Describing Knowledge Organization Systems in BARTOC and JSKOS

Andreas Ledl[1] and Jakob Voß[2]

[1] University Library of Basel, Basel
[2] Verbundzentrale des GBV (VZG), Göttingen

Abstract. This paper introduces a cooperation between the Basel Register of Thesauri, Ontologies & Classifications (BARTOC) and project coli-conc to provide information about Knowledge Organization Systems, which "encompass all types of schemes for organizing information and promoting knowledge management" [6], in uniform form. The result is a proper metadata scheme, the JSKOS data format, and an API to connect and access connecting terminology registries so terminologies can be discovered and explored at one place.

Keywords: knowledge organization systems · terminology registries · metadata schemes

1 Introduction

Over the last twenty-five years a large amount of Knowledge Organization Systems (KOS) such as classifications, thesauri, authority files, and term bases have been published online and new ones are added almost daily. Several terminology registries have emerged to identify, describe and make accessible these KOS, ideally in a human- and machine-readable way. These registries replaced link lists, which usually contained information about only a few well-known controlled vocabularies without elaborated search interfaces or bibliographic description of KOS. The BARTOC terminology registry[3] has quickly evolved to one of the largest collections of information about distinct KOS. This paper summarizes the description of KOS in BARTOC and project coli-conc,[4] and provision of its metadata as Linked Open Data and the uniform JSKOS data format.

---

[3] http://bartoc.org/
[4] https://coli-conc.gbv.de/

2 The Basel Register of Thesauri, Ontologies & Classifications (BARTOC)

According to Golub et al., who identified four types of KOS registries (Metadata Registries, basic or full Terminology Registries, Service Registries and Data Registries), BARTOC is a basic Terminology Registry, because it contains "only the metadata of KOS vocabularies" [4, 1903]. Furthermore, it is a meta registry of KOS registries (see figure 2), linking to 68 other portals.[5] BARTOC differs from other terminology registries on five counts: it includes *any kind* of KOS from *any subject area* in *any language*, *any publication format*, and *any form of accessibility*. This means that it needs universal systems for formal cataloging, classification and subject indexing of knowledge organization systems.

2.1 The origins of BARTOC

The idea for BARTOC has its roots in two classic areas of Library & Information Science: creating bibliographies and teaching information literacy. On the one hand, it is the latest contemporary descendant of intensive efforts in the 20th century to publish printed surveys of the work on KOS. On the other hand, controlled vocabularies are needed to tag pieces of information and to apply complex search strategies like the "block building approach", where a topic is broken down into separate sections to analyze the scope (termino)logically.

It was clear from the start that BARTOC would address the international library community, but also terminologists and scientists from all over the world. Since its launch in November 2013, it has had a total of 300'000 visits and 2.5 million page views.

2.2 BARTOC's current metadata scheme

BARTOC contains "a relatively sufficient amount of metadata" [1]. The metadata scheme used to describe KOS in BARTOC originates from the early days when BARTOC was just a blog called "Thesaurusportal".[6] With migration of the database to Drupal CMS the

---

[5] http://bartoc.org/en/terminology-registries
[6] http://www.profi-wissen.de/hilfsmittel-fuer-alle-denkbaren-recherchegebiete-thesaurus-porta/

schema was extended with a mapping to RDF, so KOS description in BARTOC can be used as Linked Open Data. Table 1 lists all current metadata fields including their mapping to JSKOS (see section 3 and figure 3 later) and RDF. The mapping to RDF makes use of schema.org, FOAF, and SKOS ontology.

