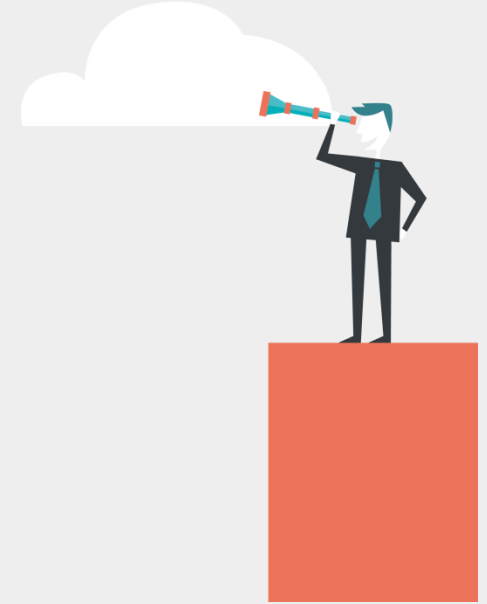
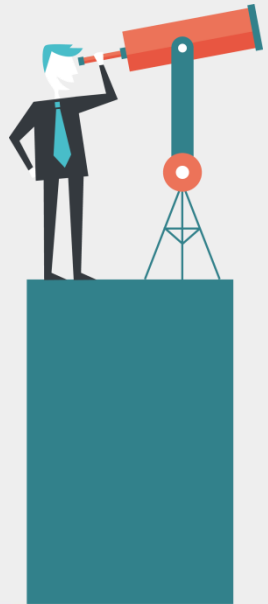


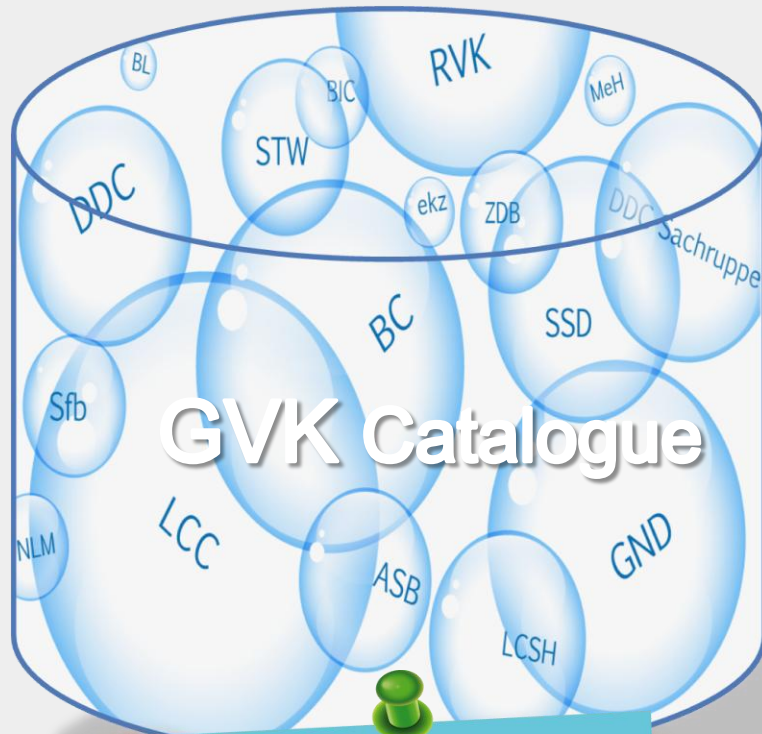
14<sup>th</sup> European Networked Knowledge  
Organization Systems (NKOS) Workshop  
18<sup>th</sup> September 2015, Poznań

# The Cocoda Mapping Tool

U. Balakrishnan, J. Voß  
Verbundzentrale des GBV (VZG)

- Background of Project coli-conc
- Focus of the Project
- Issues related to a DDC-RVK Mapping
- Methods of Mapping
- Course Correction
- Introduction to the Software
  - Demands on the Tool
  - Web Layout
  - Software Concept





## GVK Catalogue

### Aim

Creation of exhaustive concordances between Dewey and other library classification schemes

Classification systems in German speaking regions	
Universal Classification Systems	No. of classes
UDC (Universal Decimal Classification )	ca. 65.000 classes (English version)
DDC (Dewey Decimal Classification)	over 44.000 classes with 10 main classes
RVK (Regensburg Classification)	850.000 classes with 33 main classes
BC (Basic Classification)	2100 classes with 89 main classes
LCC (Library of Congress Classification)	21 main classes
Subject classification	No. of classes
DDC-Sachgruppen der DNB	10 main classes with 94 subclasses
MSC (Mathematics Subject Classification)	87 main classes
PACS (Physics and Astronomy Classification Scheme)	10 main classes
FKDigBib (Subject classification for digital library)	10 main classes
KfM (Classification for music library)	ca. 800 classes
Subject Classification at the Universities	No. of classes
TUM-classification (Science and technology classification of the TU Munic)	52 classes each with 999 notations
Subject classification of the University library Duesseldorf	45 classes
Bremer classification of the State and University library Bremen	ca. 57 main classes
GOK (Goettingen Online Classification)	ca. 33 main classes
Standard-Thesaurus Wirtschaft von der ZWB	6.000 Terms and notations
Subject classification University library Trier	36 main classes
Technical University Dortmund	28 main classes
University library Paderborn	26 main classes
University library Marburg	35 main classes
University library Bonn	24 main classes
University library Heidelberg	22 main classes
Subject classification and nomenclature of individual languages Library of the Institute of General Linguistics at the Uni Münster	23 main classes
Subject Classification at the public libraries	No. of classes
SEB (Scheme for protestant libraries)	
SKB-E (Scheme for catholic public libraries)	
KfKJ (Scheme for children and youth libraries)	Less than 1.000 classes
ASB (General classification for public libraries)	ca. 2.200 classes with 23 main classes
ÖSÖB (Austrian classification for public libraries)	
SfB (Classification for libraries)	ca 14.400 classes with 30 main classes
KAB (Classification for general libraries)	ca. 2.700 classes
SSD (Classification of the city library Duisburg)	
ESSB (Single classification for South Tyrolean)	16 main classes

