coli-conc : Mapping Knowledge Organisation Systems

EDUG 2019 - Symposium
National Library of Sweden

Uma Balakrishnan & Stefan Peters

VZG | Verbundzentrale des GBV
www.gbv.de

Agenda

- Overview of the project
- Rational and Aim
- Challenges of mapping
- Roadmap to success
- coli-conc Services
- Mapping Tool Cocoda
- Next Steps

Overview

- coli-conc is a service of the Verbundzentrale des GBV, Germany that provides:
  - free access to and exchange of knowledge organisation systems (KOS) and their mappings. This includes:
    - a KOS registry that gives access to knowledge organisation systems and their metadata in a uniform format
    - an open concordance registry to share concept mappings
    - the mapping tool **Cocoda** to easily create and evaluate mappings
    - free software to import and export KOS and mapping data

Rational (1)

- Huge diversity in the landscape of library classifications
- Absence of a uniform classification system
- Lack of mappings/concordances among the existing library KOS
- Demand for a complete concordance between Regensburg Verbund Classification and Dewey Decimal Classification for over two decades

Rational (2)

- Absence of a single platform to access KOS (actively in use) and retrieve available mappings
- Absence of a uniform exchange format for access and management of both KOS and mappings
- Absence of a tool to aid creation of mappings between any two KOS

KOS manged by the coli-conc services

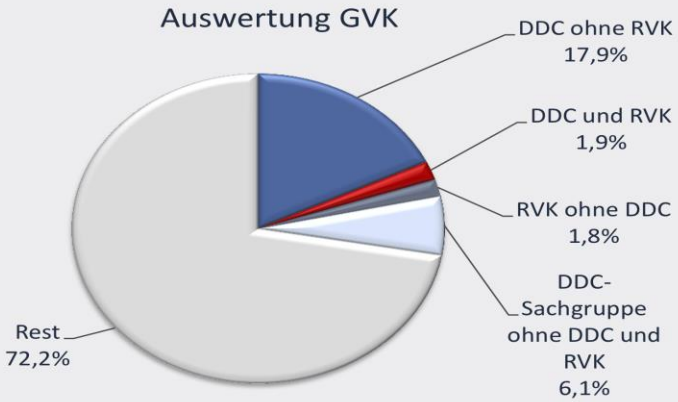
- Classification schemes
- Subject heading schemes
- Thesauri
- Authority lists

Aim

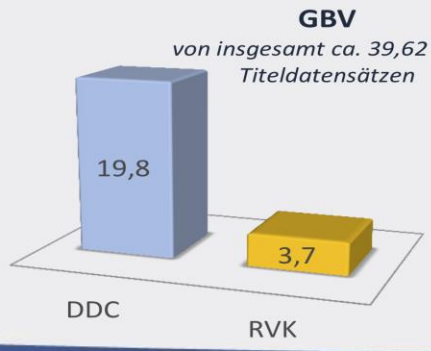
- Enrich the K10plus Union Catalogue
- Improve interoperability and enhance retrieval quality
- Facilitate search expansion
- Facilitate subject cataloguing and support automatic classification

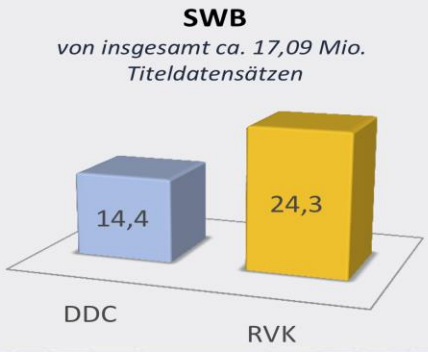
Percentage of DDC and RVK in the union catalogues 2012

