Digital Library Federation. ISBN 1-887334-76-9

Voß, J. (2016a): *Classification of Knowledge Organization Systems with Wikidata.* In: : Proceedings of the 15th European Networked Knowledge Organization Systems Workshop (NKOS 2016), p 15-22. CEUR Workshop Proceedings, Volume 1676. http://ceur-ws.org/Vol-1676/paper2.pdf

Voß, J. (2016b): Entity Lookup Microservice API (ELMA). v0.0.4 http://gbv.github.io/elma/

Voß, J. (2016c): *JSKOS data format for Knowledge Organization Systems*. v0.1.0 http://gbv.github.io/jskos/

Voß, J. (2016d): *Open Source web applications for Knowledge Organization Systems.* coli-conc report 7. http://doi.org/10.5281/zenodo