# Primary Source and Target Schemes : DDC and RVK

## Why RVK?

- wide-spread in Germany
- Local needs are better covered
- Legacy data transfer
- DDC is subject to licence

## RVK

- 850.000 classes
- 33 main classes
- Granularity varies in different subject fields
- Synthesized notations are prebuilt and integrated into the online system

## DDC

- ca. 46.000 classes
- 10 main classes
- not all synthesized notations/numbers are represented in the online system

# Structural difference DDC : RVK

## Notational Building Sequence „Medicine & health”

**614.4273**  
Diseases--humans--  
incidence—United States, ...



**614.42**  
Incidence (from  
Main Schedule)



**73**  
United States  
(from Table T2)

**RU 10585**  
Spreading of diseases  
and influence of the  
geographical  
environment  
(Geomedicine)



**R**  
Geography

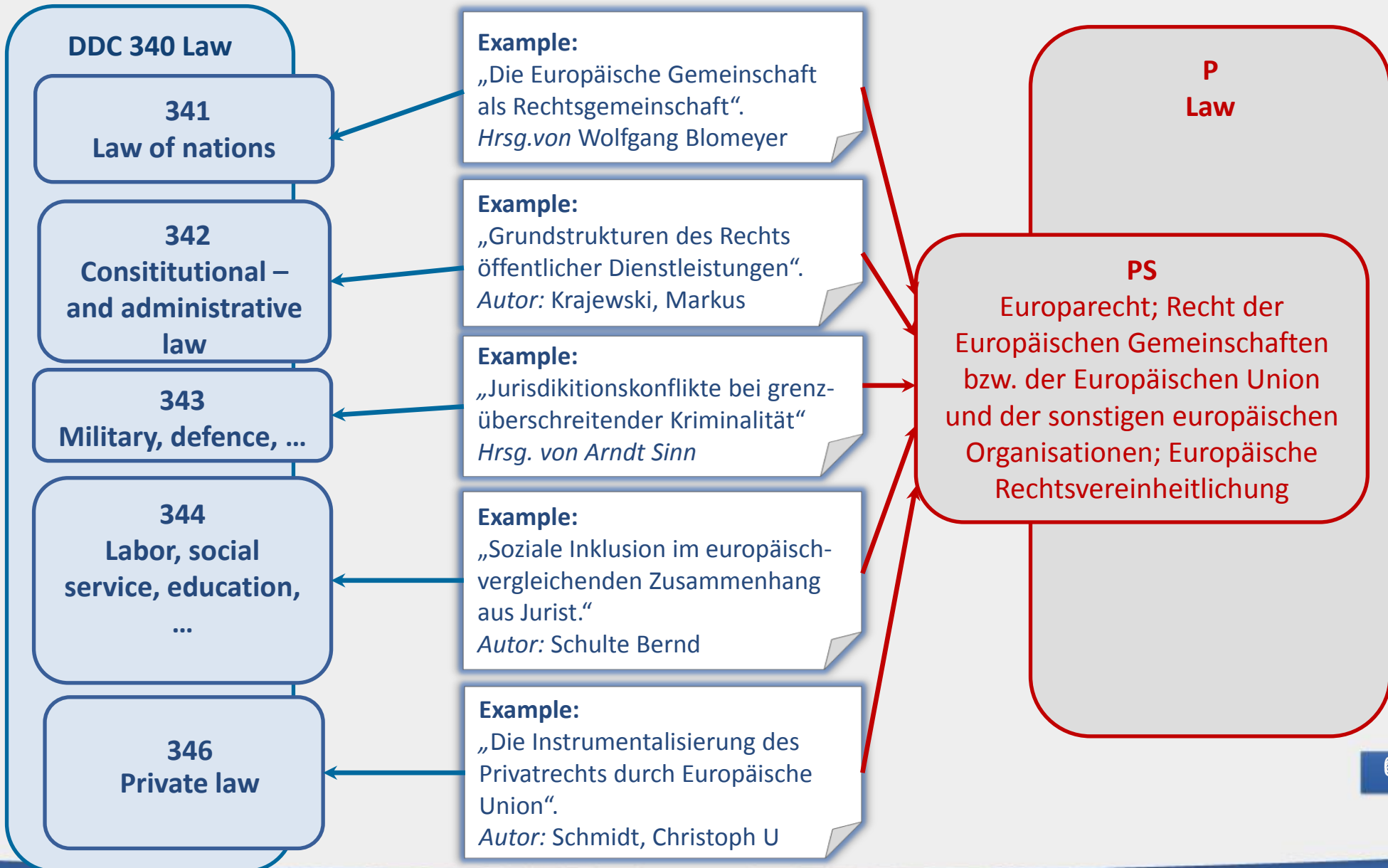


**RU 10000**  
USA



**585**  
Subject index for  
Geography -S1R  
Spreading of  
diseases and  
influence of the  
geographical  
environment  
(Geomedicine)

