Mapping Project coli-conc

Progress, learning & next steps

U. Balakrishnan, J. Agne, J. Voß
Content

- Aim & Project Start
- Partners
- coli-conc-Key Objectives & Approaches
- Work Packages
- Survey
- JSKOS Dataformat
- Literature
Aim & DFG Project start

- **Aim**: Infrastructure to aid semi-automatic creation and management of mappings between library KOS and enable a facile access to the same.

- **Project start**: 15. December 2015
Partners

Freie Universität Berlin

Universität Regensburg

Universität St. Gallen

ZBW – Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

ulb. – Universitäts- und Landesbibliothek Tirol

gesis

Wikimedia Deutschland

VZG
coll-conc-Key Objectives & Approaches

Effective Creation and Management of Mappings
- Development of a Mapping tool
- Provision of KOS and their Mappings
- Creation or adoption of concordance algorithms

Improvement of the Quality of the Mappings
- Development and Implementation of measures for Quality Assessment
- Involvement of Experts and User groups

Facilitate Use and Exchange of KOS and their Mappings
- Collection of KOS and their existing Mappings
- Provision of KOS and their Mappings
- Documentation and Outreach
Work Packages

Evaluation and Study
- Current status of KOS and their mappings
- Existing software for KOS management & creation of mappings

Development of a Web Interface
- Representation & navigation of KOS
- Compilation and processing of mappings
- Display and workspace on auto-generated mapping candidates
- Usability-tests of the web interface

Provision
- Storage and provision of KOS (KOS-DB)
- Management of concordances (KK-DB)
- Integration with the existing KOS-software, gateways and cataloging tools
- Testing and documentation of know-how for installation of the application in cloud services
- Publishing mappings as linked open data

Concordance Algorithm

Quality Assessment
- Evaluation of existing standards and know-how for quality assessment
- Conception and testing of crowd-sourcing strategy
- Development of a server component to assess and provide quality measurements

Outreach
- Documentation on KOS and mapping tool
- Hold Workshops with experts and users
- Conception and implementation of authentication and authorization techniques
- Conception of crowd sourcing approach
survey
General information

• Aim
  • Evaluation of ...
  • ... the distribution of KOS
  • ... the software of KOS
  • ... the concordance projects

• Distribution
  • INTEBIB, KIM, ÖBIB,
    FAGEI, EGSE, RVK-Mailingliste
  • Facebook, Twitter
  • hardcopy (library congress Leipzig 2016)

• Tool: qset

• Timeline: 08.02.2016 – 08.04.2016
Participants
over 200
Questionnaires

KOS:
• Distribution of KOS
• Software for KOS management
• Reasons for choosing KOS
• Purposes of KOS
• How subject indexing is done
• KOS system changes

Mapping Projects:
• Mapping projects
• Building concordances
• KOS
• Subjects
• Types of concordances
• Mapping relationships
• Mapping types
• Mapping tool
Results Part I - KOS
KOS in Germany I – Survey 2016

- Classification systems:
  - 36% RVK
  - 21% DDC
  - 10% BC
  - 33% others including: Houseown UDC LoC

- Subject headings & thesauri:
  - 68% GND
  - 19% others
  - 4% LCSH
  - 3% MeSH
  - 5% STW

2016
### KOS in Germany II – Updated List of KOS

**VZG Terminologie Webapplication**

<table>
<thead>
<tr>
<th>Universal Classification systems</th>
<th>No. of classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC (Library of Congress Classification)</td>
<td>65,000 classes</td>
</tr>
<tr>
<td>DDC (Deutsche Decimal Classification)</td>
<td>16 main classes and over 44,000 classes in total</td>
</tr>
<tr>
<td>RVK (Regensburg Classification)</td>
<td>34 main classes and over 855,000 classes in total</td>
</tr>
<tr>
<td>LC (Basic Classification)</td>
<td>89 main classes and 2,100 subclasses in total</td>
</tr>
<tr>
<td>CDU (German Library Classification)</td>
<td>21 main classes</td>
</tr>
<tr>
<td>OSL (Austrian Public Library Classification)</td>
<td>36 main classes</td>
</tr>
<tr>
<td>SFR (Classification of Public Libraries)</td>
<td>10 main classes and over 2,200 subclasses</td>
</tr>
<tr>
<td>KAS (Classification of General Libraries)</td>
<td>2,700 classes</td>
</tr>
</tbody>
</table>