Table 1. Metadata schema and mappings of KOS description in BARTOC

Field	JSKOS	RDF
URI	`uri`	subject URI
Title	`prefLabel`	`skos:preflabel`, `schema:name`
Alternative or English Title	`altLabel`	`skos:altLabel`, `schema:name`, `dct:title`, `foaf:name`
Author	`creator`	`dct:creator`, `schema:creator`
Abstract	`scopeNote`	`skos:scopeNote`, `dct:description`
Coverage	`subject`	`dct:subject`
Type	`type`	`dct:type`, `rdf:type`
Format	-	`dct:format`
Size	`extent`	`dct:extent`
License	-	`dct:license`, `schema:license`
Access	-	`dct:rights`
DDC	`subject`	`dct:subject`, `schema:about`
DDC Main Class	-[7]	
Wikidata	`identifier`	`skos:exactMapping`, `dct:identifier`
Link	`url`	`schema:url`, `foaf:page`
Language	`language`	`schema:inLanguage`, `dct:language`
Topic	`subject`	`dct:subject`, `schema:about`
Year of Creation	`created`	`dct:created`
Term Translations	-[7]	
VIAF	-[8]	
Address	-[8]	
Location	-[8]	

2.3 Alignment with NKOS AP

Both BARTOC and JSKOS origin in a bottom-up process by actual description of knowledge organization systems. For this reason the current state is not finished until it has been tested sufficiently in several real-world applications. The Networked Knowledge Organization Systems Dublin Core Application Profile (NKOS AP), created between 2010 and 2015 followed the opposite direction by theoretical investigation of KOS and their registries. The resulting metadata

---

[7] Only used for searching.
[8] Not refering to the KOS but to its publisher.

scheme is expected to be "very important to terminology registries, service registries, vocabulary users (machine or human), and retrieval systems" [11]. A comparison of the current metadata scheme of BARTOC, JSKOS, and NKOS AP resulted in an overlap at 13 of 28 fields for BARTOC and 18 for JSKOS (table 2).

Table 2. Mapping of NKOS AP to BARTOC and JSKOS

NKOS AP field	BARTOC	JSKOS
`dct:title`	Title	`prefLabel`, `altLabel`
`dct:creator`	Author	`creator`
`dct:publisher`	Author	`publisher`
`dct:description`	Abstract	`scopeNote`
`dct:subject`	Coverage, Topic, DDC	`subject`
`dct:type`	Type	`type`
`dct:language`	Language	`languages`
`dct:identifier`	URI, Wikidata	`uri`, `identifier`
`dcat:contactPoint`	Link	`url`
`dct:license`	License	`license`
`nkos:sizeNote`	Size	`extent`
`dct:format`	Format	-
`dct:created`	Year of Creation	`created`
`dct:issued`	-	`issued`
`dct:modified`	-	`modified`
`wdrs:describedBy`	-	`subjectOf`
`dct:isPartOf`	-	`partOf`
`prov:wasDerivedFrom`	-	`versionOf`
`nkos:serviceOffered`	-	`concepts`, `types`
`dct:audience`	-	not defined yet
`nkos:basedOn`	-	not defined yet
`nkos:updateFrequency`	-	to be discussed
`nkos:usedBy`	-	to be discussed
`nkos:alignedWith`	-	to be discussed
`frbrer:isRealizationOf`	-	to be discussed
`frbrer:isEmbodimentOf`	-	to be discussed
`dct:relation`	-	to be discussed
`adms:sample`	-	to be discussed

2.4 Use of controlled vocabularies to describe KOS

One particularly special feature of BARTOC, compared to other terminology registries, is its use of controlled vocabularies to describe KOS. It is considered as 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" [1]. The KOS used to describe other KOS in BARTOC are described below. Each of them also has a BARTOC record in, given with its corresponding URI.

EuroVoc (http://bartoc.org/en/node/15) 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. EuroVoc subject headings can be selected as *Topic* in Advanced Search.

DDC (http://bartoc.org/en/node/241) is the most widely used library classification system, translated in more than 30 languages. DDC codes up to the third hierarchy level enable grouping different KOS according to a certain field or topic. To make the search interface more easily accessible to wide-ranging groups of users, BARTOC provides DDC numbers and/or captions for content statistics, in the Advanced Search and in faceted search. The service is based on a subscription model. DDC was further expressed as Linked Data [7] and project coli-conc investigates the connection of DDC to other classification systems so it can be used as mapping backbone with other systems and content.

