

Uma Balakrishnan, Stefan Peters, Jakob Voß

# coli-conc – Eine Infrastruktur zur Nutzung und Erstellung von Konkordanzen

## 1 Einleitung

coli-conc<sup>1</sup> ist eine Dienstleistung der Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG). Sie stellt webbasierte Dienste für einen effektiveren Austausch von Wissensorganisationssystemen und für die effiziente Erstellung und Wartung von Mappings zur Verfügung. Der Schwerpunkt liegt auf den im deutschsprachigen Raum verbreiteten bibliothekarischen Klassifikationen und Normdateien, vor allem den bedeutenden Universalklassifikationen wie Dewey Dezimalklassifikation (DDC), Regensburger Verbundklassifikation (RVK), Basis-klassifikation (BK) und den Sachgruppen der Deutschen Nationalbibliografie (SDNB). Dieser Bericht beschreibt den Hintergrund, die Architektur und die Funktionalitäten von coli-conc sowie das Herzstück der Infrastruktur – das Mapping-Tool *Cocoda*. Außerdem wird auf Maßnahmen zur Qualitätssicherung eingegangen und ein Einblick in das neue Mapping-Verfahren mit dem Konzept-Hub gewährt.

## 2 Hintergrund

Die Sacherschließungslandschaft des deutschsprachigen Raums ist von einer Vielzahl von Klassifikationen, Thesauri und Normdateien (zusammengefasst kurz KOS für *Knowledge Organization Systems*) geprägt. Bemühungen in den letzten Jahrzehnten, ein einheitliches System zu etablieren, blieben erfolglos. Allerdings haben die Nutzung und der Austausch von Wissensorganisationssystemen in der bibliothekarischen Welt in den vergangenen Jahren durch Entwicklung geeigneter Standards und Werkzeuge deutlich zugenommen. Zudem hat die Standardisierung des *Simple Knowledge Organization Systems* (SKOS) im Rahmen von *Linked Open Data* (LOD) dazu beigetragen, dass Wissensorganisationssysteme zunehmend frei zur Verfügung stehen (Hanke 2014). Darüber hinaus wurden in den letzten Jahren mehrere Programme zur Verwaltung und Bereitstellung von KOS entwickelt (iQvoc, Poolparty, VocBench, OpenSKOS, Skosmos, xTree, etc.) sowie mehrere Portale zur Sammlung und Publikation

---

1 <https://coli-conc.gbv.de> (28.12.2020).

vorhandener KOS aufgebaut (Voß, Ledl und Balakrishnan 2016). Die einzelnen KOS oder Teile davon werden dabei idealerweise zum Download und über eigene APIs bereitgestellt.

Diese Entwicklungen haben ermöglicht, KOS vermehrt zur strukturierten Erschließung von Dokumenten und Daten anzuwenden. Jedoch ist die Situation bei Konkordanzen (auch bekannt als Cross-Konkordanzen oder Terminologie-Mappings), die Begriffe unterschiedlicher KOS aufeinander abbilden (Mayr 2010; Keil 2012), anders. Konkordanzen zwischen verschiedenen Systemen sind sehr begrenzt verfügbar. Dabei sind sie unabdingbar für die Überwindung der Heterogenität in mehrsprachigen Datenquellen und für die Verbesserung des Information Retrievals.

Generell ist die Konkordanzerstellung durch ein rein intellektuelles Verfahren sehr aufwendig und zeitintensiv und stößt angesichts der Tiefe, Größe und kontrastiven Natur verschiedener KOS insbesondere bei fein gegliederten Klassifikationssystemen wie z. B. der DDC mit etwa 49 000 Klassen und der RVK mit etwa 850 000 Klassen an ihre praktischen Grenzen. Gleichzeitig ist aber eine vollständige Konkordanz speziell zwischen diesen bedeutenden Systemen (Hermes 1998) und eine einheitliche Erschließung mit mehreren KOS ein langjähriges Desiderat der Expert:innen-Community.

Aufgrund der ersichtlichen Schwierigkeiten bei der Konkordanzerstellung, trotz vieler Konkordanz-Projekte seit 2000 wie z. B. MACS (seit 1997), Renardus (1999–2002), CrissCross (2006–2010) und KoMoHe (2004–2007), fehlten vollständige Konkordanzen zwischen den im deutschsprachigen Raum aktiv genutzten Wissensorganisationssystemen (Balakrishnan 2013).