# Conceptual differences (1)



## Conceptual differences (2)

**615**  
Pharmacology and therapeutics

**DDC 610**  
Medicine and health

**616.024**  
Domestic medicine

**Topic: Pharmacology**

**Example:**  
„Essentials of Pharmacology“  
*Author:* Oldham, Frances, k.  
*RVK notation:* **XI 1400**  
*DDC notation:* **615**

**Topic: Toxicology**

**Example:**  
„Toxicology in the use, misuse and abuse of food, drugs and chemicals“  
*Author:* Chambers, Philipp L.,...  
*RVK notation:* **VT 5308**  
*DDC notation:* **615.9**

**Topic: Drugs**

**Example:**  
„Aspirin and other Salicylates“  
*Author:* Vane, John, R.,...  
*RVK notation:* **XI 4000, VW 5100**  
*DDC notation:* **615.783**

**Topic: Domestic Medicine**

**Example:**  
„Selbstmedikation für die Kitteltasche,...“  
*Author:* Lennecke, Kirsten  
*RVK notation:* **VR 5800**  
*DDC notation:* **616.024**

**RVK: WW-XV**  
**Medicine**

**XI**  
Pharmacology and Toxicology

**RVK: V**  
**Chemistry and Pharmacy**

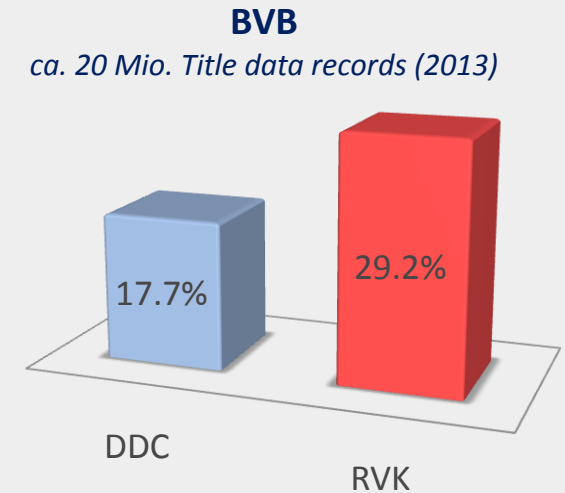
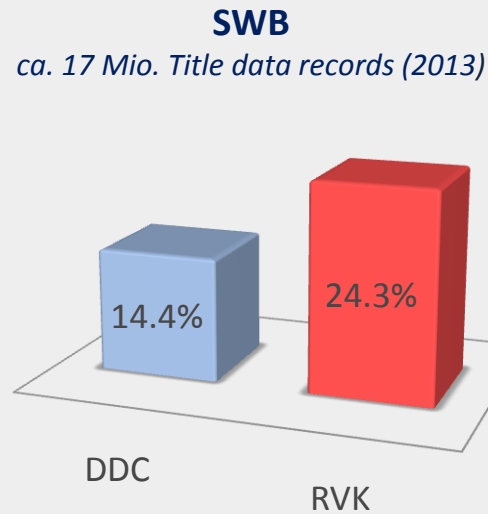
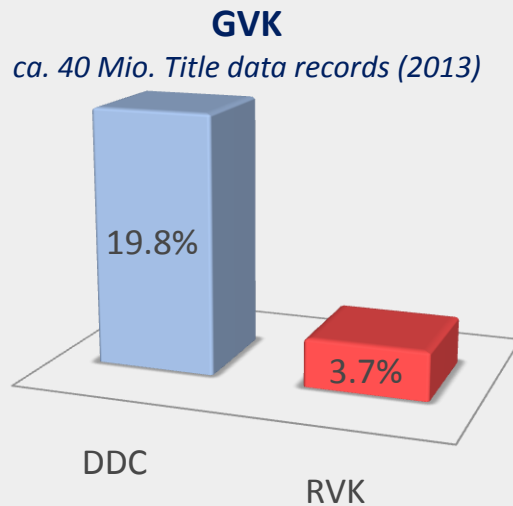
**VT**  
Pharmacology and Toxicology

