

Zum Schluss wurde im Zusammenhang mit dem Business Meeting am späten Nachmittag das zehnjährige EDUG-Jubiläum mit lebhaften Dewey-Diskussionen und Champagner gefeiert.

Die einzelnen Vorträge befinden sich auf dem EDUG-Server unter <http://edug.pansoft.de/tiki-index.php?page=Annual%20meetings>. Ein ausführlicher EDUG-Bericht 2016 findet sich unter: <https://coli-conc.gbv.de/publications/>.

## DFG-Projekt coli conc - Projektphase I, Uma Balakrishnan

Das VZG-Konkordanz-Projekt „coli-conc“ befasst sich seit dem Jahr 2011 mit der Erstellung von Konkordanzen zwischen Klassifikationen. Bis jetzt sind für ca. 16,4 % der DDC-Klassen Mappings sowohl mit der RVK (Regensburger Verbundklassifikation) als auch mit der BK (Basisklassifikation) intellektuell erstellt worden. Die dabei abgedeckten Fachbereiche umfassen unter anderem Medizin, Recht, Philosophie und die dritte Ebene der DDC-Klassen.

Im Jahr 2014 verlagerte sich der Schwerpunkt von der manuellen Erstellung zur Entwicklung einer Infrastruktur, die die Bildung, den Austausch und die Wartung von Konkordanzen zwischen bibliothekarischen Wissensorganisationssystemen (Knowledge Organization Systems, KOS) erleichtern und eine Plattform für kollaborative Arbeit zwischen den ExpertInnen ermöglichen soll.

Um dieses Ziel zu erreichen, wurde das Projekt in sechs Teilaufgaben und mehrere Arbeitspakete gegliedert.

### Bestandsaufnahme

- Bestandsaufnahme vorhandener Quellen für KOS und Konkordanzen
- Bestandsaufnahme vorhandener Software zur Verwaltung von KOS

### Bereitstellung

- Entwicklung von Server-Komponenten zur Speicherung und Bereitstellung von KOS (KOS-DB)
- Entwicklung von Server-Komponenten zur Verwaltung von Konkordanzen (KK-DB)
- Anbindung an vorhandene KOS-Software, -Portale und Katalogisierungswerzeuge
- Erprobung und Dokumentation von Verfahren zum Einsatz der Anwendung in Cloud-Diensten und Bereitstellung aller Inhalte als (Linked) Open Data

### Qualitätssicherung

- Bestandsaufnahme vorhandener Verfahren zur Bewertung von Konkordanzen
- Konzeption und Test der Einbindung von Crowdsourcing-Verfahren für die Beurteilung von Konkordanzen
- Entwicklung von Server-Komponenten zur Ermittlung und Bereitstellung von Bewertungen (KK-Measure)

## Mapping-Tool

- Entwicklung der Benutzeroberfläche zur Navigation in vorhandenen KOS
- Entwicklung der Benutzeroberfläche zum Anlegen und Bearbeiten von Mappings
- Entwicklung der Benutzeroberfläche für Mapping-Vorschläge und -Bewertungen
- Usability-Tests der Benutzeroberfläche

## Mapping-Verfahren

- Umsetzung von Mappingverfahren als Server-Komponenten zur Ermittlung und Bereitstellung von Mapping-Vorschlägen im Mapping-Tool
- Ermittlung und Verbesserung von vorhandenen Mappingverfahren für bibliothekarische Klassifikationssysteme und Vorbereitung ihrer technischen Umsetzung

## Einbindung größerer Nutzergruppen

- Dokumentation und Öffentlichkeitsarbeit zu KOS und Konkordanzen
- Durchführung von Workshops mit Fachleuten und Nutzern
- Konzeption und Umsetzung von Authentifizierungs- und Authorisierungsverfahren
- Konzeption von Crowdsourcing-Verfahren

Im Dezember letzten Jahres erhielt „coli-conc“ eine DFG-Förderung für zunächst zwei Jahre. Dies ermöglichte es, das Projektteam um eine Softwareentwicklerin, Morsheda Akter und eine Systembibliothekarin, Jana Agne, zu erweitern und damit in die Realisierungsphase überzugehen.

Ferner kooperieren Universitäts- und wissenschaftliche Bibliotheken wie z.B. das Leibniz Institut für Sozialwissenschaften (GESIS), Köln, die UB Regensburg und das Leibniz Informationszentrum Wirtschaft (ZBW) Kiel sowie BARTOC (<http://bartoc.org/>) als Projektpartner mit coli-conc.

In den letzten vier Monaten konnten wesentliche Fortschritte bei den Teilaufgaben „Bestandsaufnahme“ und „Bereitstellung“ erzielt werden.

Als erster Schritt wurde Anfang des Jahres eine Umfrage in der Fachwelt durchgeführt, um den aktuellen Stand der Verbreitung der Sacherschließungselemente zu evaluieren. Insbesondere die Klassifikationssysteme, deren Mappings und laufende Mapping-Projekte lagen im Fokus dieser Untersuchung.