Automatische Verfahren waren Gegenstand der Forschung (siehe OAEI Library Track, 2012–2014<sup>2</sup> und z. B. Pfeffer 2013), aber diese Verfahren waren unvollständig und eher für Mapping-Vorschläge geeignet (Lauser 2008; Reiner 2010). Vorhandene, manuell erstellte Konkordanzen decken oft lediglich einen kleinen Teil der von ihnen gemappten KOS ab. Vor allem im Bereich bibliothekarischer Klassifikationen gab es weder Werkzeuge noch etablierte Verfahren für die Erstellung von Konkordanzen. Bislang existierte keine Übersicht vorhandener Konkordanzen und Konkordanz-Projekte, weil die Ergebnisse der Konkordanz-Projekte meist nur intern genutzt wurden und öffentlich nicht zugänglich waren (Balakrishnan 2013). Zudem lagen die vorhandenen Konkordanzen zu großen Teilen in nicht-standardisierten Formaten vor, ohne genauere Angaben über Methodik und Aktualität. Zu den wenigen Ausnahmen frei zugänglicher Konkordanzen gehören das Projekt VIAF von OCLC (OCLC 2014), eine Sammlung von Konkordanzen der Zentralbibliothek Wirtschaft (ZBW) und das Projekt

---

2 <http://oaei.ontologymatching.org/2014/library/> (28.12.2020).

Wikidata der Wikimedia Foundation (Voß 2014, Vrandečić 2014). Das Potential offener Community-Projekte wie Wikidata für die Erstellung von Konkordanzen ist zwar groß (vgl. Hanke 2014 und Soergel 2011) und die in diesen Projekten erarbeiteten Folksonomies lassen sich auf etablierte KOS abbilden (Voß 2007; 2012) – offen bleibt jedoch auch hier die Qualität der so erstellten Konkordanzen. Die geringe Verfügbarkeit von Konkordanzen hat auch zur Folge, dass sich bisher, trotz Bemühungen zur Standardisierung (siehe SKOS, ISO 25964-2, Keil 2012, Soergel 2011), weder Austauschformate noch Verfahren für die nachhaltige Pflege und Bereitstellung durchgesetzt haben.

coli-conc identifiziert und schließt diese Lücke, indem es ermöglicht, mit der Entwicklung einer Infrastruktur, eines neuen Datenformats und unter Zusammenarbeit mit Fachexpert:innen, die Komplexität dieser Aufgabe zu bewältigen.

Die Hauptgründe für die Entwicklung der coli-conc Dienste waren:

- Effektivere Erstellung und Pflege von Konkordanzen zwischen bibliothekarischen KOS
- Verbesserung der Qualität von Konkordanzen
- Förderung des Austauschs und der Nutzung von Konkordanzen und KOS
- Unterstützung der Fachreferent:innen bei der Sacherschließungsaufgabe

Die Ziele umfassten:

- Entwicklung frei kombinierbarer Open-Source-APIs und Nutzung dieser für webbasierte Dienste, um einen einheitlichen Zugriff auf KOS und KOS-Zuordnungen sowie Tools zur Erstellung computergestützter intellektueller Mappings zwischen vielen KOS in mehreren Sprachen zu bieten
- Sammeln und Verwalten von KOS-Daten und Konkordanzen in einer einheitlichen Struktur (JSKOS-Format) auf einer zentralen Plattform
- Erarbeitung von Best Practices gemeinsam mit Sacherschließungsexpert:innen zur Qualitätssicherung von KOS und Mappings
- Unterstützung bei der Nachnutzung vorhandener KOS-Daten und -Mappings für verschiedene Dienste, einschließlich:
  - Wissensbasis für automatische bzw. semi-automatische Sacherschließung, wie z. B. Digitaler Assistent
  - Integration in bzw. Verlinkung zu Katalogisierungssoftware, wie z. B. WinIBW
  - Wissensbasis zum Durchsuchen von Hierarchien und/oder für automatisierte oder computergestützte Erweiterung der Recherche durch Hinzufügen von Synonymen und verwandten Begriffen in den Katalogen