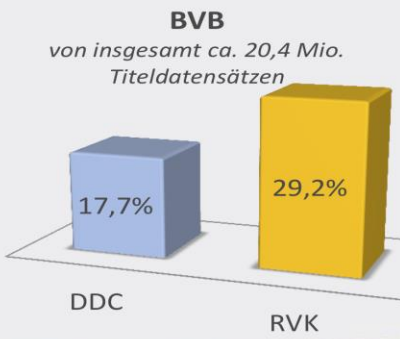
Titeldatensätze in den Verbundkatalogen mit DDC- und RVK-Notationen



von insgesamt ca. 39,62 Mio. Titeldatensätzen







VZG
www.gbv.de

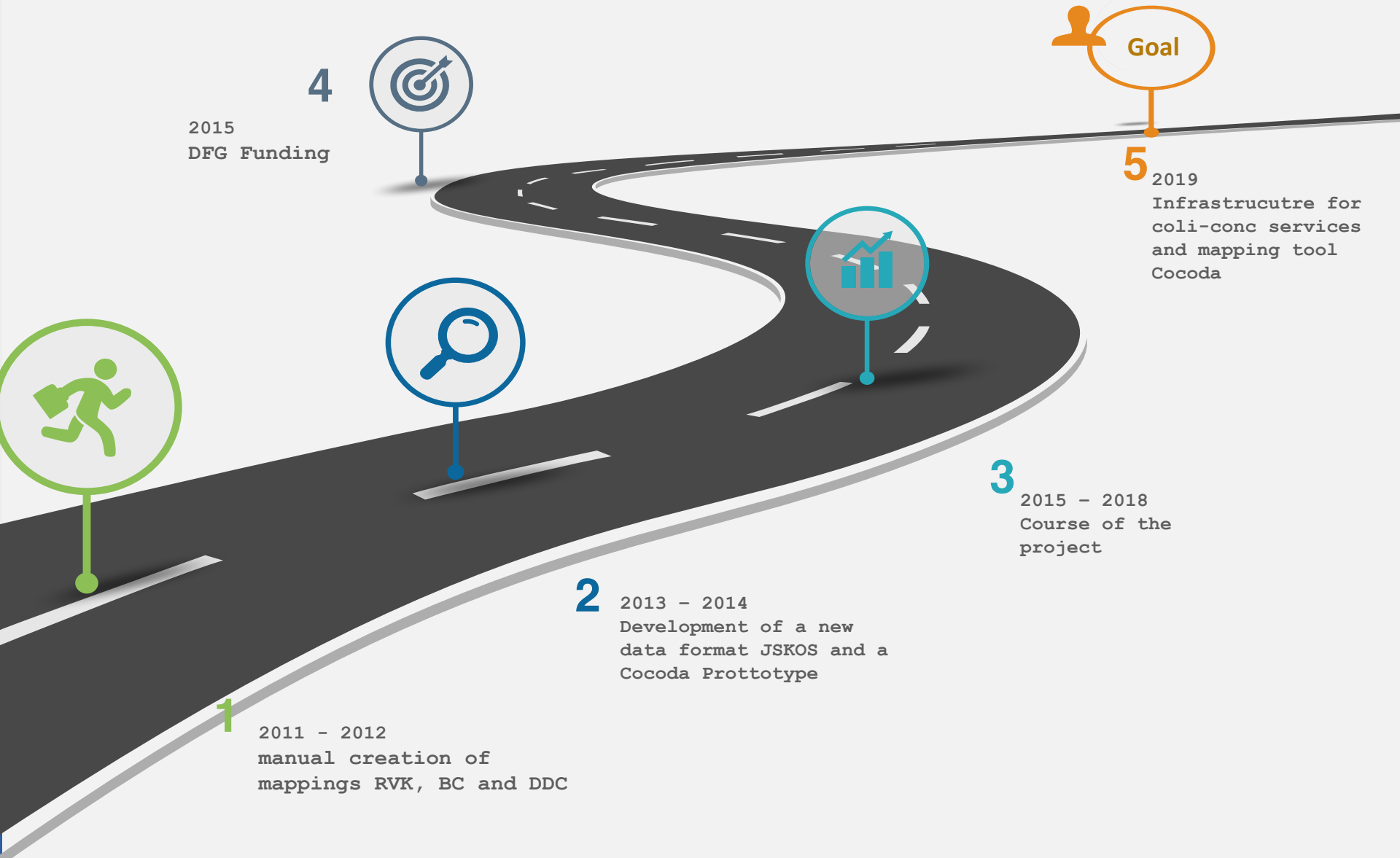
Challenges of mapping DDC and RVK

- History and development
- Language
- Cultural aspects
- Structural aspects
- Domain expertise

