

Support of Metadata Mapping with coli-conc Infrastructure

DCMI Conference, 15 October 2021

Uma Balakrishnan

Jakob Voß

Stefan Peters



What is coli-conc?

coli-conc.gbv.de

- A service of the Head Office of the GBV Common Library Network, Germany
- It offers **an integrated system for the collection, management and mapping of KOS**
- **It provides**
 - free and uniform access to KOS and their mappings
 - free software to import and export KOS and mapping data
 - a tool for creating and editing mappings with the mapping tool **Cocoda**
 - storage of metadata of the KOS, content of the KOS, and mappings between the KOS
- Funded by **the German Research Foundation**

Partners



Objectives

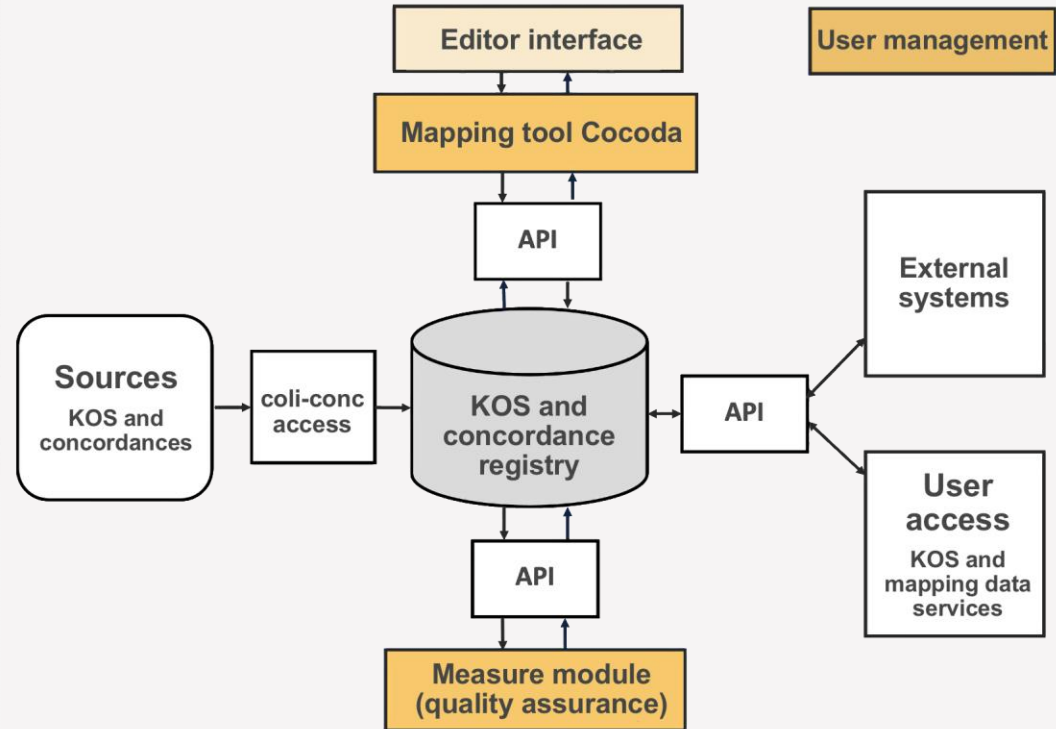


- Manage heterogeneity and create semantic network of KOS
- Catalog enrichment (K10plus)
- Facilitate KOS sharing and catalog resources
- Serve as knowledge base
- Develop and establish format standards
- Enhance the quality of mappings

coli-conc Infrastructure

coli-conc Infrastructure Overview

- Modular system
- Components can be used individually and integrated into existing systems and software
- Founded on common data format JSKOS and JSKOS API



Elements of the System Architecture

Cocoda web-based application

- User and editor interface for KOS, concordances, and mappings
- Developed in JavaScript (Vue framework)
- Open source

Backend Services

- Terminology Services (DANTE API, BARTOC, Skosmos, MarcXML...)
- Mapping Services and Database (collected concordances)
- Mapping Suggestions (co-occurrences and queries)
- Quality Services and Statistics (planned)

coli-conc Components

- The JSKOS data format
- The KOS and concordance registry and database
- Input processes
- Mapping tool Cocoda
- Web-based user interface
- coli-rich - an application for automatic enrichment of catalogues