- Anreicherung von Daten, Ermöglichung von Interoperabilität und Abfragen zwischen verschiedenen Research-Systemen
- Weiterentwicklung sowie Verbesserung von KOS

### 3 Die coli-conc-Architektur

Für den Aufbau der Infrastruktur von coli-conc (s. Abb. 1) wurden existierende Programme (KOS-Software, KOS-Portale, Mapping-Anwendungen und -Algorithmen, Katalogisierungssoftware, etc.) evaluiert, ausgewählt bzw. erweitert und neue Komponenten insbesondere für die Speicherung, Bereitstellung und Analyse von Konkordanzen erstellt. Die Systemarchitektur ist sehr modular, so dass verschiedene Komponenten unabhängig voneinander genutzt und vorhandene Programme in die Infrastruktur eingebunden werden können. Grundlage hierfür ist das gemeinsame Datenformat JSKOS und die JSKOS-API. Alle im Rahmen des Projekts entwickelten Bestandteile werden als Open Source veröffentlicht.<sup>3</sup>

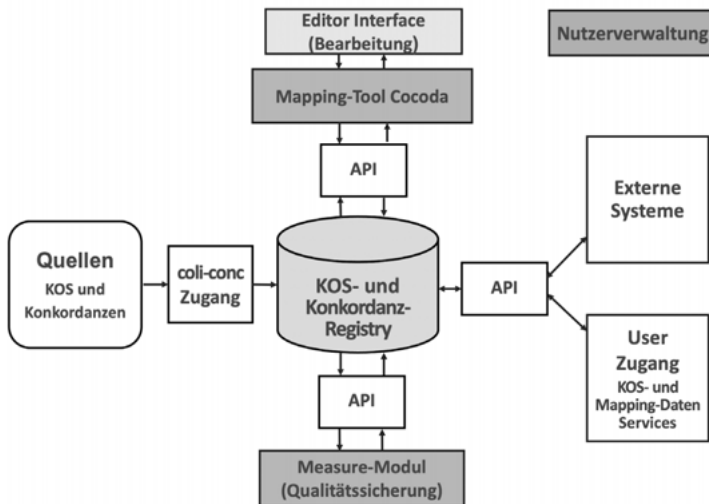


Abb. 1: coli-conc-Architektur

<sup>3</sup> <https://coli-conc.gbv.de/publications/#software> (28.12.2020).

### 3.1 Das JSKOS-Format

Das Datenformat JSKOS<sup>4</sup> wurde für eine einheitliche Darstellung von KOS und KOS-Daten in coli-conc, die aus unterschiedlichen Quellen stammen, entwickelt (Voß, Ledl & Balakrishnan 2016; Voß 2017). Das Format basiert im Wesentlichen auf SKOS und JSON-LD. Unterstützt werden neben KOS und KOS-Konzepten auch KOS-Verzeichnisse, Mappings, Konkordanzen, Mapping-Bewertungen und Daten zur Verwendung von Konzepten (*Concept Occurrences*). Die in reinem SKOS nur rudimentär abbildbaren Mapping-Daten werden in JSKOS erweitert um:

- Mappings zwischen KOS-Konzepten als eigenständige Objekte mit Metadaten wie Provenienz
- Mappings mit mehreren Konzepten (1-zu-n Beziehungen)
- Null-Mappings, falls keine passenden Konzepte in einem KOS existieren
- erweiterbare Beziehungstypen, die über die Standard-Thesaurus-Relationen für Hierarchien und Assoziationen hinausgehen
- Konfidenzlevel für die Mappings
- eine strenge Definition der Codierung wiederholbarer und nicht wiederholbarer Felder