Focus of the project

- Library KOS of Germany & German speaking countries
- Develop semi-automatic methods for mapping
- Provide standards and services to ease use and exchange of KOS and mappings
- Facilitate collaborative work
- Enhance the quality of the existing mappings
- Aid efficient management of mappings and KOS on a single platform

Roadmap to Success



1 2011 - 2012
manual creation of mappings RVK, BC and DDC

2 2013 – 2014
Development of a new data format JSKOS and a Cocoda Prottotype

3 2015 – 2018
Course of the project

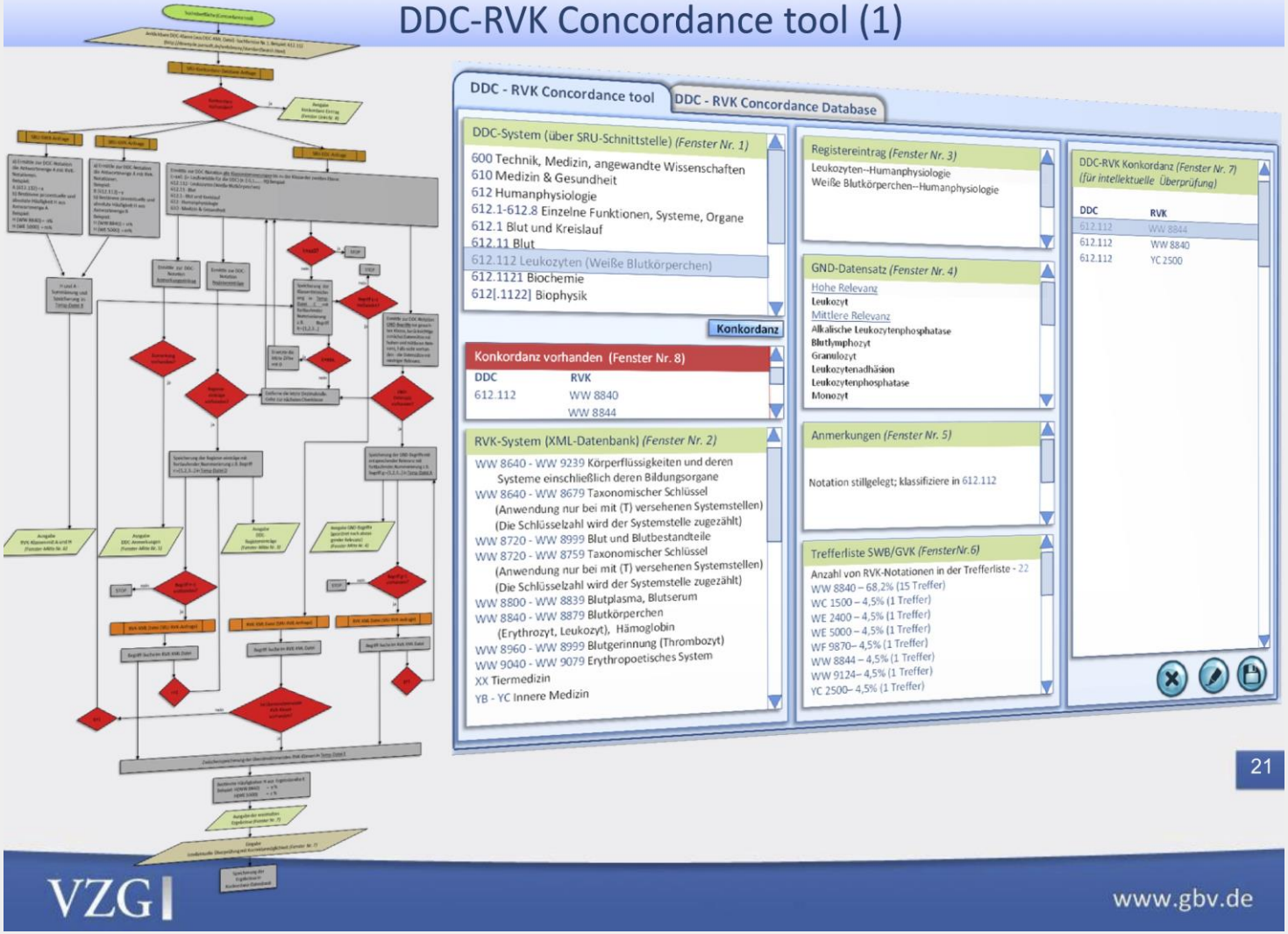
4 2015
DFG Funding

5 2019
Infrastrucutre for coli-conc services and mapping tool Cocoda

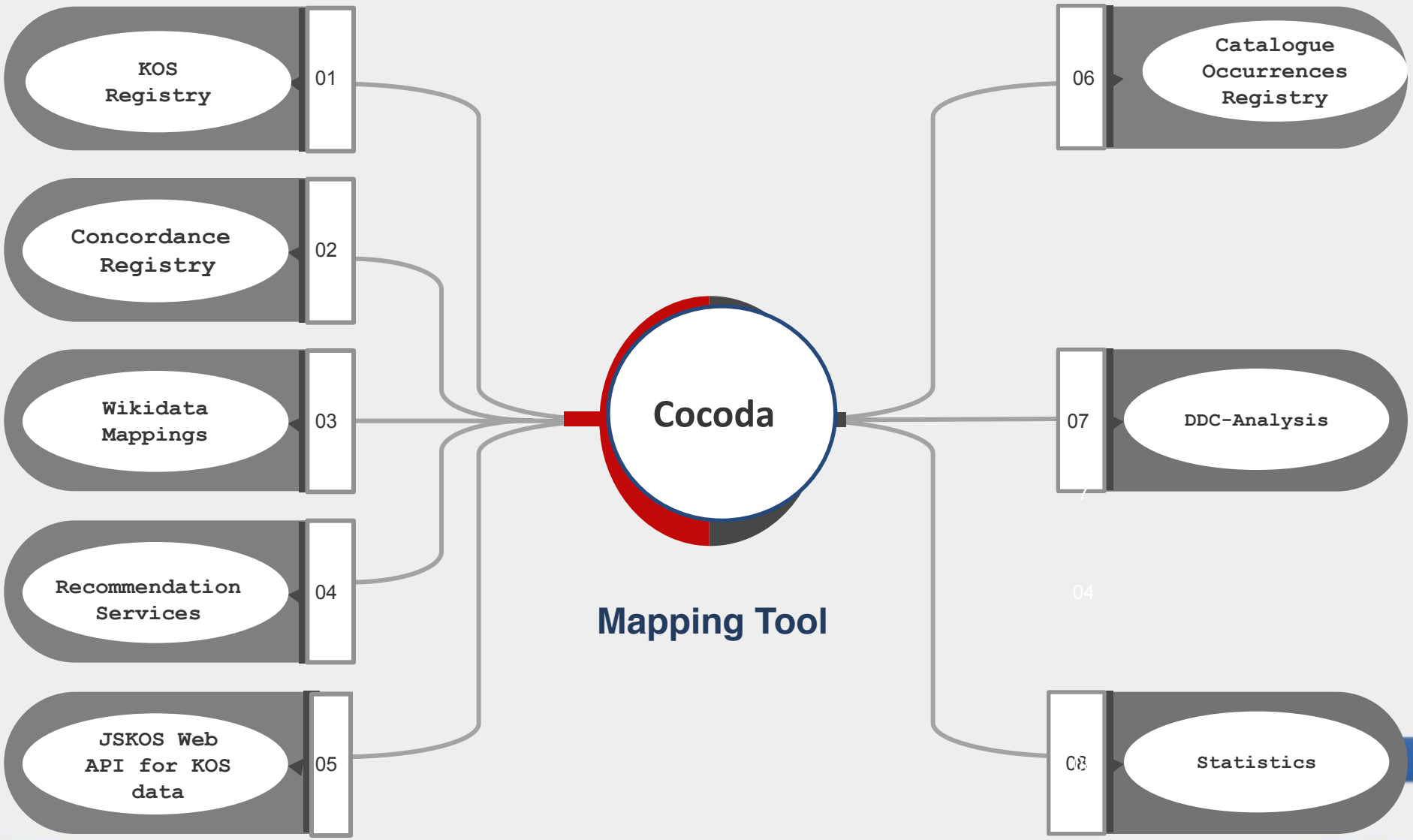
DDC-RVK Mapping Algorithmus and Initial Draft of the Cocoda-Interface

