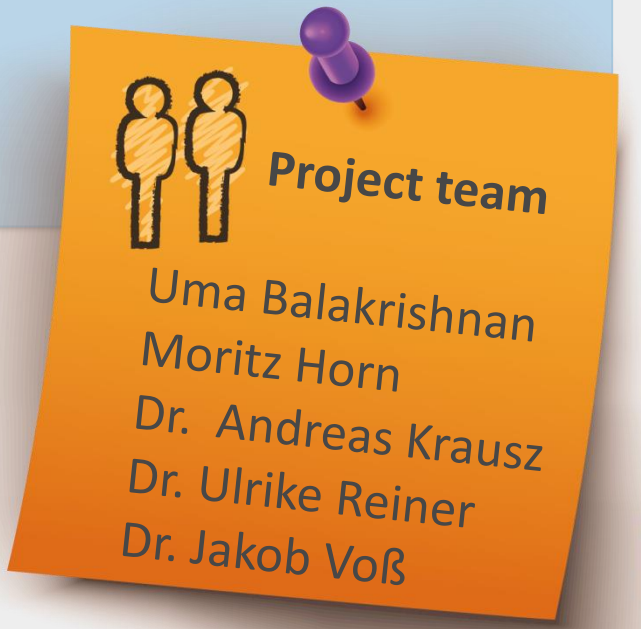
EDUG – Mapping Working Group Meeting
15. April 2015, Naples

Mapping Tool „Colibri Concordance Database (Cocoda)“

Uma Balakrishnan
Dr. Andreas Krausz

Project „coli-conc“

- To build exhaustive concordances between DDC and other library classification systems, in particular with Regensburg Classification
- Development of an infrastructure to facilitate creation, management and exchange of concordances between library knowledge organization systems



Classification systems in use in German speaking countries

Landscape of library classification systems in Germany, Austria and Switzerland

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 classess
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 Classifications at the Universities	**No. of classes**
TUM-classification (Science and technologyl 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 Classifications at the Universities	**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 libarries)	
SfB (Classification for libraries)	ca 14.400 classes with 30 main classes
KAB (Classification for general libraries)	ca. 2.700 classes
SSD (Classfication of the city library Duisburg)	
ESSB (Single classfication for South Tyrolean)	16 main classes

Primary Source and Target Schemes : DDC and RVK

RVK

- 850.000 classes
- 33 main classes
- Granularity varies in different subject fields
- Synthetic notations are pre- built and integrated into the online system