## Zusammenfassung der Umfrageauswertung

- Die Regensburger Verbundklassifikation (RVK) ist mit 36 % das im deutschsprachigen Raum am stärksten genutzte Klassifikationssystem und gewinnt seit dem Jahr 1999 stetig an Nutzern. Sie wird gefolgt von hauseigenen Systemen mit 33 %, der Dewey Dezimalklassifikation (DDC) an dritter Stelle mit 26 % und der Basisklassifikation (BK) mit 10%. (vgl.: Umlauf, 1999 <http://www.ib.hu-berlin.de/~kumlauf/handreichungen/h67/>).

- Bei den Thesauri und Subject Headings ist die Gemeinsame Normdatei (GND) das am häufigsten genutzte System mit 68 % gefolgt von hauseigenen Systemen (19 %), dem Standard Thesaurus Wirtschaft (STW) mit 5 %, den Medical Subject Headings (MeSH) mit 4 % und den Library of Congress Subject Headings (LCSH) mit 3 %.
- Die Klassifikationssysteme RVK, DDC und BK werden seit den 90er Jahren vermehrt genutzt.
- Es gibt keine Hinweise auf eine Vereinheitlichung der KOS im deutschsprachigen Raum. Viele Bibliotheken nutzen weiterhin ihre hauseigenen Systeme. Ein Systemwechsel wird von nur wenigen Bibliotheken (20 %) in Betracht gezogen. Wenn ein Wechsel geplant oder durchgeführt wird, ist eine Tendenz zur RVK zu verzeichnen.
- Die Sacherschließung erfolgt weiterhin überwiegend intellektuell (95 %).
- Fast alle Bibliotheken verwenden zur Verwaltung von KOS Pica (53 %) und Aleph (21 %), die nicht vorrangig zur Verwaltung von KOS gedacht sind (vgl.: Liste von Open-Source Software: <http://dx.doi.org/10.5281/zenodo.48740>).
- Nur ein kleiner Teil der Befragten (12 %) hat eigene Konkordanzprojekte durchgeführt oder geplant, aber 30 % der Gesamteilnehmer haben schon eine Konkordanz erstellt und das überwiegend zwischen zwei Klassifikationssystemen (69 %).
- Die meisten Konkordanzen (80 %) werden manuell bzw. intellektuell erstellt.
- Das automatische oder semiautomatische Verfahren wird hauptsächlich für die Erstellung von Mappings zwischen Thesauri verwendet.
- Es fehlt an Methoden, Standards und Tools für die Erstellung, Bearbeitung und Nachnutzung von Mappings, vor allem zwischen bibliothekarischen Klassifikationssystemen.

Ausführliche Ergebnisse der Umfrage sind auf der coli-conc Homepage (<https://coliconc.gbv.de/publications/>) veröffentlicht.

#### KOS-Metadatenbank- Webinterface

Auf Basis der Ergebnisse der Umfrage und weiterer Recherchen wurde eine „KOS Webanwendung“ entwickelt, die einen einheitlichen Zugang zu Metadaten der im deutschsprachigen Raum verbreiteten KOS bereitstellt. Die Datenbank dieser Webanwendung enthält derzeit 32 KOS, die mit Daten von BARTOC.org angereichert wurden (siehe Abb. 2). Zusätzlich wurde die Option implementiert, die KOS-Metadaten im JSKOS- oder Excel-Format herunterzuladen. Diese KOS fungieren als Basis für das Konkordanz-Tool.

Der KOS-Service wird fortlaufend erweitert und steht auf der coli-conc Projektseite unter: <http://coliconc.gbv.de/terminologies/> zur Verfügung.

#### Die Verwaltung und Darstellung der KOS

Für die KOS-Verwaltung existieren verschiedene Software-Tools in Form von Thesauri, Editoren oder Vokabular-Management-Systemen. Daher ist in diesem Projekt keine Software-Eigenentwicklung geplant, stattdessen wird eine bestehende KOS-Software eingesetzt.

Zu diesem Zweck wurde eine Liste mit Open Source KOS-Management-Software für eine weiterführende Evaluation zusammengestellt (<http://dx.doi.org/10.5281/zenodo.48740>). Diese Liste zeigt, dass Java und PHP die meistgenutzten Programmiersprachen in diesem Bereich sind.

## Das JSKOS Datenformat und die JSKOS-API

Trotz einiger Standardisierungsbemühungen (SKOS und SPARQL) hat die begrenzte Verfügbarkeit von Konkordanzen zu einem Mangel an befriedigenden allgemeinen Datenformaten und Verfahren für den Zugriff auf KOS und ihre Mappings geführt. Die Notwendigkeit eines flexibleren Datenformats führte im Rahmen des Projektes zu der Entwicklung des JSKOS-Datenformates (auf Grundlage von SKOS) und der JSKOS-API. Das JSKOS-Datenformat wurde in einer ersten Version unter <https://gbv.github.io/jskos/> publiziert. Die Konvertierung der KOS nach JSKOS ist für die in diesem Projekt meistgenutzten Systeme DDC, RVK und GND fast abgeschlossen. Die Daten wurden in eine Datenbank importiert. Zusätzlich werden Wrapper getestet, um den Zugang zu bestehenden KOS zu erlangen.