https://www.gbv.de/Verbundzentrale/Publikationen/publikationen-der-vzg-2013/pdf/Balakrishnan_131120_RVK_WS_Konkordanz.pdf

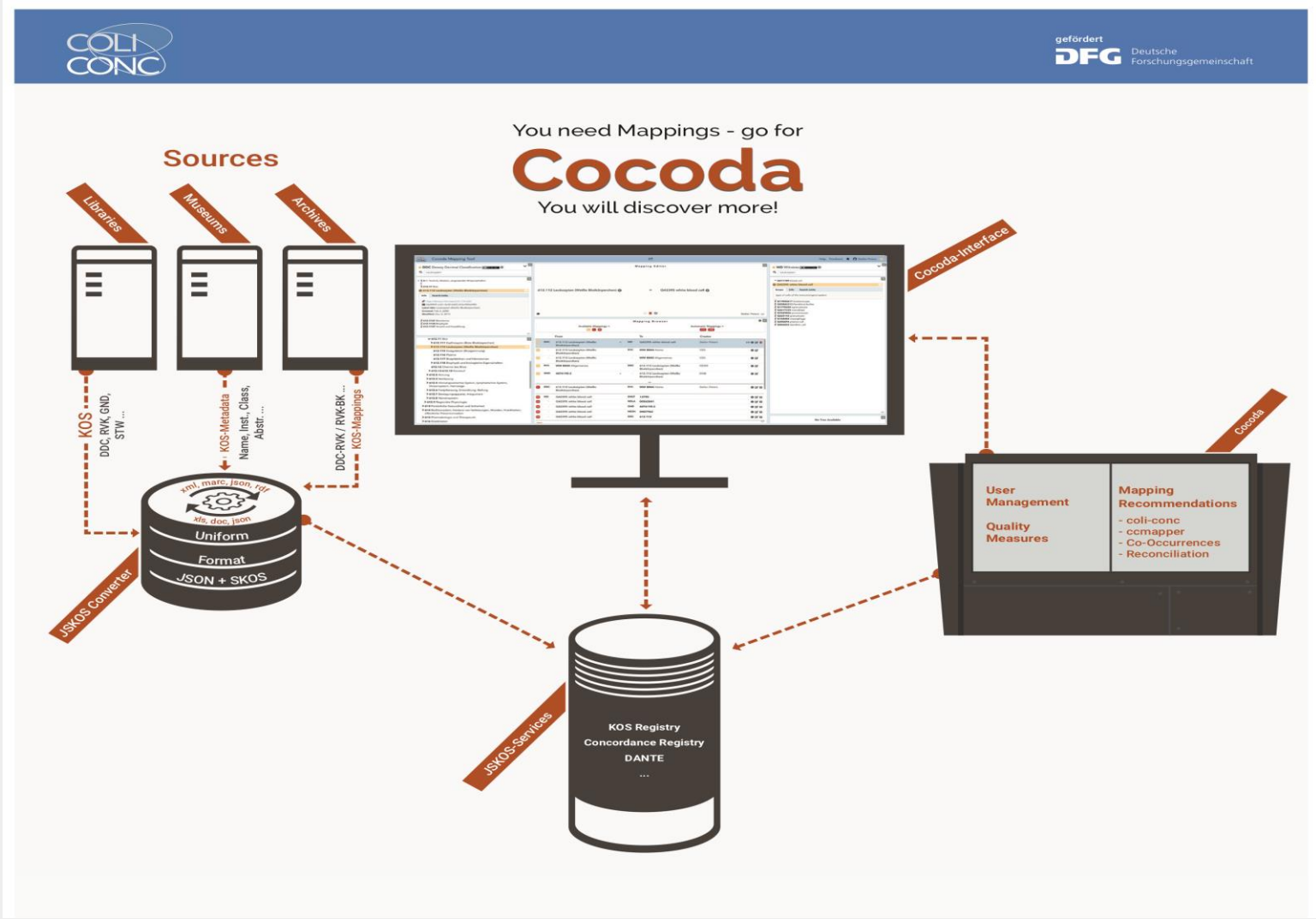
DDC-RVK Concordance tool (1)