**VW**  
Pharmaceutical Biology

**VR**  
General Pharmacy



# Percentage of the DDC and RVK in the Union Catalogues

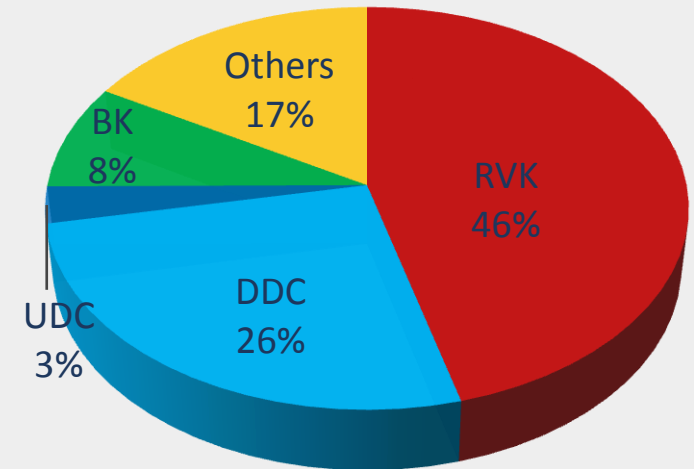




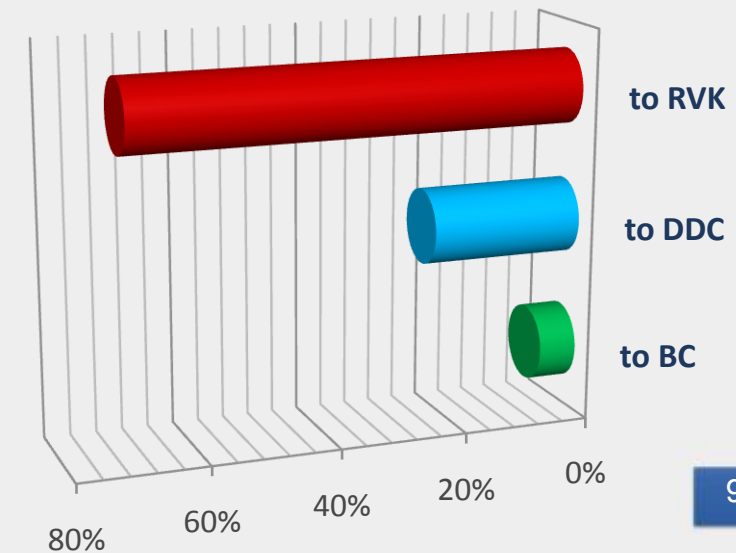


- Current status of DDC-X-concordance
- Field of application and the reasons for the use of DDC
- Methods & Problems in building a DDC-X concordance
- Interest in a DDC - RVK concordance

Existing Mapping works		
Concordance	Subject area	Contact
DDC – BK	Chemistry	TUB TUHH
	Politics	SUB Hamburg
	The thousand classes of the third summary	VZG
DDC – EZB	41 EZB-Fachgruppen	VZG
DDC – RVK	Library- and Information science	HdM Stuttgart
	Social science	UB Greifswald
	Medicine & Health, Law, the thousand classes of the third summary level	VZG
RVK – DDC	Biology, Chemistry, Geology, Paleontology, Physics, Mathematics	GESIS
	Psychology	SLUB Dresden
RVK – BK	German literature, Politics, Law	UB Wien
RVK – MSC	Mathematics	UB Regensburg
RVK – PACS	Physics	UB Regensburg
SWD – DDC		DNB
	Library- and Information science	HdM Stuttgart
SWD-RVK	Library- and Information science	HdM Stuttgart
RVK-BK-MSC-PACS	Mathematics, Physics	ULB Tirol
DDC-MSC-BKL	Mathematics	TIB Hannover

