

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

Austria

Andreas Blumauer
Managing Director & CEO
Semantic Web Company GmbH / PoolParty

France

Ghislain Atemezing, PhD
R&D Engineer, Mondeca

Germany

Project coli-conc
Head Office of the GBV Common Library
Network

Jakob Voß
Uma Balakrishnan
Stefan Peters

Greece

Dr. Haralabos Papatheodorou
English to Greek medical translator

India

Dr. Biswanath Dutta
Associate Professor, Indian
Statistical Institute, Bangalore

Italy

Claudio Gnoli
Science and Technology Library,
University of Pavia

Kristin Kolshus
Food and Agriculture Organization
of the United Nations (FAO)

The Netherlands

Gerard Coen
Data Archiving and Networked
Services (DANS), The Hague

Norway

Dan Michael O. Hegge
University of Oslo Library

South Korea

Prof. Ziyong Park
Department of Knowledge & Information,
Hansung University

Dr. So Young Yoon
Department of Knowledge & Information,
Hansung University

Spain

Prof. Amadeu Pons
Facultat d'Informació i Mitjans Audiovisuals (UB)

Switzerland

Dr. Andreas Ledl
University of Applied Sciences and Arts
Northwestern Switzerland

Luxemburg

Mihai Paunescu
Publications Office of the EU

USA/North America

Heather Hedden
Data & Knowledge Engineer, Semantic Web
Company

United Kingdom

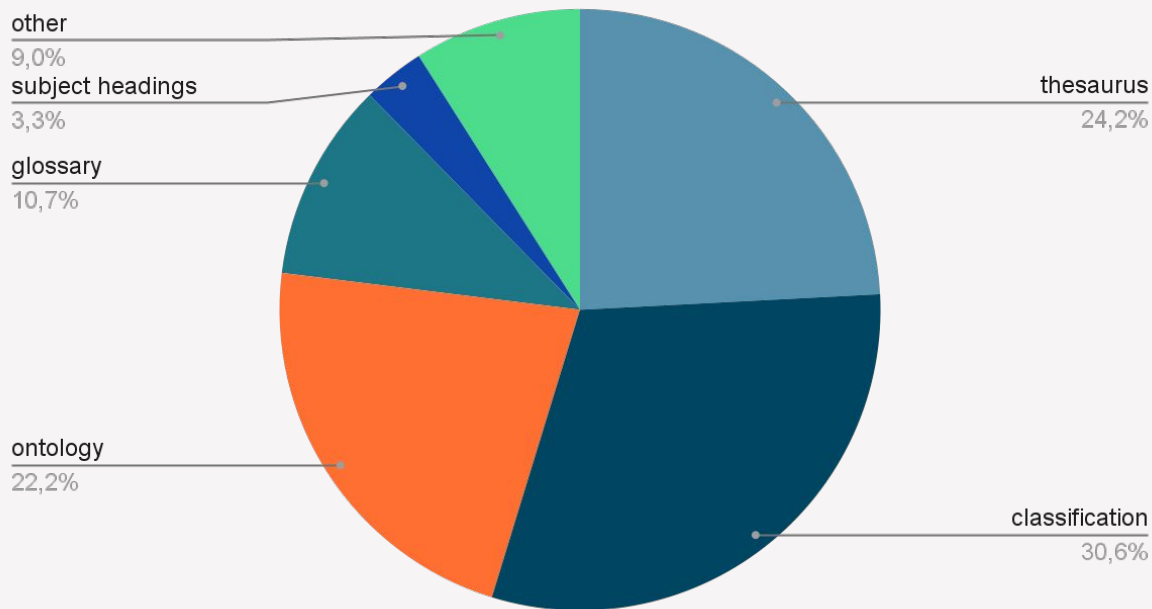
Jeannine Beeken
UK Data Service

Additional contributors include Anne Schuchardt, Christian Rüter, David-Benjamin Rohrer, Erik Stubkjaer, Jarmo Saarikko, Ralph Hafner, Sabrina Gaab, Susanne Arndt, Volkan Çağdaş