coli-conc Services



Cocoda Workflow



Mapping Tool
Cocoda
(Stefan Peters)

10 May 2019
EDUG 2019, Stockholm

Cocoda Mapping Tool

Stefan Peters
Verbundzentrale des GBV (VZG)
stefan.peters@gbv.de

What is Cocoda?

- Cocoda is a web-based tool for semi-automatic creation of mappings between different knowledge organization systems (KOS).
- Goal: Move away from Excel spreadsheets and multiple browser tabs for research, and integrate everything into one interface.
- Primarily developed with library classifications (DDC, RVK, BK, GND, ...) in mind, but supports all kinds of other KOS as well.

Background

- Developed by the GBV Head Office (VZG) in Göttingen, Germany.
- Part of Project coli-conc: https://coli-conc.gbv.de
- The mapping tool and its related services are available as open source repositories on GitHub.
- The data is openly available through APIs.

Data Sources

- Data sources use the JSKOS format (SKOS in JSON-LD, https://gbv.github.io/jskos/).
- In Cocoda, different sources are configurable for
  - vocabularies
  - concepts
  - existing mappings
  - mapping recommendations

Converting and Providing Data

- There are tools to convert data into JSKOS, but we're working on improving those.
- You can either host your own instance of our server to offer JSKOS data, or send us the data so we can import it.
- We are also planning to make the tools more flexible (e.g. to be able to easily add your own data sources without dealing with config files).

Mappings

- Project coli-conc collected existing concordances with around 385,000 mappings.
- Additionally, mapping data from Wikidata is available.
- First use in production by Project KENOM (https://www.kenom.de):
  - Mapping of local systems to GND, Wikidata, Nomisma, and others.
  - ~28,000 mappings created since November 2018.

Semi-automatic Mapping

- Cocoda aims to help with manual and **semi-automatic** mapping.
- Different algorithms that provide mapping recommendations:
  - coli-conc recommendations: simple string matching in target vocabulary
  - CCMapper: created in cooperation with Pansoft, used only for RVK-DDC mappings
  - Catalog co-occurrences
  - OpenRefine Reconciliation API: uses a more complicated string matching (only for GND and Wikidata)

Multilanguage Support

- Cocoda has support for multiple interface languages, currently English and German.
- New languages can be added easily.
- Supports vocabulary data in multiple languages, e.g. Wikidata.
- Currently, most of the available vocabularies in Cocoda are in German.

Additional Services

- KOS Registry (with API)
- Concordance Registry (with API)
- Wikidata/GND Mappings
- Concept Co-Occurrences
- coli-ana (DDC analysis)
  - in development

KOS Registry

About
Partners
Services
KOS Types
KOS Registry
Concordance Registry
Wikidata Mappings
GND Mappings
KOS data registries
(Co-)Occurrences
DDC Analysis
Cocoda Mapping Tool
ccmapper
Publications
Software
Data
Licenses

KOS Registry

This list contains a collection of knowledge organization systems relevant to project coli-conc. The metadata is a subset of BARTOC, also available for download. See also BARTOC dumps for full downloads.