Konvertierungen von und nach JSKOS sind mit anderen RDF-Serialisierungen, mit dem MARC-21-Format für Klassifikationen (Heggø 2017) sowie eingeschränkt mit CSV-Daten möglich. JSKOS ist vorwiegend für Webanwendungen konzipiert, es kann aber auch als internes Format für NoSQL-Datenbanken wie MongoDB genutzt werden.

```
{
  "type": [ "http://www.w3.org/2004/02/skos/core#narrowMatch" ],
  "fromScheme": { "uri": "http://dewey.info/scheme/edition/e22/" },
  "toScheme": { "uri": "http://bartoc.org/en/node/454" },
  "from": {
    "memberSet": [ {
      "uri": "http://dewey.info/class/387/e22/",
      "notation": [ "387" ],
      "preflabel": { "en": "Water, air, space transportation" }
    } ]
  },
  "to": {
    "memberSet": [ {
      "uri": "http://id.loc.gov/authorities/subjects/sh85121579",
      "preflabel": { "en": "Shipping" }
    } ]
  }
}
```

**Abb. 2:** Ein Mapping im JSKOS-Format

<sup>4</sup> <https://gbv.github.io/jskos/> (28.12.2020).

## 3.2 JSKOS-API

Für einen bequemen und direkten Zugriff auf KOS-Daten (Notationen, Konzepte, Begriffe und Beziehungen) verschiedener Quellen sind Webdienste erforderlich. Einige Systeme stellen KOS-Daten über APIs zur Verfügung (Voß und Balakrishnan 2016), aber die Abdeckung und Qualität der Terminologiedienste unterscheiden sich stark. Zur einheitlichen Abfrage von KOS-Daten aus verschiedenen Terminologiediensten wurde nach Analyse bestehender KOS-APIs die JSKOS-API entwickelt. Die API ist für die zentrale Datenbankkomponente JSKOS-Server, als API-Wrapper für vorhandene Terminologiedienste wie Skosmos und lobid<sup>5</sup> sowie für den Terminologiedienst DANTE<sup>6</sup> umgesetzt (Dührkoph 2017). KOS-Daten aus all diesen Quellen können somit einheitlich im Web abgerufen werden. Zur weiteren Vereinfachung des Zugriffs gibt es außerdem eine JavaScript-Bibliothek.<sup>7</sup>

## 3.3 Die coli-conc-Datenbank

Zur Speicherung von KOS-Daten und Mappings aus verschiedenen Quellen und Projekten wird die Datenbankanwendung JSKOS-Server eingesetzt.<sup>8</sup> Die coli-conc-Datenbank ist eine zentrale Instanz von JSKOS-Server, in der die KOS- und Mapping-Daten im JSKOS-Format eingespeist und zur Verfügung gestellt werden.<sup>9</sup> Dabei wird jeder Datensatz mit Metadaten (importierte KOS, Konkordanzen, Creator, Datum, ...) beschrieben. Inhaltlich lässt sich die Datenbank in die zwei Verzeichnisse KOS-Registry und Konkordanz-Registry aufteilen:

- Die *KOS-Registry* enthält verschiedene KOS und deren Metadaten. Die beschreibenden Metadaten wie Kurzbeschreibung, Herausgeber:innen und Aktualität bilden einen Teil des Terminologieverzeichnis BARTOC.<sup>10</sup> Diese Registry unterstützt die Suche und Filterung nach KOS-Typ, Sprache, Lizenz, etc. und ermöglicht das Herunterladen von KOS-Metadaten in verschiedenen Formaten.
- Die *Konkordanz-Registry* enthält Mappings und Konkordanz-Informationen, die im Rahmen des coli-conc-Projektes gesammelt oder erstellt werden. Die

<sup>5</sup> <https://lobid.org> (28.12.2020).

<sup>6</sup> <http://api.dante.gbv.de> (28.12.2020).

<sup>7</sup> <https://github.com/gbv/cocoda-sdk> (28.12.2020).

<sup>8</sup> <https://github.com/gbv/jskos-server> (28.12.2020).

<sup>9</sup> <https://coli-conc.gbv.de/api/> (28.12.2020).

<sup>10</sup> Das an der Universitätsbibliothek Basel entwickelte *Basic Register of Thesauri, Ontologies & Classifications* wird seit Ende 2020 von der VZG betrieben: <https://bartoc.org> (28.12.2020).

Mappings in der *Konkordanz-Registry* können einzeln nach Systemen, Konzepten, Notationen und Autor:innen durchsucht und exportiert werden. Zudem bietet diese Registry eine Übersicht aller Konkordanzen in einzelnen Bereichen oder Systemen, die auch in JSKOS oder als CSV-Datei zum Herunterladen bereitgestellt sind.