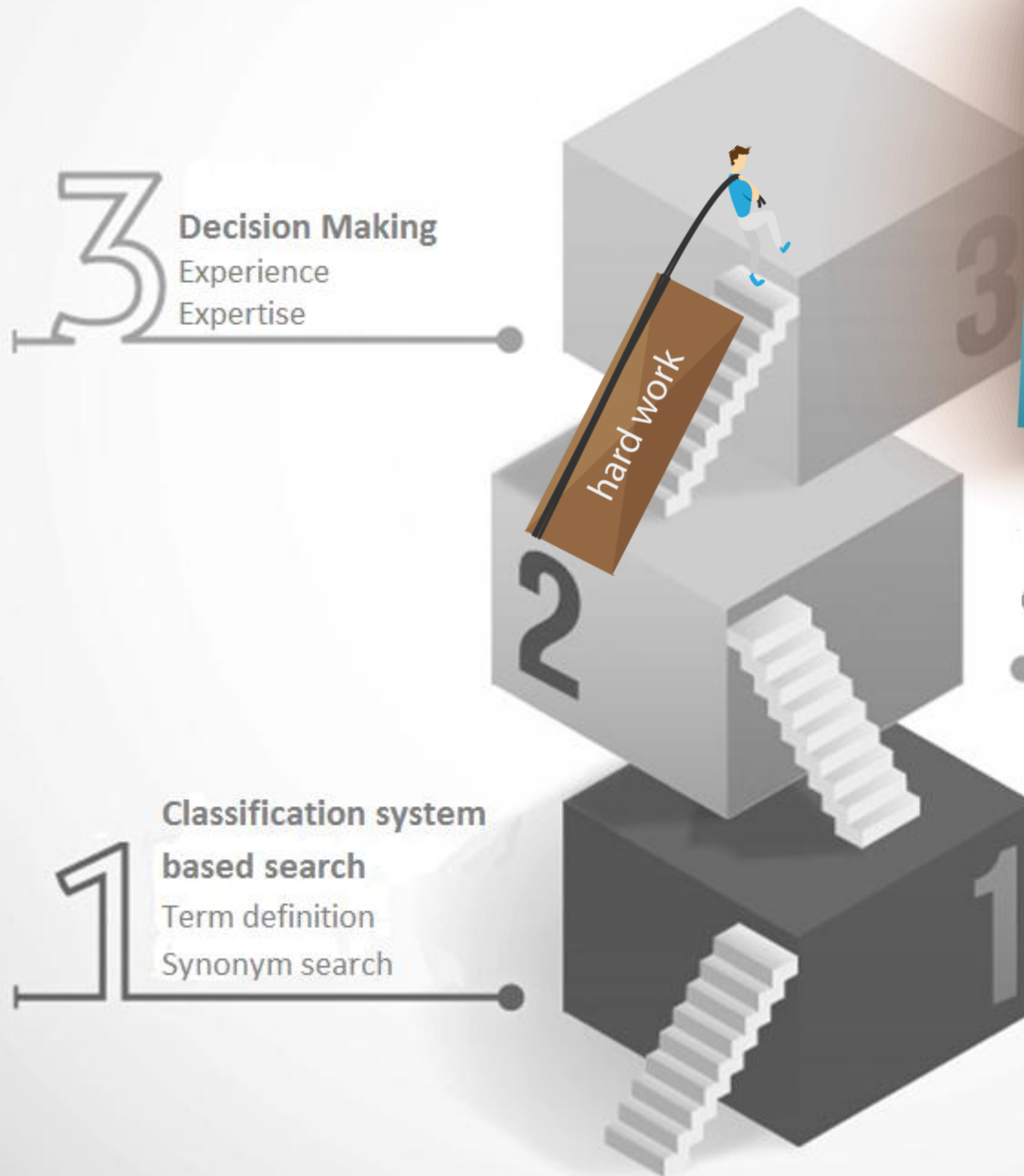


Classification schemes in the libraries that participated in the survey



Shift to other classification schemes

# Mapping Methods



## Work done so far:

### Complete concordance

- DDC - EZB
- DDC - BK for the thousand classes of the third summary of the DDC
- DDC - RVK for the thousand classes of the third summary of the DDC
- DDC - RVK for the DDC subject area Medicine & Health
- DDC - RVK for the DDC subject area „Law“

### Partial Cocordance

- DDC - RVK for the DDC subject area „Philosophy“ (ca. 14% of the current DDC-classes)

# Course Correction

Facilitate exchange and use of concordances and KOS

- Collection of the existing mappings and KOS
- Provision of the above

Enhance the speed of building concordances between library KOS and ease their management

- Develop a mapping tool
- Make the concordances and KOS easily accessible
- Draft algorithms for automatic generation of mapping candidates

Improve the quality of the concordances

- Develop and implement measures for quality control
- Involve and expand the user groups



# Demands on the tool

Integration of Data from different sources

Allow validation and storage of data

Multi-user web based open source tool

Easy access to and exchange of information  
Serve as collaboration platform

Presentation of Data and mapping candidates on a single screen

Clear overview of the context of the selected term through display of

- the hierarchical structure of the classes
- scope notes
- Register Index Entries
- linked vocabularies and
- synonym suggestions

Mapping suggestions through

- evaluation of the co-occurrences of assigned notations/terms in the title data records
- automatic generation of mappings
- integration of the concordance database
- inclusion of the results of a manual mapping



# Web Interface Cocoda Prototype

Cocoda

[Log in](#)

Source Scheme: DDC ▾

Search Options

Search by: **Term** Notation

Search by terms (typeahead)

none

DDC

GND

RVK

Wikidata

Top Concepts >

Source Scheme

Active Mapping

No concepts selected for mapping

Mappings

Target Scheme: RVK ▾

Search Options

Search by: **Term** Notation

Search by terms (typeahead)

none

DDC

GND

RVK

Wikidata

Top Concepts >

Target Scheme

# Web Interface Cocoda Prototype

Cocoda

[Log in](#)

Source Scheme: DDC ▾

Search Options ▾

Search by: **Term** Notation

Leukozyten (Weiße Blutkörperchen)

612.112 Leukozyten (Weiße Blutkörperchen)

↳ Blut  
↳ Biochemie  
↳ Biophysik  
↳ Anzahl und Auszählung

Leukozyten--Humanphysiologie

Weiße Blutkörperchen--Humanphysiologie

Map ➔ Look up database all ▾

Top Concepts ▾

- 0 Informatik, Informationswissenschaft & allgemeine Werke ⓘ
- 1 Philosophie & Psychologie ⓘ
- 2 Religion ⓘ
- 3 Sozialwissenschaften ⓘ
- 4 Sprache ⓘ
- 5 Naturwissenschaften ⓘ
- 6 Technik, Medizin, angewandte Wissenschaften ⓘ
- 7 Künste und Unterhaltung ⓘ
- 8 Literatur ⓘ