☑ Thesaurus | ☑ Name authority list | ☑ Subject heading scheme | ☑ Classification scheme | expand all | collapse all

title

Mathematics Subject Classification (MSC)

Physics and Astronomy Classification Scheme (PACS)

Classification Digital Library

Classification for music libraries

Klassifikation des Deutschen Patent- und Markenamts (DEKLA)

Systematik für Evangelische Büchereien (SEB)

Nonfiction classification of Catholic public libraries (SKB)

Systematik für Kinder- und Jugendbibliotheken (SKJ)

Systematik der Stadtbibliothek Duisburg (SSD)

Einheitssystematik für Südtiroler Bibliotheken (ESSB)

Gemeinsame Normdatei (GND)

Standard Thesaurus Wirtschaft (STW)

Universal Decimal Classification (UDC)

Dewey Decimal Classification (DDC)
"The Dewey Decimal Classification (DDC) system, devised by library pioneer Melvil Dewey in the 1870s and owned by OCLC since 1988, provides a dynamic structure for the organization of library collections. Now in its 23rd edition, and available in print and Web versions, the DDC is the world's most widely used library classification system."

DDC: 001, 025

Concordance Registry

VZG | Verbundzentrale des GBV (VZG)

COLI CONC

Concordances

We collect concordances and mappings in a public database run with jskos-server. The database can be queried via this web interface and via an API at https://coli-conc.gbv.de/api/.

Concordances | Search Mappings

from x ⇄ to x creator x

from	to	description	creator	date	download	mappings
DDC	RVK	DDC 1000er-Klassen	VZG	2013	JSKOS CSV	2,033
RVK	BK	Recht	ULB Tirol	2013	JSKOS CSV	15,045
DDC	BK	Full DDC up to three digits	VZG		JSKOS CSV	1,223
RVK	DDC	TA-TD Allgemeine Naturwissenschaft	GESIS		JSKOS CSV	68
DDC	BK	Chemie	TUB Hamburg-Harburg		JSKOS CSV	99
DDC	RVK	Mathematik	Philipp Zumstein	2016	JSKOS CSV	40
RVK	BK	Germanistik	ULB Tirol	2013	JSKOS CSV	23,379
RVK	DDC	Chemie und Pharmazie	GESIS		JSKOS CSV	771
RVK	DDC	U Physik	GESIS		JSKOS CSV	1,335
DDC	RVK	Bibliotheks- und Informationswissenschaften	HdM Stuttgart	2009	JSKOS CSV	378
RVK	GND	GND-Indexterme der RVK	UB Regensburg	2015	JSKOS CSV	94,195
DDC	ixtheo	DDC zu IxTheo-Klassifikation	Timotheus Chang Whae Kim		JSKOS CSV	586
RVK	BK	Wirtschaft	ULB Tirol	2013	JSKOS CSV	758
DDC	BK	Politikwissenschaft	SUB Hamburg		JSKOS CSV	655
DDC	RVK	Philosophie und Psychologie (1--)	VZG		JSKOS CSV	1,058