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	15,034		
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	10,882		
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	376		
RVK	GND	GND-Indexterme der RVK	UB Regensburg	2015	JSKOS CSV	93,430		
DDC	IXTHEO	DDC zu IxTheo-Klassifikation	Timotheus Chang Whae Kim	2018	JSKOS CSV	573		
RVK	BK	Wirtschaft	ULB Tirol	2013	JSKOS CSV	732		
DDC	BK	Politikwissenschaft	SUB Hamburg		JSKOS CSV	655		
DDC	RVK	Philosophie und Psychologie (1-)	VZG		JSKOS CSV	1,046		
DDC	RVK	Medizin und Gesundheit	VZG	2011	JSKOS CSV	2,777		
GND	LCSH	project macs	DNB	2016	JSKOS CSV	43,396		
STW	GND	Ausschnitt mit 1-zu-1 Mappings	ZBW		JSKOS CSV	15,239		
RVK	DDC	CL-CZ Psychologie	Manuela Queitsch		JSKOS CSV	407		
RVK	DDC	TE-TZ Geologie und Paläontologie	GESIS		JSKOS CSV	1,162		
RVK	DDC	W Biologie	GESIS		JSKOS CSV	5,347		
RVK	DDC	SA-SP Mathematik	GESIS		JSKOS CSV	227		
GND	DDC	CrisCross-SWD-DDC-Mapping	DNB		JSKOS CSV	170,638		
RVK	DDC	Ethik	VZG	2015	JSKOS CSV	23		
RVK	DDC	Philosophie und Psychologie	VZG	2016	JSKOS CSV	357		
DDC	RVK	Recht	VZG	2012	JSKOS CSV	2,190		
DDC	RVK	DDC 100er-Klassen	VZG	2013	JSKOS CSV	738		

27 Concordances Total: 370,798

Abb. 3: Konkordanz-Registry

### 3.4 Das Mapping-Tool Cocoda

Das Kernstück der coli-conc-Infrastruktur ist das Mapping-Tool *Cocoda*.<sup>11</sup> Die Webanwendung unterstützt und vereinfacht das intellektuelle Mapping-Verfahren und erhöht die Effizienz der Erstellung von Mappings. Die erstellten

<sup>11</sup> <https://coli-conc.gbv.de/cocoda/> (28.12.2020).

Mappings sind über die Konkordanz-Registry grundsätzlich öffentlich verfügbar, außer sie werden lokal im eigenen Browser gespeichert. Cocoda besteht unter anderem aus Modulen für folgende Funktionen.

### 3.4.1 KOS-Browsing und -Anzeige

Das Modul *KOS-Browsing und -Anzeige* erlaubt die Suche von und in KOS mit Anzeige von Hierarchien und Konzeptbeziehungen zum Browsen und zur Anzeige aller Informationen zu einem ausgewählten Konzept. Links und rechts der Cocoda-Anwendung stehen Quell- und Ziel-Vokabular zur Auswahl. Derzeit sind 170 Vokabulare in Cocoda eingebunden und lassen sich für das Mapping auswählen. Unter den wichtigsten Systemen sind DDC Deutsch, RVK, BK, die Gemeinsame Normdatei (GND), Wikidata, Iconclass, die Hessische Systematik, die Oberbegriffsdatei (OBG), die Bremer Online Systematik (BOS), die DFG-Fachsystematik, die Sachgruppen der DNB und die THEMA-Klassifikation des Buchhandels. Die Einbindung weiterer Vokabulare ist nach Festlegung von URIs für die einzelnen KOS und nach Vergabe einer freien Lizenz möglich. So wurde beispielsweise die Klassifikation der Internationalen Bibliographie für Theologie und Religionswissenschaft (IxTheo) hinzugefügt, um sie auf DDC zu mappen. Das KOS-Browsing-Modul ermöglicht es, in einem einzelnen Vokabular einen bestimmten Begriff für das Mapping durch Browsen auszuwählen und auch direkt nach einem Begriff oder nach Notationen zu suchen. Bei Klassifikationen wird die Baumansicht der Hierarchie angezeigt. Der Zugriff auf verschiedene Vokabulare erfolgt über den Normdatendienst DANTE und über die eigene coli-conc-Datenbank.

