Preparing BARTOC for the future of Knowledge Organisation

NKOS Consolidated Workshop
2023-11-09
Jakob Voss, Uma Balakrishnan
Outline

- Introduction
- Numbers
- Basic Features
- Recent developments
- Next steps
- Challenges
Introduction

- Basel/Basic Register of Thesauri, Ontologies & Classifications
- **Initiated** by Andras Ledl in 2013/2014 at University of Basel
- Hosted by VZG since 2020
- **Most comprehensive web catalog** of knowledge organization systems (KOS) also known as *vocabularies* or *terminologies*
Vocabulary Metadata

- Information about the vocabulary (abstract, publisher, subject domain...)
- Access related Information (format, license,...)
- Content (language, size…)
- Identifiers other technical data

*Everything is accessible as Linked Open data, via API & in JSKOS format*
BARTOC Editors

- 18 editors from 14 countries
- additional contributors
- yearly Editorial Meeting
Numbers
Vocabularies in BARTOC

3446 vocabularies

- Thesaurus: 24.2%
- Classification: 30.6%
- Ontology: 22.2%
- Glossary: 10.7%
- Subject headings: 3.3%
- Other: 9.0%

Records created:
- 2021: 111
- 2022: 27
- 2023: 23

Records updated:
- 2021: 187
- 2022: 153
- 2023: 273
Terminology registries in BARTOC

126 catalogs and databases of terminologies (including BARTOC)
https://bartoc.org/registries

- 102 registries in 2021, 30 with API
- 122 (minus 5) registries in 2023, 56 (minus 1) with API

Plan: Query registries via API & merge data of their terminologies in BARTOC
## API access to vocabularies

This table lists APIs that provide access to vocabularies and their implementations in various registries. The total number of vocabularies and their listings in registries are also provided.

<table>
<thead>
<tr>
<th>API</th>
<th>implemented</th>
<th>vocabularies</th>
<th>listed in registries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skosmos</td>
<td>yes</td>
<td>278</td>
<td>~ 500</td>
</tr>
<tr>
<td>JSKOS API</td>
<td>yes</td>
<td>98</td>
<td>~ 350</td>
</tr>
<tr>
<td>SPARQL</td>
<td>no (depends on)</td>
<td>15</td>
<td>~ 2.000</td>
</tr>
<tr>
<td>Medical Subject Headings (SPARQL)</td>
<td>yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Library of Congress</td>
<td>yes (partial)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Reconciliation API</td>
<td>yes</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Skohub</td>
<td>yes</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Medical Subject Headings SPARQL</td>
<td>yes (partial)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>MyCore Classification XML file</td>
<td>yes</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Library of Congress API</td>
<td>yes (partial)</td>
<td>2</td>
<td>~ 120</td>
</tr>
<tr>
<td>Other Web Services</td>
<td>no (diverse)</td>
<td>20</td>
<td>&gt; 1.000</td>
</tr>
<tr>
<td>OntoPortal</td>
<td>no</td>
<td></td>
<td>~ 1.300</td>
</tr>
<tr>
<td>Ontology Lookup Service (OLS)</td>
<td>no</td>
<td></td>
<td>~ 500</td>
</tr>
</tbody>
</table>

* *In total 388* in total *~ 5.000* |

* *without duplicates*
Basic Features
<table>
<thead>
<tr>
<th>Name</th>
<th>Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library of Congress Name Authority File</td>
<td><img src="API" alt="Library of Congress Name Authority File" /></td>
</tr>
<tr>
<td>&quot;The Library of Congress Name Authority File (NAF) file provides authoritative data for names of persons, organizations, events, places, and titles...&quot;</td>
<td></td>
</tr>
<tr>
<td>Library of Congress Subject Headings</td>
<td><img src="API" alt="Library of Congress Subject Headings" /></td>
</tr>
<tr>
<td>&quot;LCSH in this service includes all Library of Congress Subject Headings, free-floating subdivisions (topical and form), Genre/Form headings, Children's (AC) headings, and validation strings for which authority records have been created...&quot;</td>
<td></td>
</tr>
<tr>
<td>Library of Congress Demographic Group Terms</td>
<td><img src="API" alt="Library of Congress Demographic Group Terms" /></td>
</tr>
<tr>
<td>&quot;This vocabulary will be used to describe the creators of, and contributors to, resources, and also the intended audience of resources...&quot;</td>
<td></td>
</tr>
<tr>
<td>Library of Congress Cultural Heritage Organizations</td>
<td><img src="API" alt="Library of Congress Cultural Heritage Organizations" /></td>
</tr>
<tr>
<td>A system of classification for Cultural Heritage Organizations developed by the Library of Congress.</td>
<td></td>
</tr>
<tr>
<td>Library of Congress Genre/Form Terms</td>
<td><img src="API" alt="Library of Congress Genre/Form Terms" /></td>
</tr>
<tr>
<td>&quot;The Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT) is a thesaurus that describes what a work is versus what it is about...&quot;</td>
<td></td>
</tr>
<tr>
<td>Library of Congress Children's Subject Headings</td>
<td><img src="API" alt="Library of Congress Children's Subject Headings" /></td>
</tr>
<tr>
<td>&quot;The Library of Congress Subject Headings Supplemental Vocabularies: Children's Headings (LCSHAC) is a thesaurus which is used in conjunction with LCSH...&quot;</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Links</td>
</tr>
<tr>
<td>------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Finnish Public Libraries Classification System</td>
<td>🌐 API</td>
</tr>
<tr>
<td>Helsinki City Library Classification System</td>
<td>🌐 API</td>
</tr>
<tr>
<td>Jefferson's Library Classification</td>
<td></td>
</tr>
<tr>
<td>Library of Congress Classification (LCC)</td>
<td>🌐</td>
</tr>
</tbody>
</table>
Deep Link into Catalogues and Tools
Technical Infrastructure