Cocoda Live Demo

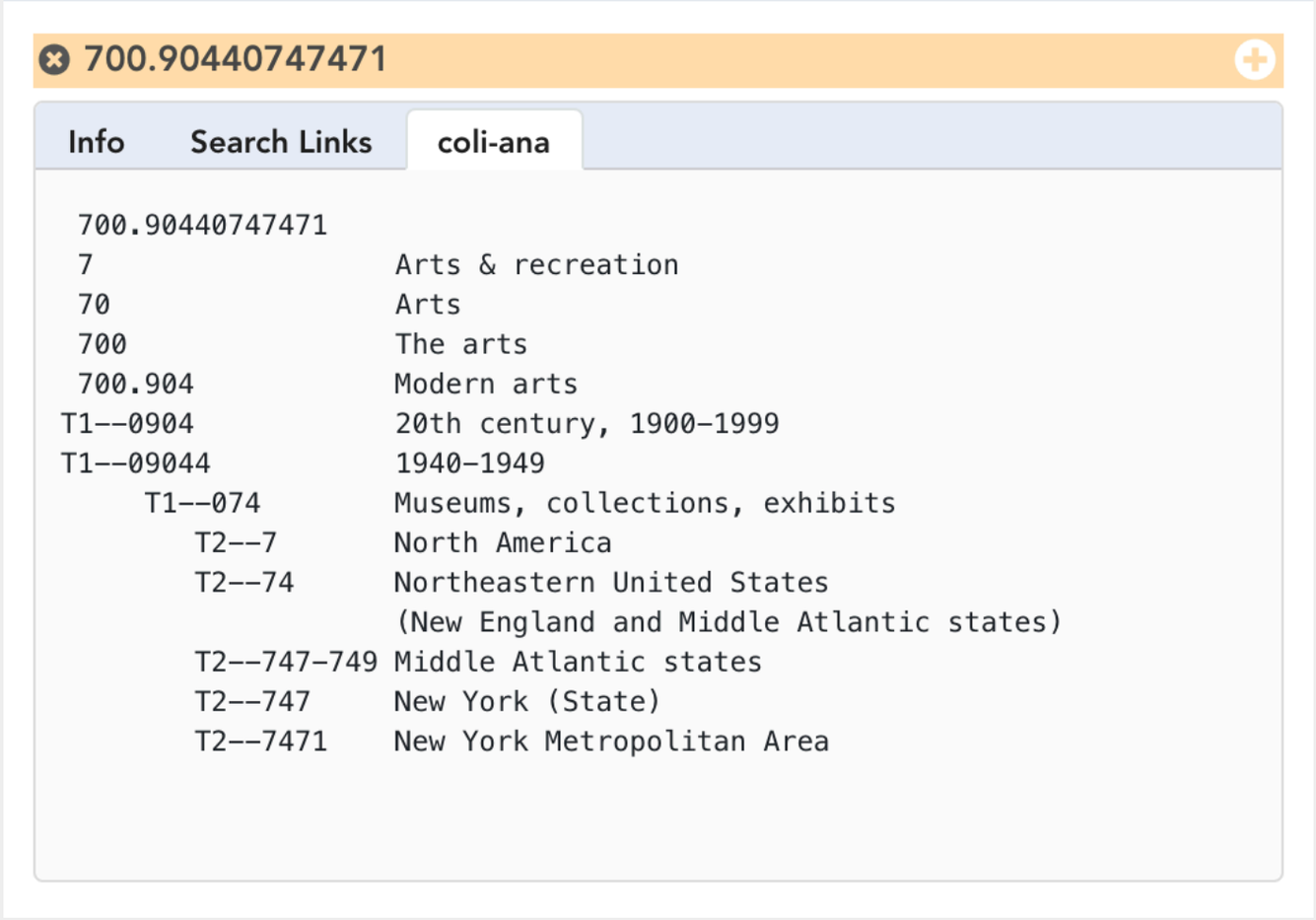
- https://coli-conc.gbv.de/cocoda/app/
- Note: DDC in English is currently not available in the release version of the tool, it is only used here for demo purposes.

Future Plans

- Add more KOS.
- Offer tools to easily convert one's own systems into JSKOS format.
- Improve the tool (usability, tutorial, new features).
  - Already collected many ideas: https://github.com/gbv/cocoda/issues
  - e.g.: coli-ana integration (see next slide)
- More intelligent mapping recommendations.

Example:
coli-ana in Cocoda

- Decomposition of synthesized DDC notations right in Cocoda:



More Information

- Project homepage: https://coli-conc.gbv.de
- GitHub repository for Cocoda: https://github.com/gbv/cocoda
- Cocoda documentation and manual (WIP): https://gbv.github.io/cocoda/

Next Steps

- Ingest further KOS into the mapping tool Cocoda
- Link and Integrate Cocoda with cataloguing tools (WinIBW, Digitaler Assistent,..) as well as library portals
- Expand services to cultural Institutions, such as museums & archives
- Improve Usability
- Modify and apply extant algorithms for more KOS
- Generate usage statistics on KOS and mappings
- Import and include results of DDC-analysis (project coli-ana) for DDC-built notations
- Implement workflows for quality control and collaborative work