### 3.4.2 Mapping-Anzeige und -Vorschläge

Dieses Modul bietet die Suche und Bewertung vorhandener Mappings und die Auswahl von Mapping-Vorschlägen aus verschiedenen Quellen. Dies beinhaltet:

- *Existierende Mappings*: Zum einen werden Mappings aus der zentralen Konkordanz-Registry geladen, in die bestehenden Konkordanzen importiert sowie von Nutzer:innen generierte Mappings gespeichert werden. Zum anderen besteht eine Anbindung zu Wikidata, durch die die in Wikidata gespeicherten Mappings angezeigt und nach Verknüpfung mit einem Wikidata-Account auch erstellt, bearbeitet und gelöscht werden können.
- *Mapping Generator*: Cocoda setzt verschiedene Algorithmen für Vorschläge passender Mappings ein.
- *Implizite Mappings*: Diese Funktion leitet automatisch Zuordnungen aus Kookkurrenzen in den Katalogen ab und bietet diese als Mapping-Vorschläge an.



### 3.4.3 Bearbeitung von Mappings

Das Bearbeitungsmodul erlaubt, vorhandene Mappings zu bearbeiten und neue Mappings zu erstellen. Es bietet zudem die Möglichkeit, die Mapping-Vorschläge im Quell- und Ziel-KOS zu überprüfen, um eine bessere Übereinstimmung zu finden. Die Bedienoberfläche gestattet mehreren Personen gleichzeitig an dem Tool zu arbeiten und die Qualität der Mappings zu bestimmen.

### 3.4.4 Statistiken zur Qualitätssicherung

Das bislang nur eigenständig nutzbare Measure-Modul<sup>12</sup> erzeugt Statistiken über KOS-Arten, Häufigkeit der Nutzung der KOS und einzelner Begriffe, Informationen über Mappings pro KOS, pro Fachbereich, deren Häufigkeit, neue Einträge, Lücken in den Mappings und den Status der Mapping-Arbeit in einzelnen KOS. Die Einbindung des Moduls in Cocoda ist vorgesehen.

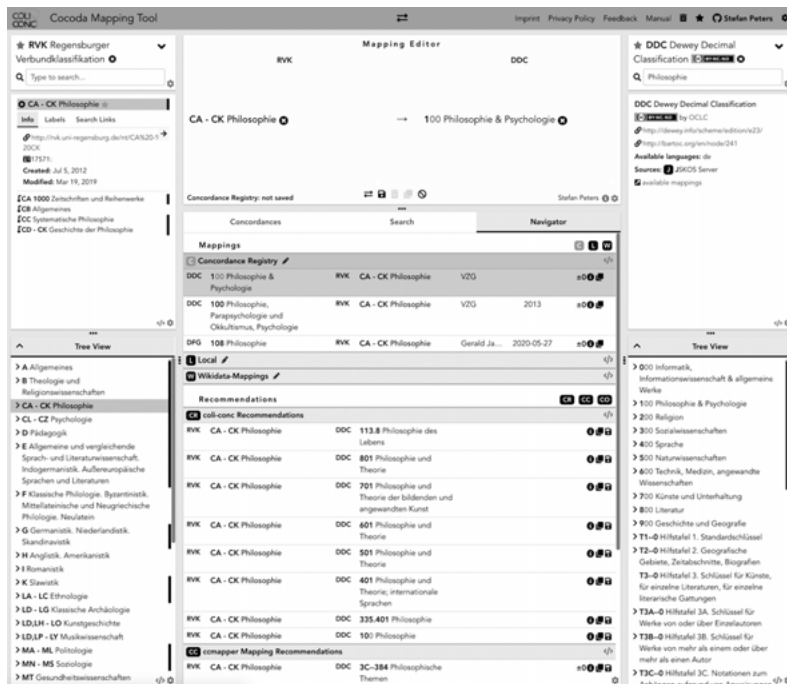


Abb. 4: Mapping-Tool Cocoda

12 <https://github.com/gbv/jskos-metrics> (28.12.2020).