Active Mapping

No concepts selected for mapping

Target Scheme: RVK ▾

Search Options ▾

Search by: **Term** Notation

Search by terms (typeahead Wikidata)

none  
DDC  
GND  
RVK  
Wikidata

Top Concepts ➔

# Web Interface Cocoda Prototype

Cocoda

Log out

Source Scheme: DDC ▾

## Search Options ▾

Search by: **Term** Notation

Leukozyten (Weiße Blutkörperchen)

612.112 Leukozyten (Weiße Blutkörperchen)

- ↳ Blut
- ↳ Biochemie
- ↳ Biophysik
- ↳ Anzahl und Auszählung

Leukozyten--Humanphysiologie

Weiße Blutkörperchen--Humanphysiologie

Map ➔

Look up database

all ▾

## Top Concepts ▾

- 0 Informatik, Informationswissenschaft & allgemeine Werke ⓘ
- 1 Philosophie & Psychologie ⓘ
- 2 Religion ⓘ
- 3 Sozialwissenschaften ⓘ
- 4 Sprache ⓘ
- 5 Naturwissenschaften ⓘ
- 6 Technik, Medizin, angewandte Wissenschaften ⓘ
- 7 Künste und Unterhaltung ⓘ
- 8 Literatur ⓘ

## Active Mapping

No concepts selected for mapping

Target Scheme: RVK ▾

## Search Options ▾

Search by: **Term** Notation

Blutkörperchen (Erythrozyt, Leukozyt), Häm

WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin

- ↳ Blut und Blutbestandteile
- ↳ Allgemeines
- ↳ Einzelne biologische Disziplinen
- ↳ Systematische Spezifizierung

Add +

Replace all ➔

## Top Concepts ▾

- A Allgemeines ⓘ
- B Theologie und Religionswissenschaften ⓘ
- CA - CK Philosophie ⓘ
- CL - CZ Psychologie ⓘ
- D Pädagogik ⓘ
- E Allgemeine und vergleichende Sprach- und Literaturwissenschaft. Indogermanistik. Außereuropäische Sprachen und Literaturen ⓘ
- F Klassische Philologie. Byzantinistik. Mittellateinische und Neugriechische Philologie. Neulatein ⓘ



# Web Interface Cocoda Prototype

Cocoda

Log out

Source Scheme: DDC

## Search Options

Search by: **Term** **Notation**

612.112

612.112 Leukozyten (Weiße Blutkörperchen)

↳ Blut  
↳ Biochemie  
↳ Biophysik  
↳ Anzahl und Auszählung

Leukozyten--Humanphysiologie

Weißer Blutkörperchen--Humanphysiologie

Map

Look up database

all

## Top Concepts

- 0 Informatik, Informationswissenschaft & allgemeine Werke
- 1 Philosophie & Psychologie
- 2 Religion
- 3 Sozialwissenschaften
- 4 Sprache
- 5 Naturwissenschaften
- 6 Technik, Medizin, angewandte Wissenschaften
- 7 Künste und Unterhaltung
- 8 Literatur

Active Mapping

612.112



WW 8840 - WW 8879

WW 8720 - WW 8999

Clear all

Show mapping object

## Mapping Candidates

### Catalog Occurrences

Used notation: 612.112  
Used database: GVK/SWB  
Results (total) for 612.112: 42  
Corresponding notations in RVK:

Notation	Hits	% of total
WW 8840	22	52.4 %
YC 2500 - YC 2599	11	26.2 %
WF 9895	8	19.0 %
XG 6700 - XG 6728	1	2.4 %

### Suggested Target Concepts

WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin  
WW 8720 - WW 8999 Blut und Blutbestandteile  
WW 8845 - WW 8879 Systematische Spezifizierung

## Concordance database

Target Scheme	Concept	Creator	Date	Relevance
RVK	Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	VZG	2012	
GND	Leukozyt	CrissCross	2010	high (0.8)
GND	Alkalische Leukozytenphosphatase Blutlymphozyt Granulozyt Leukozytenadhäsion	CrissCross	2010	medium (0.5)

Target Scheme: RVK

## Search Options

Search by: **Term** **Notation**

ww 8840

WW 8840 Allgemeines

↑ Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin

Add

Replace all

Add

Replace all

## Top Concepts

- A Allgemeines
- B Theologie und Religionswissenschaften
- CA - CK Philosophie
- CL - CZ Psychologie
- D Pädagogik
- E Allgemeine und vergleichende Sprach- und Literaturwissenschaft. Indogermanistik. Außereuropäische Sprachen und Literaturen
- F Klassische Philologie. Byzantinistik. Mittellateinische und Neugriechische Philologie. Neulatein

# Web Interface Cocoda Prototype

Cocoda

Log out

Source Scheme: DDC

## Search Options

Search by: **Term** **Notation**

612.112

612.112 Leukozyten (Weiße Blutkörperchen)

↳ Blut  
↳ Biochemie  
↳ Biophysik  
↳ Anzahl und Auszählung

Leukozyten--Humanphysiologie

Weißer Blutkörperchen--Humanphysiologie

Map

Look up database

all

## Top Concepts

- 0 Informatik, Informationswissenschaft & allgemeine Werke
- 1 Philosophie & Psychologie
- 2 Religion
- 3 Sozialwissenschaften
- 4 Sprache
- 5 Naturwissenschaften
- 6 Technik, Medizin, angewandte Wissenschaften
- 7 Künste und Unterhaltung
- 8 Literatur