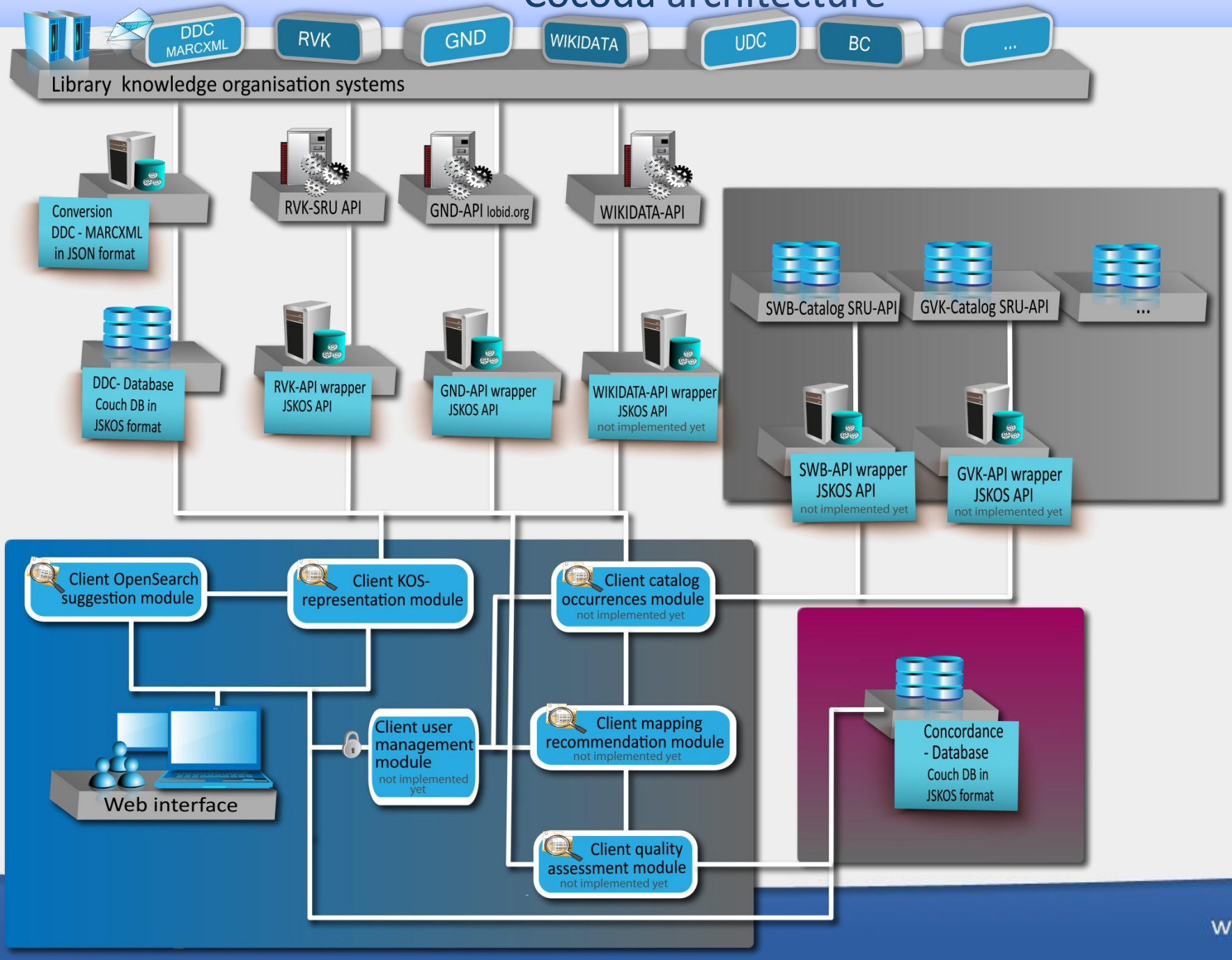
DDC

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

Other reasons:

- wide-spread in Germany
- Local needs are better covered
- Legacy data transfer
- The quality of DDC Data (GNL) not persuasive
- DDC is subjected to licence
- RVK is available as authority data

Cocoda architecture



Web interface „Cocoda“ prototype

Cocoda

Log in

Source Scheme: DDC

Search Options

Search by: Term Notation

612.112

612.112

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

Active Mapping

612.112 → WW 8840 - WW 8879, WW 8720 - WW 8999

Clear all | Create download link

Concordance database

Target Scheme	Notation	Creator	Date	Relevance
RVK	Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	VZG		
GND	Leukozyt	CrissCross		high (0.8)
GND	Alkalische Leukozytenphosphatase Blutlymphozyt Granulozyt Leukozytenadhäsion Leukozytenphosphatase Monozyt	CrissCross		medium (0.5)
GND	Leukozytenintegrine	CrissCross		low (0.2)

Catalog Occurrences

Used notation: 612.112
Used database: GVK
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

Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin
Blut und Blutbestandteile
Systematische Spezifizierung

Target Scheme: RVK

Search Options

WW 8720 - WW 8999 **Blut und Blutbestandteile**

Blut / Bestandteile, Blut

↖ Körperflüssigkeiten und deren Systeme einschließlich deren Bildungsorgane
↘ Taxonomischer Schlüssel
↘ Blutplasma, Blutserum
↘ Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin
↘ Blutgerinnung (Thrombozyt)