#### Subject classification

- **DDC-Sachgruppen der UNI**
  - 10 main classes with 94 subclasses

- **MSC (Mathematics Subject Classification)**
  - 87 main classes

- **PACS (Physics and Astronomy Classification Scheme)**
  - 10 main classes

- **FDS 98 (Subject Classification for Digital Libraries)**
  - 10 main classes

- **RIV (Classification of Music libraries)**
  - ca. 800 classes

- **Classification of the German Patent and Trademark Office**

- **SEB (Classification of Protestant Libraries)**

- **Non-fictional Classification of Catholic Public Libraries**
  - < 1,000 classes

- **SSO (Classification of the Public Library in Stade)**
  - 16 main classes

- **Systematic of the library of the State University of Hamburg**
  - 16 main classes

- **Classification of Social Studies - GESIS**
  - 16 main classes

- **Classification of the finance court Cologne**
  - 16 main classes

- **FIV subject classification**
  - 16 main classes

- **FIV classification of the region**
  - 16 main classes

- **Psychoanalytic classification**
  - 16 main classes

- **Journal of Economic Literature Classification System**
  - 16 main classes

- **OSUAR (Online Systematic for Maps)**
  - 16 main classes

- **Classification of the institute of contemporary history**
  - Ca. 34 main classes

- **SMWW (Systematik Musikbibliographischer Bibliothek)**
  - ca. 34 main classes

- **FDC (GFDC Global forest decimal classification)**
  - ca. 34 main classes

- **InTheo (Index theologicus)**
  - ca. 34 main classes

#### Subject Classifications at the Universities

- **TUM-Classification (Science and technology classification of the TU Munich)**
  - 52 classes each with 999 notations

- **Subject classification of the University Library in Dusseldorf**
  - 45 classes

- **Bremen classification of the State and University Library Bremen**
  - ca. 57 main classes

- **GOX (Goettingen Online Classification)**
  - ca. 33 main classes

- **Systematis des Göttinger Band-Rekatalogis**
  - 36 main classes

- **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 Hanover**
  - 24 main classes

- **University Library Heidelberg**
  - 22 main classes

- **Subject classification and nomenclature of individual libraries**
  - Institute of General Linguistics at the University of Münster
  - 23 main classes

#### Thesauri

- **Standard-Thesaurus Wirtschaft von der ZWBO**
  - 6,000 Terms and notations

- **APA (Thesaurus of the American Psychological Association)**

- **MDH (Medical Subject Headings)**

- **CAB thesaurus (Thesaurus of the Centre for Agriculture and Biosciences)**

- **LCM (Library of Congress Subject Headings)**

- **Rameau (Répertoire d'autorité matière encyclopédique et alphabétique unifiée = Encyclopedic directory and unified alphabetical subject authority)**

- **RISAX (Rules for the keyword index)**

- **ACM (Computing Classification System)**

---

[http://coli-conc.gbv.de/terminologies/]
KOS in Germany III – Surveys 1999 & 2011

1999
- RVK: 46%
- DDC: 26%
- Others: 17%
- BC: 8%
- UDC, DDC: 5%
- GHBS NRW: 8%
- RVK: 33 % houseown

2011
- RVK: 46%
- DDC: 26%
- Others: 17%
- BC: 8%
- UDC: 3%
A look at Dewey

- External data transfer
- Free Online Access
- Free of license
- Improve retrieval
- Usability of legacy data
- Other reasons
- Management-related reasons

- 1947
- 1996 (2)
- 2000
- 2002
- 2005
- 2009
- 2010
- 2011
- 2014

Home systems
- RVK (15)
- DDC (10)
- BC (5)

1999
- RVK
- DDC
- BC

2011
- RVK
- DDC
- BC

2016
- RVK
- DDC
- BC
KOS System Change

system change? 81% NO
to

60% RVK

10% undecided
15% others
5% Thesaurus
5% BC
5% DDC
Software used for KOS Management