KOS Types Vocabulary (http://bartoc.org/en/node/1665) was developed by the DCMI NKOS Task Group [3] and is, as far as we see, the only controlled vocabulary for KOS types. It differentiates between 14 different types of KOS (categorization scheme, classification scheme, dictionary, gazetteer, glossary, list, name authority list, ontology, semantic network, subject heading scheme, synonym ring, taxonomy, terminology, and thesaurus).

Wikidata (http://bartoc.org/en/node/1940) is a general purpose database and authority file that anyone can edit. By now BARTOC only contains mappings to corresponding KOS records in Wikidata to provide links to Wikipedia articles.

Additional vocabularies are used for format, license, and languages of KOS but they have not been published as terminologies yet.
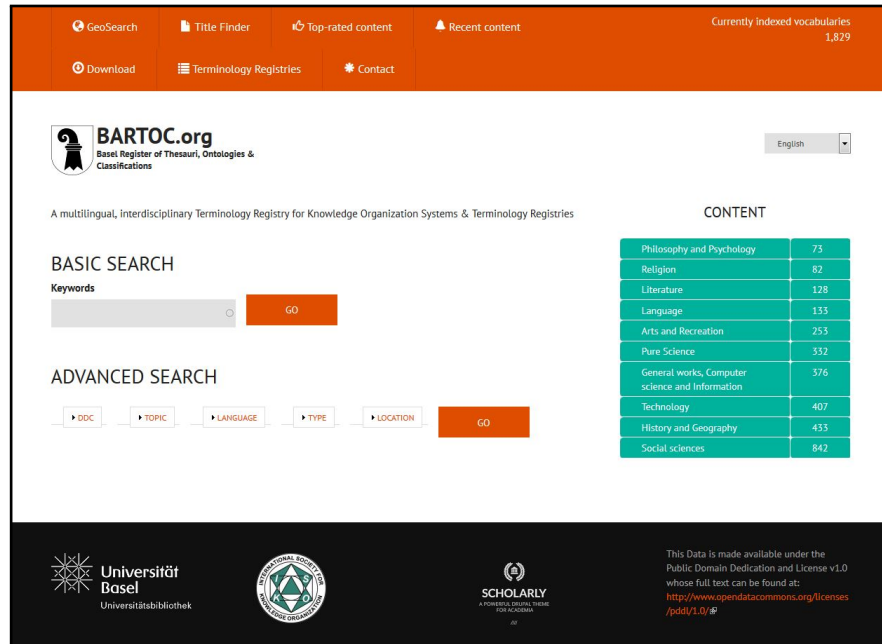
Fig. 1. Screenshot of BARTOC's search interface

3 JSKOS data format for Knowledge Organization Systems

The coli-conc project at Verbundzentrale des GBV (VZG) is funded by German Research Foundation (DFG) to facilitate management and exchange of concordances between knowledge organization systems. This includes the collection and provision of information about KOS and its concepts in a uniform 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 on the Web but it comes with the complexity of RDF and it requires extensions to cover more than basic properties. To better support use of KOS data, especially in web applications, the JSKOS data format for Knowledge Organization Systems is precisely defined, tested, and documented [10]. JSKOS is also compatible with JSON-LD so it can be mapped to and from SKOS/RDF, if needed.[9]

---

[9] JSON-LD defines general mapping rules from JSON to RDF. General JSON-LD, however, has too many degrees of freedom, in contrast to JSKOS.

3.1 JSKOS metadata scheme

In a nutshell, JSKOS supports the following object types:

– **Concepts** as basic entities of all KOS are covered well by SKOS. JSKOS only adds general fields from Dublin Core and common fields found in authority records.
– **Concept Schemes** are equivalent to KOS. In addition to descriptive fields a link to an API can be provided for querying concepts from this concept scheme. Figure 3 shows EuroVoc as example of a concept scheme expressed in JSKOS.
– **Concept Types** can be used to broadly group concepts, for instance concepts about places, people, events, and abstract topics.
– **Mappings** and **Concordances** describe mappings between concepts or concept schemes. This is a major contribution of JSKOS because support of mappings in plain SKOS is very limited.
– **Registries** collect concepts, concept schemes, concept types, mappings, concordances and/or other registries. Registries have no counterpart in SKOS neither.