### 3.5 Accountverwaltung

Zur Accountverwaltung vergibt oder speichert Cocoda keine Passwörter, sondern greift auf vorhandene Identity Provider zurück (Single Sign-on). Zur Mitarbeit bei der Erstellung und Pflege von Konkordanzen mit Cocoda reichen beispielsweise eine eigene ORCID, ein Account bei einem Projekt der Wikimedia Foundation (z. B. Wikipedia oder Wikidata) oder bei GitHub. Lediglich für erweiterte Funktionen wie die Freigabe von Mappings zur Kataloganreicherung müssen vorhandene Accounts freigeschaltet werden.

## 4 Anbindung an externe Systeme

Wie oben erwähnt sind KOS und KOS-Mappings für die strukturierte Erschließung von Dokumenten sehr nützlich. Katalogisierungssysteme und Sacherschließungssoftware können diese Daten verfügbar machen. Die modulare Softwarearchitektur des Mapping-Tools *Cocoda* und weiterer im Rahmen von *coli-conc* entwickelter Anwendungen vereinfachen sowohl die unabhängige Nachnutzung einzelner Dienste und der gewonnenen Daten als auch die Anbindung der Infrastruktur an externe Systeme. Diese Anbindung wird durch die frei zugänglichen Softwarekomponenten und die öffentlichen APIs ermöglicht. So konnten unter anderem der Normdatendienst DANTE, das Erschließungswerkzeug *easyDB*, das Mapping-Tool *CCMapper*, der Vokabular-Browser *Skosmos* und die Wissensdatenbank *Wikidata* eingebunden werden. Auch die Authentifizierung von Personen geschieht gegen externe Identity Provider wie *ORCID*, *Wikidata* und *GitHub* u. a. mittels des *OAuth*-Protokolls.

## 5 Maßnahmen zur Qualitätssicherung

Die im Rahmen von *coli-conc* entwickelten Standards und Werkzeuge bilden eine gute Grundlage zur Sammlung, Pflege und Bereitstellung von Mappings. Um deren Qualität zu gewährleisten, wurden weitere Maßnahmen ergriffen und eingeführt:

- Einbindung der Fachcommunity durch die *coli-conc*-Expertengruppe
- Ausbau von Kooperationen mit dem Normdatendienst DANTE und BARTOC (VZG), der Wikimedia Foundation (Wissensdatenbank *Wikidata*) und den Herausgeber:innen von Erschließungssystemen wie der RVK (Universitäts-

bibliothek Regensburg), DDC (OCLC) und mit der European DDC Users Group (EDUG)

- Funktionen zur Bewertung, Prüfung und Korrektur von Mappings
- Einbindung von Vorschlägen für Mappings auf der Grundlage verschiedener Verfahren und Quellen
- Dokumentation von Standards zur Beschreibung von Mappings
- Entwicklung einer offenen Registry zur Speicherung, Verwaltung und Zugriff auf Konkordanzen und Wissensorganisationssysteme in einem einheitlichen Format sowie
- Entwicklung von Verfahren zur Erzeugung von Statistiken über die Vollständigkeit, Aktualität und Korrektheit von Mappings und KOS.

Darüber hinaus wird die Integration weiterer Systeme an die coli-conc-Infrastruktur vorangetrieben. Dies beinhaltet:

- Katalogisierungs- und Sacherschließungssoftware, wie WinIBW und den Digitalen Assistenten (DA), um die direkte Nutzung und Verbesserung von Mappings bei der Katalogisierung zu gewährleisten
- Mapping-Provider wie <sameAs><sup>13</sup> und Hub<sup>14</sup>
- Authentifizierung via DFN-AAI (Shibboleth)
- Systeme zur Publikation von Wissensorganisationssystemen wie Skosmos, Skohub und Linked Open Vocabularies (LOV)

Auch die Weiterentwicklung des Import- und Update-Verfahrens für KOS und Mappings, von Schnittstellen wie coli-rich für automatische Anreicherung der Verbund- und Online-Kataloge, und Discovery-Systeme sind vorgesehen.<sup>15</sup> In Entwicklung ist außerdem die weitere Auswertung von impliziten Mappings aus den Kookkurrenzen im K10plus-Verbundkatalog sowie die Integration der Bestandteile automatisch zerlegter DDC-Notationen des Projektes coli-ana<sup>16</sup> in das Tool Cocoda.