The JSKOS data format

Facilitate representation, use, and exchange of KOS and mappings

Requires

- An easy-to-use data format (JSKOS)
- An easy-to-use access method (JSKOS-API)

JSKOS is

- based on JSON(-LD) ⇒ compatible with SKOS/RDF
- can represent concordances, mappings, KOS, and KOS data
- has extra features (confidence level of mappings, elements for concepts and co-occurrences, mappings with multiple concepts and ordered lists)
- **Specifications:** <http://gbv.github.io/jskos/>

```
{
  "from": {
    "memberSet": [
      {
        "uri": "http://dewey.info/class/200/e23/"
      }
    ]
  },
  "to": {
    "memberSet": [
      {
        "uri": "http://uri.gbv.de/terminology/thema/QRA"
      }
    ]
  },
  "fromScheme": {
    "uri": "http://bartoc.org/en/node/241"
  },
  "toScheme": {
    "uri": "http://bartoc.org/en/node/1043"
  },
  "creator": [
    {
      "uri": "https://github.com/stefandesu",
      "prefLabel": {
        "en": "Stefan Peters"
      }
    }
  ],
  "type": [
    "http://www.w3.org/2004/02/skos/core#exactMatch"
  ],
  "created": "2021-10-01T08:25:51.542Z",
  "modified": "2021-10-01T08:26:07.817Z",
  "uri": "https://coli-conc.gbv.de/api/mappings/22cb5fdc-196c-4fb0-ad73-9c157cd4b4e9"
}
```

KOS Registry

- Subset of **BARTOC.org**
- KOS currently in use in German-speaking countries
- Enriched with metadata
- Uniform access in JSKOS
- Link to mapping tool Cocoda

Concordance Registry

- Contains currently 27 concordances between different KOS, such as GND-DDC, STW-GND, DDC 1000-RVK, BC-DDC, GND-RVK, RVK-BC, MSC-BC, DDC-LCSH,...
- Total of over 371.000 mappings and concordances
- Wikidata mappings
- Stored in JSKOS format

<http://coli-conc.gbv.de/terminologies/>

<http://coli-conc.gbv.de/concordances/>

Mapping Tool Cocoda

Freely accessible at:

<https://coli-conc.gbv.de/cocoda/app/>

COLI-CONC Cocoda Mapping Tool (dev)

Imprint Accessibility Privacy Policy Feedback Manual Stefan Peters

★ DDC Dewey Decimal Classification

Q Mathematics

500 Science
510 Mathematics
510 Mathematics

Info Search Links Labels coli-ana GND

<http://dewey.info/class/510/e23/>

Mapping Editor

DDC → PB Mathematics

510 Mathematics → PB Mathematics

Concordance Registry: not saved

Stefan Peters

Concordances Search Navigator

Mappings

Concordance Registry 5 of 32

DDC	510 Mathematics	RVK	SA - SP Mathematik	VZG	2013	±0
DDC	510 Mathematics	RVK	QH 100 - QH 170 Mathematik	VZG	2013	±0
DDC	510 Mathematics	BK	31.00 Mathematik: Allgemeines	VZG		±0
RVK	CM 2500 Mathematische Psychologie	DDC	510 Mathematics	Manuela ...		±0
RVK	CM 3000 Methoden der Psychologie (Testtheorie, Skalierung, Faktorenanalyse u.ä.)	DDC	510 Mathematics	Manuela ...		±0

Local Wikidata-Mappings

Recommendations

DDC	510 Mathematics	THEM	PB Mathematics
DDC	510 Mathematics	THEM	PB Mathematics
DDC	510 Mathematics	THEM	P Mathematics and Science
DDC	510 Mathematics	THEM	KJQ Business mathematics and systems
DDC	510 Mathematics	THEM	PB Mathematics
DDC	500 Natural sciences & mathematics	THEM	PB Mathematics