Figure 2 illustrates the application of JSKOS objects to BARTOC. The website contains both a terminology registry and a meta registry of other terminology registries. Each KOS in BARTOC can be described as JSKOS Concept Scheme. The concepts of each KOS are not included in BARTOC but project coli-conc provides converters and mappings to make them accessible via downloads and an API. The metadata fields to describe objects in JSKOS are consistent for all object types, for instance `prefLabel` is used for both concept labels and concept scheme titles (see figure 3).

3.2 JSKOS-API

Reusing terminologies does not only require a uniform data format but also methods to access and query selected parts of a KOS. Such methods can be provided either by downloading and importing the whole KOS into a database or by querying an existing service via API. Several APIs and services exist for selected KOS (for instance WebDewey[10] for DDC) but without common standard and many

---

[10] `http://dewey.org/webdewey/`. This service is based on a subscription model.
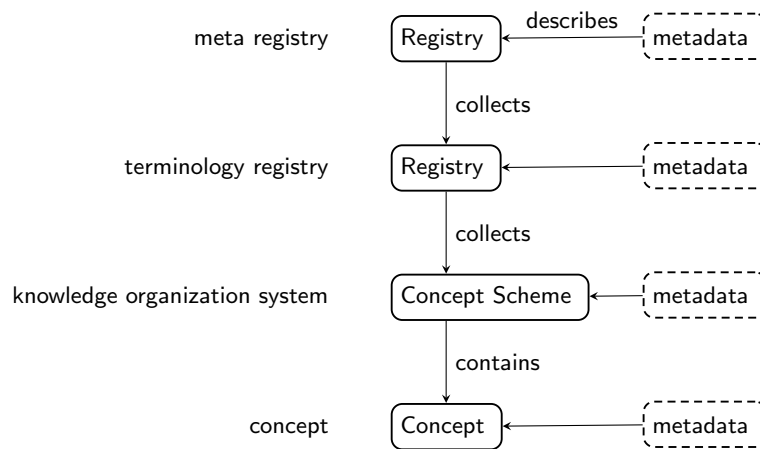
Fig. 2. Overview of metadata about KOS and registries

```
{
  "@context": "https://gbv.github.io/jskos/context.json",
  "id": "http://bartoc.org/en/node/15",
  "prefLabel": {
      "en": "Multilingual Thesaurus of the European Union"
  },
  "altLabel": { "en": "EuroVoc" },
  "url": "http://eurovoc.europa.eu/",
  "identifier": [ "http://www.wikidata.org/entity/Q1370467" ],
  "type": [
      "http://www.w3.org/2004/02/skos/core#ConceptScheme",
      "http://bartoc.org/en/taxonomy/term/1",
      "http://bartoc.org/en/taxonomy/term/2"
  ],
  "subject": [ {
      "id": "http://dewey.info/class/001",
      "prefLabel": { "en": "Knowledge" }
    }, {
      "id": "http://eurovoc.europa.eu/4060",
      "prefLabel": { "en": "European Union" }
  } ],
  "languages": [ "bg", "ca", "hr", "cs", "da", "nl", "en", "et", "fi",
    "fr", "de", "el", "hu", "it", "lv", "lt", "mk", "mt", "pl", "pt",
    "ro", "sr", "sk", "sl", "es", "sv" ]
}
```

Fig. 3. Abbreviated JSKOS record of Eurovoc terminology

terminology provider avoid the technical effort of setting up and maintain an additional web service. For this reason project coli-conc defines JSKOS-API based JSKOS and evaluation of similar APIs.

The full specification of JSKOS-API requires an ongoing overview of uses cases for terminology services [8]. A subset of the most common requirements has already be defined as Entity Lookup Microservice API (ELMA) [9]. The API provides two basic methods of access:

- **Entity Search** queries a list of concepts matching a query string with relevance ranking. The access method is intended for typeahead to select a concept of unknown URI. The response format is the same as OpenSearch Suggestions API [2].
- **Entity Lookup** queries one concept by its URI. The access method is intended to get details about a known concept.