## Active Mapping

612.112



WW 8840 - WW 8879

WW 8720 - WW 8999

Clear all

Create download link

Save to Database

## Mapping Candidates

### Catalog Occurrences

Used notation: 612.112  
Used database: GVK/SWB  
Results (total) for 612.112: 42  
Corresponding notations in RVK:

Notation	Hits	% of total
WW 8840	22	52.4 %
YC 2500 - YC 2599	11	26.2 %
WF 9895	8	19.0 %
XG 6700 - XG 6728	1	2.4 %

### Suggested Target Concepts

WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin  
WW 8720 - WW 8999 Blut und Blutbestandteile  
WW 8845 - WW 8879 Systematische Spezifizierung

## Concordance database

Target Scheme	Concept	Creator	Date	Relevance
RVK	Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	VZG	2012	
GND	Leukozyt	CrissCross	2010	high (0.8)
GND	Alkalische Leukozytenphosphatase Blutlymphozyt Granulozyt Leukozytenadhäsion	CrissCross	2010	medium (0.5)

Target Scheme: RVK

## Search Options

Search by: **Term** **Notation**

ww 8840

WW 8840 Allgemeines

↑ Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin

Add

Replace all

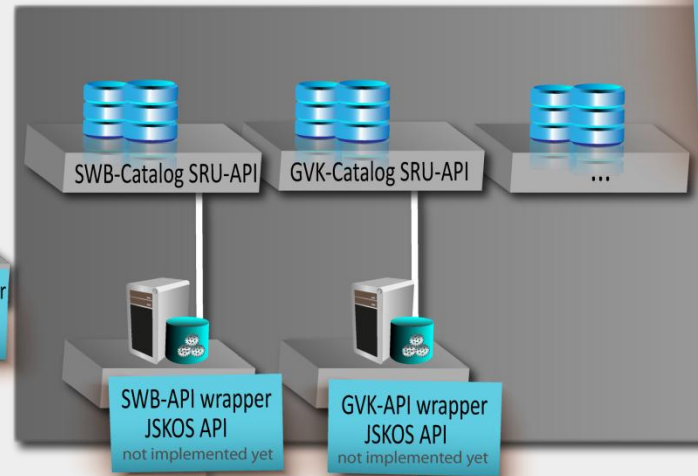
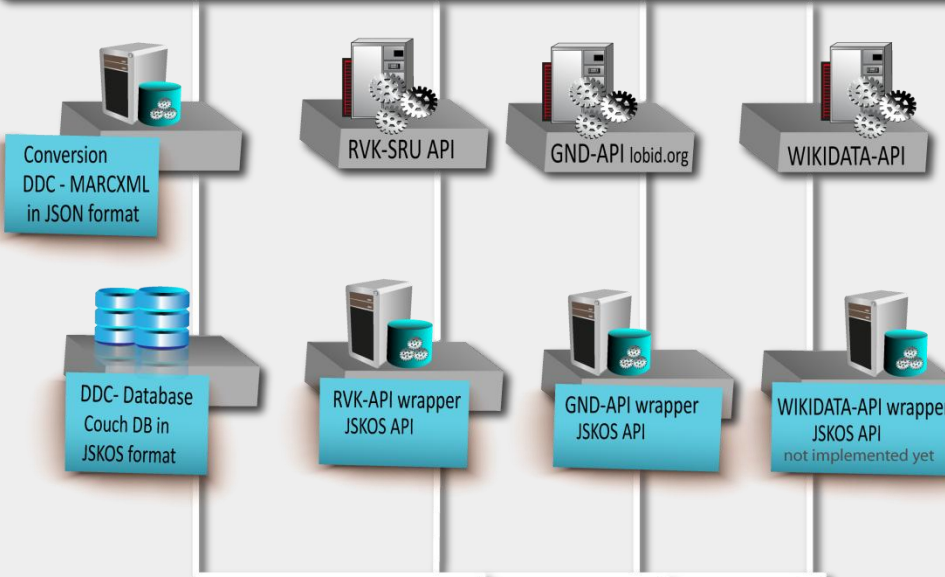
Add