Numbers

Vocabularies in BARTOC

3446 vocabularies



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

API	implemented	vocabularies	listed in registries
Skosmos	yes	278	~ 500
JSKOS API	yes	98	~ 350
SPARQL	no (depends on)	15	~ 2.000
Medical Subject Headings (SPARQL)	yes	1	
Library of Congress	yes (partial)	2	
Reconciliation API	yes	4	
Skohub	yes	5	
Medical Subject Headings SPARQL	yes (partial)	1	
MyCore Classification XML file	yes	3	
Library of Congress API	yes (partial)	2	~ 120
Other Web Services	no (diverse)	20	> 1.000
OntoPortal	no		~ 1.300
Ontology Lookup Service (OLS)	no		~ 500
		<i>in total 388*</i>	<i>in total ~ 5.000*</i>

* without duplicates

Basic Features

Vocabularies

Search [Filter](#)

Search

Full-text search across vocabulary description, ranked by relevance

1...64 </>

Name	Links
Library of Congress Name Authority File "The Library of Congress Name Authority File (NAF) file provides authoritative data for names of persons, organizations, events, places, and titles..."	🏠
Library of Congress Subject Headings "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..."	🏠
Library of Congress Demographic Group Terms "This vocabulary will be used to describe the creators of, and contributors to, resources, and also the intended audience of resources..."	API
Library of Congress Cultural Heritage Organizations A system of classification for Cultural Heritage Organizations developed by the Library of Congress....	🏠
Library of Congress Genre/Form Terms "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..."	🏠 API
Library of Congress Children's Subject Headings "The Library of Congress Subject Headings Supplemental Vocabularies: Children's Headings (LCSHAC) is a thesaurus which is used in conjunction with LCSH..."	🏠 API

Vocabularies

Search

Filter

KOS Type

Classification schema

Languages

English ×

License

Creative Commons Public Domain Dedication

Subject

DDC

0 Computer science, information & general works

DDC

Sorting

by label

filter



Name

Links

[Finnish Public Libraries Classification System](#) "The WebPLC is the webversion of the Finnish Public Libraries Classification System..."

API

[Helsinki City Library Classification System](#) "The Web-HCLCS is the webversion of the Helsinki City Library Classification System..."

API

[Jefferson's Library Classification](#) Thomas Jefferson created this classification for his private library...

[Library of Congress Classification \(LCC\)](#) "The Library of Congress Classification (LCC) is a classification system that was first developed in the late nineteenth and early twentieth centuries to organize and arrange the book collections of the Library of Congress..."

Dewey Decimal Classification

[edit](#)[About](#) [Access](#) [Content](#) [Identifiers](#)Languages [ar](#) [nb](#) [zh](#) [fr](#) [de](#) [el](#) [he](#) [is](#) [it](#) [ru](#) [es](#) [sv](#) [vi](#)

Search and browsing in vocabularies registered in BARTOC is experimental. See [cocoda-sdk](#) for technical background.

Search in vocabulary

Top concepts

- ↕ 0 Informatik, Informationswissenschaft & allgemeine Werke
- ↕ 1 Philosophie & Psychologie
- ↕ 2 Religion
- ↕ 3 Sozialwissenschaften
- ↕ 4 Sprache
- ↕ 5 Naturwissenschaften
- ↕ 6 Technik
- ↕ 7 Künste & Unterhaltung
- ↕ 8 Literatur
- ↕ 9 Geschichte & Geografie
- ↕ T1--0 Hilfstafel 1. Standardschlüssel
- ↕ T2--0 Hilfstafel 2. Geografische Gebiete, Zeitabschnitte, Biografien
- ↕ T3--0 Hilfstafel 3. Schlüssel für Künste, für einzelne Literaturen, für einzelne literarische Gattungen
- ↕ T3A--0 Hilfstafel 3A. Schlüssel für Werke von oder über Einzelautoren
- ↕ T3B--0 Hilfstafel 3B. Schlüssel für Werke von mehr als einem oder über mehr als einen Autor
- ↕ T3C--0 Hilfstafel 3C. Notationen zum Anhängen aufgrund von Anweisungen in Hilfstafel 3B, 700.4, 791.4, 808–809
- ↕ T4--0 Hilfstafel 4. Schlüssel für einzelne Sprachen und Sprachfamilien
- ↕ T5--0 Hilfstafel 5. Ethnische und nationale Gruppen
- ↕ T6--0 Hilfstafel 6. Sprachen