## Konkordanz-Datenbank

Um einen einheitlichen Zugang zu vorhandenen Mappings aus verschiedenen Projekten zu gewähren und zugleich diese auch als Mapping-Kandidaten für das semi-automatische Mapping-Vorschlagsystem zu nutzen, wurden die vorhandenen Mappings in einer Datenbank gespeichert (Abb. 3). Diese dient als zentraler Konkordanz-Speicher und ist über die Web-Oberfläche leicht zugänglich (<http://coli-conc.gbv.de/concordances/>).

Abb. 2: *KOS Metadaten - Webinterface*

The screenshot shows a web-based interface for managing knowledge organization systems (KOS). At the top, there is a search bar labeled "Search: Enter keyword" and a dropdown menu labeled "Subject Classificati". Below the search bar, a list of terminologies is displayed, each with a small circular icon and a name. A blue button labeled "Expand for Metadata" is positioned above the list. At the bottom of the list, there is a "Export JSON and Excel" button with icons for both formats. A footer note at the very bottom states: "This database is a part of project coli-conc. See BARTOC for a general register of terminologies. Bartoc website: [BARTOC](#)".

Terminology
MSC (Mathematics Subject Classification)
PACS (Physics and Astronomy Classification Scheme)
FKDigBib (Fachklassifikation digitale Bibliothek)
KfM (Klassifikation für Musikbibliotheken)
Klassifikation des Deutschen Patent- und Markenamts
SEB (Systematik für evangelische Büchereien)
Sachbuch-Systematik für Katholische öffentliche Büchereien
KfKJ (Klassifikation für Kinder- und Jugendbibliotheken)
SSD (Systematik der Stadtbibliothek Duisburg)
ESSB (Einheitssystematik für Südtiroler Bibliotheken)

Die bisher entstandenen bzw. eingespielten Mappings wurden mit dem Tool Catmandu nach JSKOS konvertiert und anschließend in eine NoSQL-Datenbank (MongoDB) importiert. Zurzeit enthält die Datenbank ca. 137.400 uni- und bidirektionale Mappings zwischen DDC, RVK, GND und BK (Abb. 4). Zudem ist eine Download Option in verschiedenen Formaten geplant.

## Ausblick

Das DFG-Projekt ist zunächst bis Dezember 2017 bewilligt.

Die nächste Phase des Projektes umfasst:

- Fortlaufende Konvertierung der KOS bzw. deren Mappings in das JSKOS-Format
- Weiterentwicklung des Datenbank-Webservice
- Weiterentwicklung des Suchalgorithmus zwischen der DDC und der RVK für die automatische Generierung von Mappingvorschlägen
- Laufende Erstellung von intellektuellen Mappings für die Fächer Informatik und Physik zwischen der DDC und der RVK

Zusätzlich beabsichtigt das Projekt in den kommenden Monaten die Analyse und Evaluierung der Sacherschließungslandschaft in den Museen.

Abb. 3: Konkordanz-Datenbank Webinterface

The screenshot shows a browser window with the URL <https://coli-conc.gbv.de/>. The page title is "coli-conc". Below it, there's a "Concordances" section with a note about an early preview of the mapping database. Under "Search Mappings", there are three input fields: "Source" (DDC), "Target" (RVK), and "Creator" (VZG). A "Search" button is present. Below these, a table lists 8118 mappings found, showing columns for Source, Target, and various identifiers like ST 302, AN 14000, etc. The table has headers for Source, Target, Creator, Last Change, and Relevance.

Abb. 4: Übersicht vorhandener Mappings

KOS und Fachgebiet	Ersteller	Anzahl Mappings
DDC-BC Chemie	Unbekannt	100
DDC-BC Politik	SUB Hamburg	851
DDC-RVK Philosophie und Psychologie	VZG	1.112
DDC-RVK (DDC 1000er Klassen)	VZG	2.065
DDC-RVK Bibliotheks- und Informationswissenschaften	HdM Stuttgart	385
DDC-RVK Ethik	VZG	251
DDC-RVK Medizin	VZG	2.814
DDC-RVK Recht	VZG	2.268
DDC-BK (DDC 1000er Klassen)	VZG	1.376
RVK-BC Recht	Uni- und Landesbibliothek Tirol	16.836
RVK-BC Wirtschaft	Uni- und Landesbibliothek Tirol	1.042
RVK-GND	UB Regensburg	94.196
RVK-BK Germanistik	Uni- und Landesbibliothek Tirol	13.249
SWD-DDC Bibliotheks- und Informationswissenschaften	HdM Stuttgart	462
SWD-RVK Bibliotheks- und Informationswissenschaften	HdM Stuttgart	396
<b>Gesamt: 137.403</b>		