- Terminologies are stored in JSKOS Server: [https://bartoc.org/api/](https://bartoc.org/api/)
- Authentication with ORCID: no passwords are stored
- Web application written in JavaScript
  - Static pages in Markdown
  - Backend server in NodeJS
  - User Interface with EJS (server) and Vue (client)
Technical Infrastructure

- Records contain information about terminology APIs
- Access to terminologies via this APIs using cocoda-sdk (API wrappers)

Everything is Open Source: https://github.com/gbv/bartoc.org
Additional Content

- Meta-Registry of other KOS-registries (LOV, Bioportal,...)
- Information about KOS-Software
- Statistics

Everything is available under free licenses (mainly CC0)
We also provide programming libraries to simplify access to KOS
Recent Developments
Recent developments

▪ New terminologies and data updates
  o Large set of Korean KOS
    (see Collecting Distributed KOSs and Building Meta Information at NKOS 2022)
  o More Information about APIs to access terminologies
▪ Switch from ILC1 to ILC2
▪ Improved Vocabulary Browsing
▪ New mailing list bartoc-editors@lists.gbv.de
▪ Grant proposal at German Research Foundation (DFG) to extend BARTOC
BARTOC Extensions

BARTOC SKOSMOS (2018)
- As Vocabulary server
- SKOSMOS instance to access
  Vocabulary data based on
  SKOS/RDF
- Halt in development since 2020
- VZG took over the instance in 2022
- Most KOS not included so far
  (experimental hosting)

BARTOC Fast (2019)
- Federated search across
  23 vocabulary services
- Currently non functional
- Planned to rebuild in 2024-

https://skosmos.bartoc.org/
Next Steps
Main goals of the proposed extension
Main goals: more and better

Improve **content and search** capabilities

- Update and extend **information** about vocabularies
- Redesign **search** with discovery and knowledge graph interface
- Overall update of technical infrastructure
Main goals: new capabilities

Integrate with other systems of vocabulary management
- Import from other terminology registries
- Integrate mappings and links to indexed content

Better support research data management
- Support editions and relations between terminologies
- Improve usability and workflows of editing
Main goals: provide service

Support *terminology management*
- Collection of tutorials, guidelines, software…
- Information about **hosting** your own vocabularies

**Outreach**
- Build a bigger active **community** to enhance and facilitate the use of KOS
- Extend and improve **collaborative work** amongst the interested institutions
Challenges
Terminologies are not hyped

- *Knowledge Organization System* (KOS) not widely known beyond library and information science
- Other trends such as AI, LLM and full-text retrieval
  But: Knowledge Graphs are still a hot topic
- Lack of resources to properly document and manage terminologies in BARTOC
Cultural Divide in terminology management

Communities with different language and domains of applications:

- **Vocabularies** in libraries, archives, museums ("humanities")
- **Ontologies** and terminologies in research data management ("sciences")

BARTOC is stronger in the first but known in both cultures.

Example: Mapping formats have been created independently with JSKOS ([https://gbv.github.io/jskos/jskos.html#concept-mappings](https://gbv.github.io/jskos/jskos.html#concept-mappings)) and SSSOM ([https://mapping-commons.github.io/sssom/](https://mapping-commons.github.io/sssom/)). Reinventing wheels should be avoided.
Bridging the divide

NFDI
Terminology Services
Research Data Management
Focus on Ontology Lookup Service (OLS) and OntoPortal
Resources

- Daily updated BARTOC statistics at [https://bartoc.org/stats](https://bartoc.org/stats)
- BARTOC Issues [https://github.com/gbv/bartoc.org/issues](https://github.com/gbv/bartoc.org/issues)
- Editors mailing list [bartoc-editors@lists.gbv.de](mailto:bartoc-editors@lists.gbv.de)
- First funding proposal [https://doi.org/10.5281/zenodo.7673393](https://doi.org/10.5281/zenodo.7673393) (German)

*Don’t hesitate to contact us!* [https://bartoc.org/contact](https://bartoc.org/contact)