Replace all

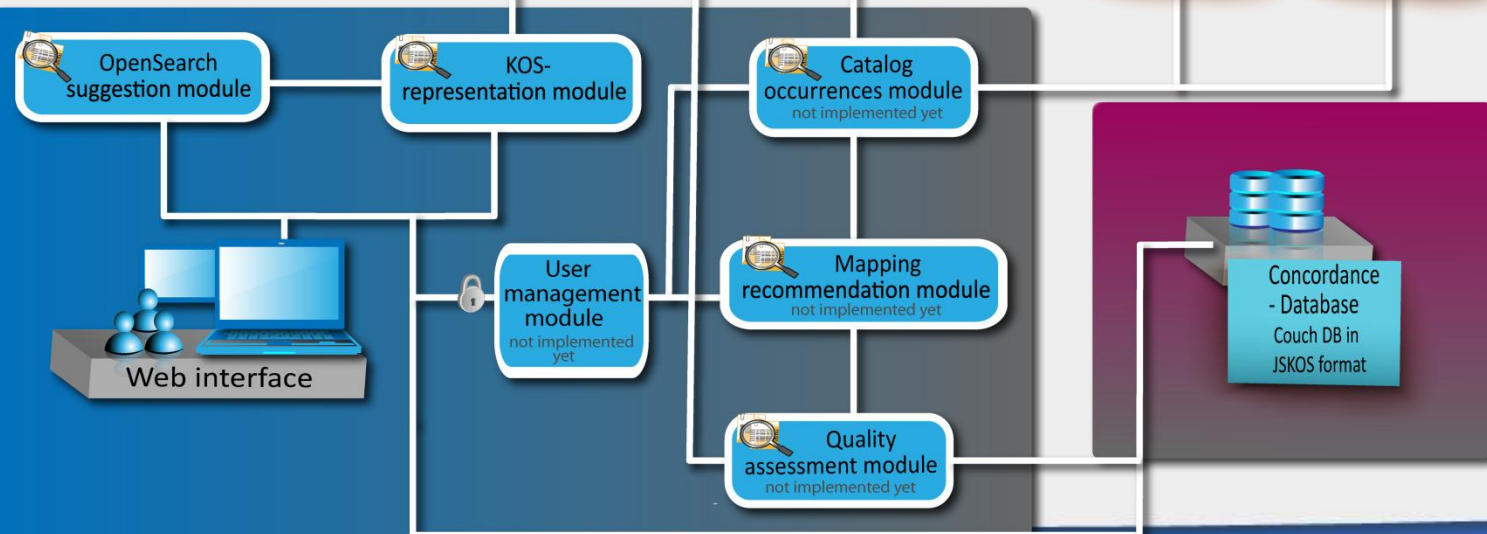
## Top Concepts

- A Allgemeines
- B Theologie und Religionswissenschaften
- CA - CK Philosophie
- CL - CZ Psychologie
- D Pädagogik
- E Allgemeine und vergleichende Sprach- und Literaturwissenschaft. Indogermanistik. Außereuropäische Sprachen und Literaturen
- F Klassische Philologie. Byzantinistik. Mittellateinische und Neugriechische Philologie. Neulatein

# Cocoda – Software Concept



- Format requirements**
- Compatible with SKOS
  - Represent both KOS and its mappings
  - Simple and easy to use in web application



- Based on SKOS extended with mapping objects
- Application of JSON-LD with consistent fields  
⇒ simple JSON with optional mapping to RDF
- Also supports existence statements
  - $\nexists$  : negation (e.g. has no related concepts)
  - $\forall$  : completeness (e.g. has narrower concepts)
- Open Standardization of JSKOS, feedback appreciated: <https://gbv.github.io/jskos/>

# JSKOS Example: Concept

```
{
  "@context": "http://gbv.github.io/jskos/context.json",
  "type": ["http://www.w3.org/2004/02/skos/core#Concept"],
  "inScheme": ["http://dewey.info/scheme/edition/e22/"],
  "uri": "http://dewey.info/class/612.112/e22/",
  "notation": ["612.112"],
  "prefLabel": {
    "en": "White corpuscles (Leukocytes)",
    "de": "Leukozyten (Weiße Blutkörperchen)"
  },
  "narrower": [ null ],
  "broader": [ {
    "notation": ["612.11"],
    "uri": "http://dewey.info/class/612.11/e22/"
  } ],
  "related": null
}
```

Null as array element: existence

Null value: negation

Missing fields: maybe (Open World Assumption)

- Based on JSKOS
- Used in Cocoda with Web Components
- Simple URL queries, possible mapping to LOD-URLs
  - `/mappings?from=DDC&to=RVK&notation=612`
  - `/concepts?notation=612`  $\Leftrightarrow$  `/concepts/612`
- Wrapping other APIs
- Open preview of specification and prototype
  - <https://gbv.github.io/jskos-api/>
  - <https://github.com/gbv/cocoda-db>
  - <http://coli-conc.gbv.de/concordances/>

## JSKOS Example: Mapping (*current draft*)

```
{  
  "from": {  
    "inScheme": ["http://dewey.info/scheme/edition/e22/"],  
    "conceptSet": [ {  
      "uri": "http://dewey.info/class/612.112/e22/",  
      "notation": ["612.112"]  
    } ]  
  },  
  "to": {  
    "inScheme": ["http://d-nb.info/gnd/7749153-1"],  
    "conceptSet": [ {  
      "uri": "http://d-nb.info/gnd/4074195-3",  
      "preflabel": { "de": "Leukozyt" }  
    } ]  
  },  
  "mappingType": "closeMatch",  
  "creator": "VZG"  
}
```



# Thank You!

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

Vectors slide no.11: © Vallepu – fotolia.com <https://de.fotolia.com/>  
Vectors slide no.9: © NLshop– fotolia.com <https://de.fotolia.com/>  
Vectors slide no. 2,3,4,10,12,18: designed by Freepik.com <http://www.freepik.com>  
Thanks to Jana Agne for creating the table at the slide no.3