JSKOS-API/ELMA services have been implemented as database application[11] and as wrappers[12] to access GND, Wikidata, ORCID, DDC and other KOS. The implementations are published as open source to be used in other applications as well.[13]

Based on JSKOS-API applications can make use of any KOS that is available in JSKOS format. As BARTOC is also mapped to JSKOS, it can be accessed by the same method. Planned applications at VZG include a tool to create and evaluate concept mappings, and a general terminology service ("Normdatendienst") to provide a uniform search and browsing interface to multiple terminologies.

4 Summary

The Basel Register of Thesauri, Ontologies & Classifications prepares thousands of Knowledge Organization Systems under one interface in order to achieve greater visibility, to highlight their features, to make them searchable and comparable, and to foster knowledge sharing.[14] BARTOC covers a lot of user tasks, allowing "to find, identify, select, obtain . . . KOS resources through the data provided" [4, 1906]. When

---

[11] See https://github.com/gbv/cocoda-db

[12] See https://jskos-php-examples.herokuapp.com/.

[13] See https://coli-conc.gbv.de/publications/ for a current list of software.

[14] See [5, 20]

a user has found an interesting terminology, he or she is directed to the publisher's site for further investigation. But once the KOS is made available via JSKOS-API, its concepts and structure can directly be explored from other places as well. The publication of more and more KOS via JSKOS-API, as being implement in project coli-conc, will allow users to directly browse and search in KOS from BARTOC. In reverse, the content of BARTOC registry will be searchable from other sites as well.

Due to the mutual benefit for both, BARTOC and coli-conc, it will be the most urgent task to improve alignment of BARTOC metadata scheme, NKOS AP metadata scheme, and JSKOS data format. The advantages of the latter, compared to plain RDF, include ease of use, a uniform description also for mappings, concordances, and registries, and a defined method to query registries and concept schemes. This way both BARTOC and JSKOS(-API) will foster the visibility, availability and usefulness of Knowledge Organization Systems in general.

References

1. Bratková, E., Kučerová, H.: Knowledge organization systems and their typology. Revue of Librarianship 25(2), 1–25 (2014)
2. Clinton, D.: Opensearch suggestions extension. Tech. rep. (2006), http://www.opensearch.org/Specifications/OpenSearch/Extensions/Suggestions/1.0
3. Dublin Core Metadata Initiative. NKOS Task Group: KOS Types Vocabulary (October 2015), http://wiki.dublincore.org/index.php/NKOS_Vocabularies
4. Golub, K., Tudhope, D., Zeng, M.L., Žumer, M.: Terminology registries for knowledge organization systems: Functionality, use, and attributes. Journal of the Association for Information Science and Technology 65(9), 1901–1916 (2014)
5. Hlava, M.: Developing an eclectic terminology registry. Bulletin of the American Society for Information Science and Technology 37(4), 19–22 (2011)
6. Hodge, G.: Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. The Digital Library Federation (2000), http://www.clir.org/pubs/reports/reports/pub91/contents.html
7. Panzer, M.: Dewey: how to make it work for you. Knowledge Quest 42(2), 22–29 (2013)
8. Voß, J.: Anforderungen an Normdatendienste. Tech. Rep. 4, coli-conc (April 2016), https://dx.doi.org/10.5281/zenodo.50180
9. Voß, J.: Entity Lookup Microservice API (ELMA). Tech. Rep. Version 0.0.3 (March 2016), http://gbv.github.io/elma/
10. Voß, J.: JSKOS data format for Knowledge Organization Systems. Tech. Rep. Version 0.1.0 (February 2016), http://gbv.github.io/jskos/
11. Zeng, M.L., Žumer, M.: Networked knowledge organization systems dublin core application profile (nkos ap) (2015), http://nkos.slis.kent.edu/nkos-ap.html