THEMA

THEMA Thema subject classification scheme

Q Mathematics

TP Mathematics and Science
PB Mathematics

Info Labels Search Links

<http://uri.gbv.de/terminology/thema/PB>

Issued: 2013

Philosophy of mathematics
Mathematical foundations
Discrete mathematics
Algebra
Groups and group theory
Number theory
Pre-calculus
Calculus and mathematical analysis
Geometry
Topology
Probability and statistics
Optimization
Combinatorics and graph theory
Applied mathematics
History of mathematics

Tree View

- 500 Science
 - 500 Science
 - 510 Mathematics
 - 510 Mathematics
 - 511 General principles of mathematics
 - 512 Algebra
 - 513 Arithmetic
 - 514 Topology
 - 515 Analysis
 - 516 Geometry
 - 518 Numerical analysis
 - 519 Probabilities & applied mathematics
 - 520 Astronomy
 - 530 Physics
 - 540 Chemistry
 - 550 Earth sciences & geology
 - 560 Fossils & prehistoric life
 - 570 Life sciences; biology
 - 580 Plants (Botany)
 - 590 Animals (Zoology)
 - 600 Technology
 - 700 Arts & recreation
 - 800 Literature
 - 900 History & geography

Tree View

- N History and Archaeology
- P Mathematics and Science
 - PB Mathematics
 - PD Science: general issues
 - PG Astronomy, space and time
 - PH Physics
 - PN Chemistry
 - PS Biology, life sciences
 - Q Philosophy and Religion
 - QD Philosophy
 - QR Religion and beliefs
 - QRA Religion: general
 - QRD Hinduism
 - QRF Buddhism
 - QRJ Judaism
 - QRM Christianity
 - QRP Islam
 - QRR Other religions and spiritual beliefs
 - QRS Ancient religions and Mythologies
 - QRV Aspects of religion
 - QRY Alternative belief systems
 - R Earth Sciences, Geography, Environment, Planning
 - S Sports and Active outdoor recreation
 - T Technology, Engineering, Agriculture, Industrial processes

Mapping tool Cocoda Components

A single-page Web application. **Four main components:**

- **KOS representation:** Concept browser
- **Mapping editor:** Create and modify mapping candidates and assign mapping type
- **Mapping browser:** Browse existing mappings and mapping suggestions
- **Quality measures:** Generator for confidence level and user statistics (planned)

KOS Representation – Concept Browser

- Dropdown menu for KOS selection
- Display of top concept hierarchy
- Hierarchical navigation and detailed display of the concepts
- Display of intra-KOS structural content
(scope notes and linked relative index terms, etc.)
- Display of mapping candidates from different sources
- Deep links into catalogues and other sources

Mapping Browser: Mapping Suggestion Module

Task: For a caption in the source KOS (e.g. DDC)

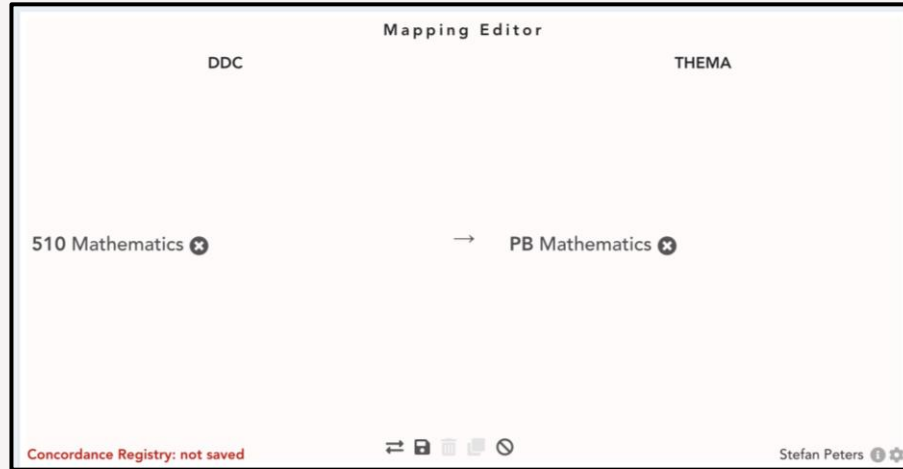
find the best mapping in the target KOS (e.g. THEMA or LCSH)

Automated recommendation services:

- **Mapping database** with the Dewey notation
- **Implicit mappings in the Union catalog** for records with the Dewey notation that also have THEMA or LCSH notation (co-occurrences)
- **Search the target system** with the Dewey captions/relative index and additional terms using query term expansion using other KOS and knowledge bases (such as Wikipedia or mapped GND terms (German Subject Headings))

Mapping Editor Module

- **Create, edit, save, delete mappings and assign** mapping type
- **Export** mappings JSKOS and CSV



coli-rich: Automatic Enrichment of the K10plus

- Existing KOS → Mappings → Enrichment of further indexing systems/KOS

- Example

045F=5010 \$a549

DDC 549 Mineralogy

DDC 549 → BC 38.30

Available Mapping in the database

045Q/01=5301

Additional PICA-Field with the Source Information

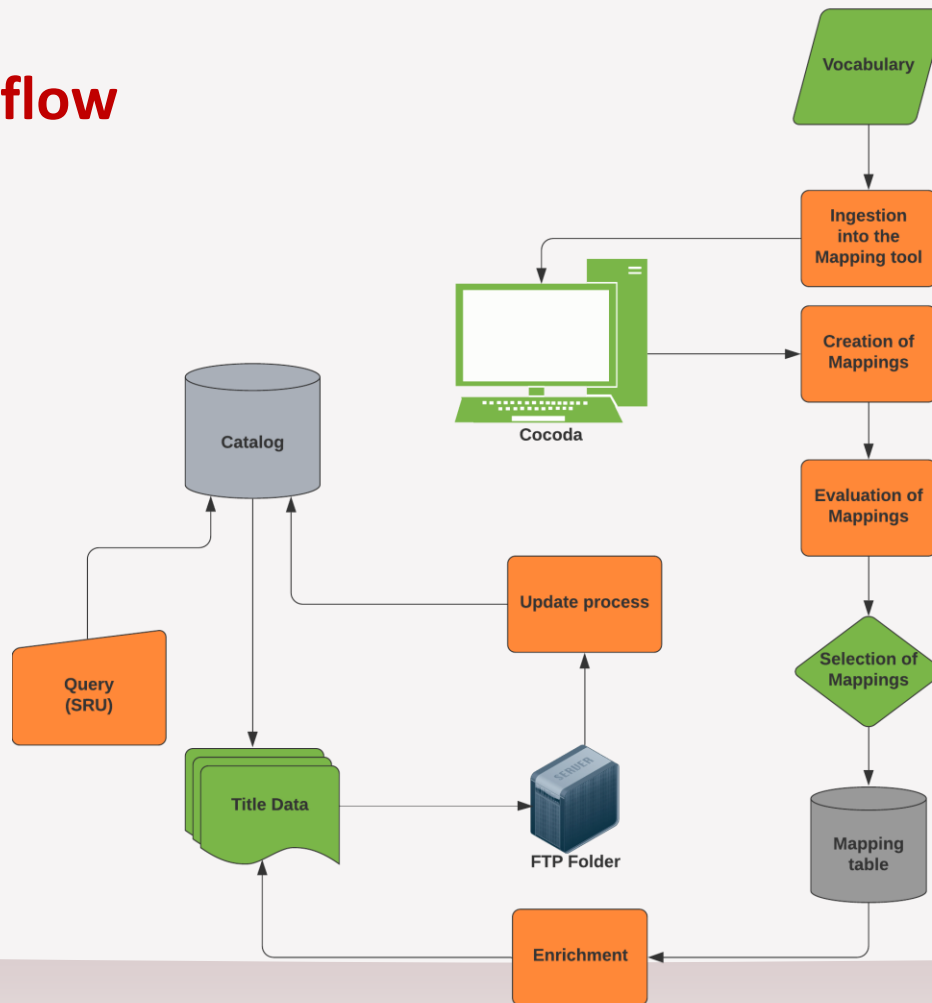
\$a38.30

\$Acoli-conc DDC-BC

\$Ahttps://coli-conc.gbv.de/api/mappings/af8ac88b-f7ab-427a-8e06-9e091d281bdc

More Information : <https://github.com/gbv/coli-rich>

coli-rich Workflow

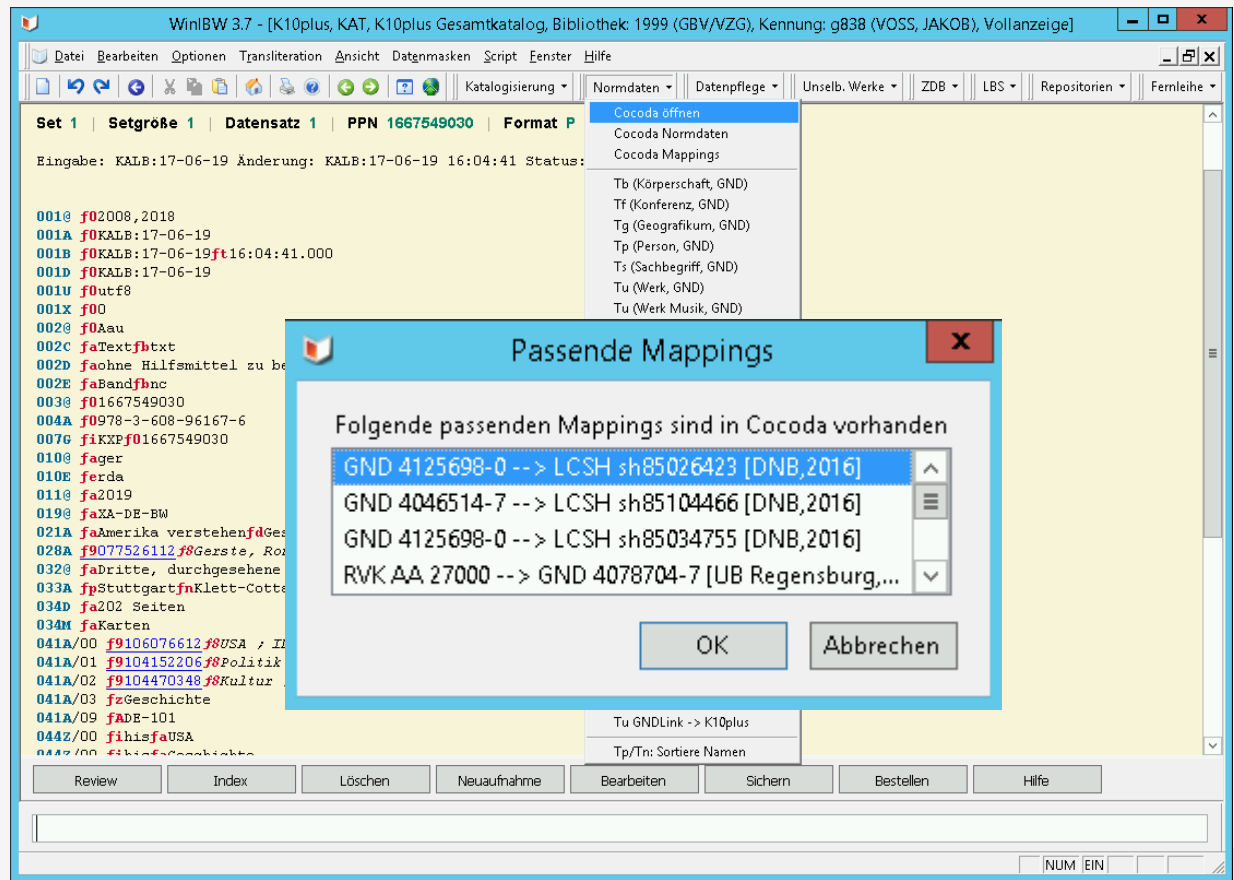


Connection to the external cataloguing and subject indexing systems

WinIBW: Proof of Concept

<https://github.com/gbv/cocoda-winibw>

Digitaler Assistent: planned



Cocoda live Demo and Tutorial

Thank You!

Webseite: <https://coli-conc.gbv.de>

Uma Balakrishnan (project lead): balakrishnan@gbv.de

Jakob Voß (technical coordinator): voss@gbv.de

Stefan Peters (software developer): peters@gbv.de

Twitter: [@coli_conc](https://twitter.com/coli_conc)

You like to collaborate or
become a partner?
Just e-mail us :-)