Add | Replace all

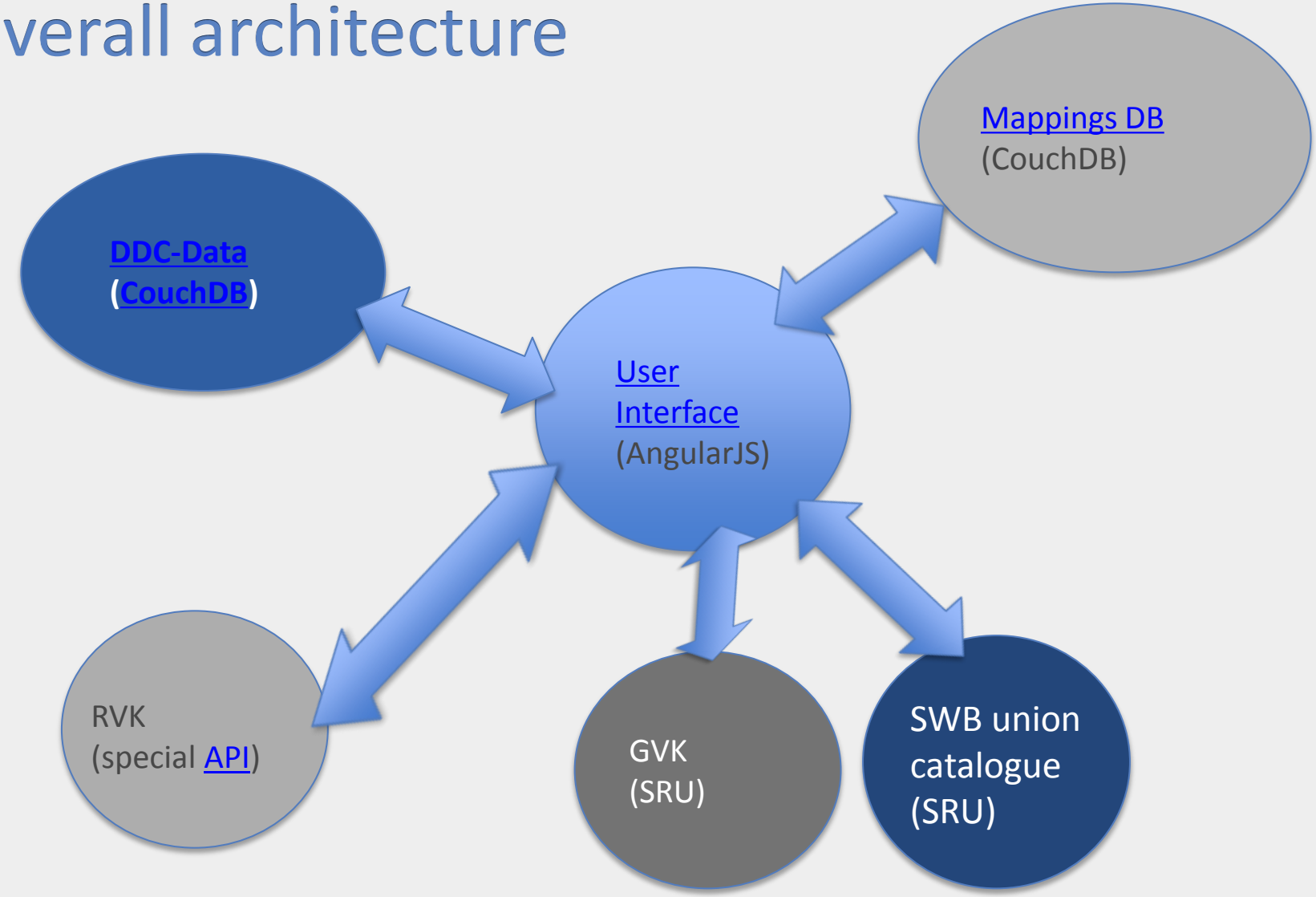
Top Concepts

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

Part 2

- Overall architecture
- Data model and technical background

Overall architecture



Why CouchDB?

Everything in CouchDB is HTTP. You can serve a CouchApp from database itself using views. You don’t need additional tools running like Apache.

continue

Why JSKOS?

- closed world statements
- order of multiple notations
- order of multiple inScheme statements

back

DDC DB data model (snippet)


```
{
   "uri": "http://dewey.info/class/641.5/e23/",
   "type": ["http://www.w3.org/2004/02/skos/core#Concept"],
   "notation": ["641.5"],
   "inScheme": ["http://dewey.info/edition/e23/"],
   "prefLabel": {
      "en": "Cooking"
   },
   "broader": [{
         "uri": "http://dewey.info/class/641/e23/",
         "notation": ["641"],
         "prefLabel": {
            "en": "Food and drink"}
      }],
   "narrower": [{
         "uri": "http://dewey.info/class/641.502/e23/",
         "notation": ["641.502"],
         …
```


back

Mapping DB data model


```
{
"creator": "CrissCross",
"mappingType": "exactMatch",
        "from": {
                "inScheme": [{"notation":"DDC"}],
                "conceptSet": [
                    {"notation": ["612.112"],
                      "prefLabel": { "de": "Leukozyten (Weiße Blutkörperchen)"} }]
                        },
        "to": {
                 "inScheme": [{"notation":"GND"}],
                 "conceptSet": [ {
                    "uri": "http://d-nb.info/gnd/4074195-3",
                    "prefLabel": { "de": "Leukozyt" },
                    "notation":["4074195-3"] } ],
                     "coordination": "OR"}
}
```


continue

CouchDB queries

DDC DB:

http://esx-151.gbv.de/?key=Jungunternehmer&view=by_name

http://esx-151.gbv.de/?key=Jungunternehmer

http://esx-151.gbv.de/?key=623.74&view=by_num

Mappings DB:

http://esx-151.gbv.de/?key=4055849-6&db=mappings&view=toNotation&exact=true

http://esx-151.gbv.de/?key=330.124&db=mappings&view=fromNotation&exact=true

http://esx-151.gbv.de/?key=DDC&db=mappings&view=fromSchemeNotation&exact=true

http://esx-151.gbv.de/?key=GND&db=mappings&view=toSchemeNotation&exact=true

back

CouchDB query (=view) definiton example

```
{
  "_id": "_design/get",
  "_rev": "2-1b04d657423514aaa349a977d2af165f",
  "views": {
    "by_from": {
      "map": "function(doc){ if(doc.from.notation){ emit(doc.from.notation, [doc.creator, doc.timestamp, doc.from, doc.to]); } }"
    },
    "by_to": {
      "map": "function(doc){ if(doc.to){ for (var i = 0; i < doc.to.length; i++){ emit(doc.to[i].notation, [doc.creator, doc.timestamp, doc.from, doc.to]); } } }"
    }
  }
}
```


continue

RVK-API query examples

http://rvk.uni-regensburg.de/api/json/children/UT 5000 - UT 9800

http://rvk.uni-regensburg.de/api/xml/ancestors/PH 3570

back

Thank you!