Dewey Decimal Classification

[edit](#)[About](#) [Access](#) [Content](#) [Identifiers](#)Languages [ar](#) [nb](#) [zh](#) [fr](#) [de](#) [el](#) [he](#) [is](#) [it](#) [ru](#) [es](#) [sv](#) [vi](#)

Search and browsing in vocabularies registered in BARTOC is experimental. See [cocoda-sdk](#) for technical background.

Search in vocabulary

Top concepts

- ↑ 9 Geschichte & Geografie
- ↑ 930-990 Geschichte einzelner Kontinente, Länder, Ortschaften; außerirdischer Welten
- ↑ 940-990 Geschichte einzelner Kontinente, Länder, Ortschaften der Neuzeit; Geschichte außerirdischer Welten
- ↑ 94 Geschichte Europas

943 Deutschland und benachbarte mitteleuropäische Länder

<http://dewey.info/class/943/e23/>Deutschland und benachbarte mitteleuropäische Länder¹⁰

📅 2000-02-02 📅 2018-03-09

- ↕ 943.0001-943.0008 Standardschlüssel für Mitteleuropa
- ↕ 943.0009 Archäologie; Zeitabschnitte; Gebiete, Regionen, Orte im Allgemeinen; Sammelbiografien von Personen aus Mitteleuropa
- ↕ 943.001-943.009 Standardschlüssel für Deutschland
- ↕ 943.01-943.08 Zeitabschnitte der Geschichte Deutschlands
- ↕ 943.1 Nordostdeutschland
- ↕ 943.2 Sachsen und Thüringen
- ↕ 943.3-943.5 Orte und Gebiete von Bayern, Südwestdeutschland, Nordwestdeutschland
- ↕ 943.6 Österreich und Liechtenstein
- ↕ 943.7 Tschechien und Slowakei
- ↕ 943.8 Polen
- ↕ 943.9 Ungarn

Deep Link into Catalogues and Tools

The image displays a complex web interface for searching and mapping concepts. On the left, a search results page from 'K10plus Verbundkatalog GBV|SWB' shows results for the search term 'b43'. The results list four items, including 'Nazi Volksgemeinschaft Technology' and 'Die österreichische kaiserliche Zensur...'. A sidebar on the right provides a 'DDC Dewey Decimal Classification' tree, showing '943 Germany and neighboring central European countries'. To the right of the sidebar is a 'Mapping Editor' tool with two boxes for mapping concepts. Further right is a 'STW Thesaurus for Economics' interface showing a hierarchy of terms like '16825-0 Central Europe'. Below the mapping editor is a table of 'Concordances' and 'Mappings' between DDC and STW terms.

DDC	STW	16825-0 Central Europe	±0
943 Germany and neighboring central European countries	RVK	NK 2415 Luxemburg	±0
943 Germany and neighboring central European countries	RVK	NK 2410 Belgien	±0
943 Germany and neighboring central European countries	RVK	NK 1900 Österreich (Österreich-Ungarn und Republik)	±0
943 Germany and neighboring central European countries	RVK	NK 2405 Niederlande	±0

Technical Infrastructure

- Terminologies are stored in **JSKOS Server**: <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 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
 - Large set of Korean KOS
(see *Collecting Distributed KOSs and Building Meta Information* at NKOS 2022)
 - 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 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-

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/jkos.html#concept-mappings>) and 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>
- BARTOC Issues <https://github.com/gbv/bartoc.org/issues>
- Editors mailing list bartoc-editors@lists.gbv.de
- First funding proposal <https://doi.org/10.5281/zenodo.7673393> (German)

Don't hesitate to contact us!

<https://bartoc.org/contact>