- PICA
- Aleph
- Other: OSIRIS/BREWIS, DSWORM, ABACUS, E-LIB, Alephino

<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pica</td>
</tr>
<tr>
<td>Aleph</td>
</tr>
<tr>
<td>aDIS/BMS</td>
</tr>
<tr>
<td>URICA</td>
</tr>
<tr>
<td>ALMA</td>
</tr>
<tr>
<td>SydneyPLUS/Lucidea</td>
</tr>
<tr>
<td>Allegro</td>
</tr>
<tr>
<td>Eigenentwicklung</td>
</tr>
<tr>
<td>DSWORM</td>
</tr>
<tr>
<td>OSIRIS/BREWIS</td>
</tr>
<tr>
<td>ABACUS</td>
</tr>
<tr>
<td>E-LIB</td>
</tr>
<tr>
<td>Alephino</td>
</tr>
<tr>
<td>DSpace</td>
</tr>
<tr>
<td>BibliothecaPlus</td>
</tr>
<tr>
<td>Office</td>
</tr>
<tr>
<td>Geocom Bibliothek (Lotus Notes basiert)</td>
</tr>
<tr>
<td>FileMaker (Eigenentwicklung)</td>
</tr>
<tr>
<td>SISIS-SunRise (lokal)</td>
</tr>
<tr>
<td>IFIS</td>
</tr>
<tr>
<td>HohSearch</td>
</tr>
<tr>
<td>STAR</td>
</tr>
</tbody>
</table>
Subject indexing - How is it done?

Software:
- SisisSunrise library system
- SIKOM-interface
- IntelligentCapture
- Averbis Extraction Platform
- Semiautomatic Digital Assistent by Eurospider

96 % intellectually
4 % automatically
Results Part II - Mapping Projects
Mappings I – Projects & Concordances

- **Projects**: 80% No, 20% Yes
- **Concordances**: 70% No, 30% Yes
Mappings – Processes & Types

Classification systems
mostly: manual/intellectual, unidirectional mappings

Thesauri
few: semi-/automatic mappings
(tools only for thesauri: OSIRIS, Amalgame, Decision Trees)
Mappings – KOS mapped to Dewey
## Concordances

### coli-conc Mapping Database:

### KOS and topic | creator | No. of Mappings
--- | --- | ---
**DDC to ...**
DDC-BC Chemistry | unknown | 100
DDC-BC politics | SUB Hamburg | 851
DDC-RVK (DDC Class 100) | VZG | 754
DDC-RVK (DDC first 1000 classes) | VZG | 2065
DDC-RVK Library and Information science | HdM Stuttgart | 385
DDC-RVK Ethics | VZG | 224
DDC-RVK Medicine | VZG | 2814
DDC-RVK Law | VZG | 2268
DDC-BK (DDC first 1000 classes) | VZG | 1376

**RVK to ...**
RVK-BC Law | University and Regional Library of Tyrol | 16036
RVK-BC Economics | University and Regional Library of Tyrol | 1042
RVK-DDC Ethics | VZG | 27
RVK-DDC Philosophy and Psychology | VZG | 358
RVK-GND | UB Regensburg | 94196
RVK-BK German studies | University and Regional Library of Tyrol | 13249

**SWD to ...**
SWD-DDC Library and Information science | HdM Stuttgart | 462
SWD-RVK Library and Information science | HdM Stuttgart | 396

**currently total: 137403**

[http://coli-conc.gbv.de/concordances/](http://coli-conc.gbv.de/concordances/)
Conclusion

1. The use of the DDC has certainly increased since 2000, however not significantly

2. The RVK is consistently well represented

3. Still there is a huge number of libraries using home classification systems that do not strive for a change or a uniform standard

4. Several software are available, for management of KOS, however methods, standards and tools for creating mappings between KOS, in particular between classification systems lacking
JSKOS data format
Objectives

1. Creation and Manage of Concordances and KOS
2. Improvement the Quality of the Mappings
3. Facilitate Use and Exchange of KOS and their Mappings

This requires

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

- Based on JSON(-LD) ⇒ compatible with SKOS/RDF but easier to use, especially in web applications
- Adds expression of concordances, mappings, and registries (which were lacking in pure SKOS or other formats)
- Reference implementations in PHP, JavaScript & Java
The JSKOS data format

- Based on JSON(-LD) ⇒ compatible with SKOS/RDF but easier to use, especially in web applications
- Adds expression of concordances, mappings, and registries (which were lacking in pure SKOS or other formats)
- Reference implementations in PHP, JavaScript & Java

```json
{
  "uri": "http://dewey.info/class/641.5/e23/",
  "notation": [ "641.5" ],
  "prefLabel": { "en": "Cooking", "de": "Kochen" },
  "narrower": [ ... ],
  ...
}
```
JSKOS-API

Query KOS and mappings to include in other applications
  • cocoda mapping tool
  • subject indexing
  • …

REST-API
  • ?notation=614.5&language=en,de
  • ?search=Cook&format=suggest
  • …
Current state

Done enough to build on

- Several Open Source implementations (see homepage)
- GND and RVK expressed in JSKOS

In progress

- JSKOS-API specification
- Implement JSKOS-API in existing KOS software
- Express DDC in JSKOS, lacking dewey.info :-(

Thank you!