---

**13** <http://sameas.org/> (28.12.2020).

**14** <https://hub.toolforge.org/> (28.12.2020).

**15** <https://coli-conc.gbv.de/coli-rich/> (28.12.2020).

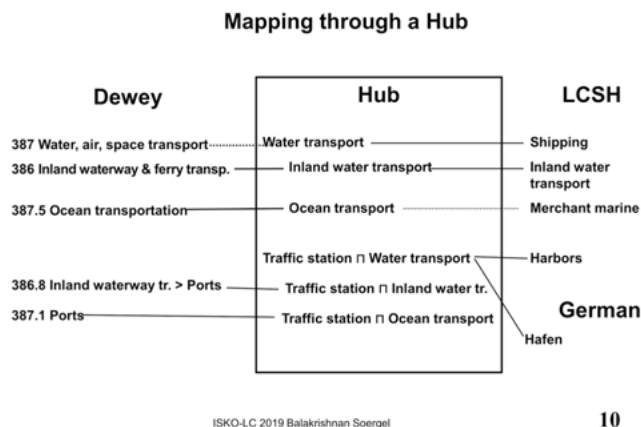
**16** <https://coli-conc.gbv.de/coli-ana/> (28.12.2020).

## 6 Entwicklung eines Mapping-Verfahrens mit dem Konzept-Hub

Im Rahmen eines Pilotprojektes wurde ein weiteres Mapping-Verfahren *Konzept-Hub* als Ergänzung zu bisherigen Verfahren prototypisch entwickelt, umgesetzt und die abgeleiteten Beziehungen mit verifizierten Beziehungen in der *coli-conc* Datenbank verglichen (Balakrishnan, Soergel und Helfer 2020).

Mit dem Konzept-Hub können sowohl das Mapping zwischen KOS als auch die Festlegung von Konzeptbeziehungen innerhalb eines KOS einfacher und besser erzielt werden (s. Abb. 5) (Soergel 2011). Im Hub werden Konzepte durch Kombinationen von Elementkonzepten ausgedrückt, genauer gesagt durch Formeln der Beschreibungslogik (description logic, DL) (Bechhofer und Goble 2001). Zwei Begriffe mit derselben DL-Formel bezeichnen wahrscheinlich dasselbe Konzept. Wenn dies nicht der Fall ist, müssen das System der Elementkonzepte und/oder das System der syntagmatischen Beziehungen, die im DL verwendet werden, verfeinert werden. Wenn die beiden Begriffe von unterschiedlichen KOS stammen, erstellt der Hub eine Zuordnung. Wenn sie aus demselben KOS stammen, sind die beiden Begriffe Synonyme.

Der Vorteil dieses Ansatzes ist, dass die Beziehungen zwischen DL-Formeln und damit auch zwischen den entsprechenden Konzepten abgeleitet werden können. Dies kann verwendet werden, um die Struktur eines KOS zu elaborieren und die Konkordanz-Beziehungen präziser und auf multilateraler Ebene zu bestimmen. Die Weiterentwicklung des Mapping-Algorithmus mittels Konzept-Hub auf einer größeren Datenmenge dient damit auch zur Qualitätssicherung.



**Abb. 5:** Mapping mit dem Konzept-Hub

## 7 Literaturverzeichnis

- Balakrishnan, Uma: Das Projekt coli-conc: Ein Bericht zur semi-automatischen Erstellung von Konkordanzen zur Dewey Dezimalklassifikation. In: VZG Aktuell (2013) Nr. 1. S. 12–16. [https://www.gbv.de/Verbundzentrale/Publikationen/broschueren/vzg-aktuell/vzg\\_aktuell\\_2013\\_01.pdf](https://www.gbv.de/Verbundzentrale/Publikationen/broschueren/vzg-aktuell/vzg_aktuell_2013_01.pdf) (28.12.2020).
- Balakrishnan, Uma, Jakob Voß und Dagobert Soergel: Towards integrated systems for KOS management, mapping, and access. Coli-conc and its collaborative computer-assisted KOS mapping tool Cocoda. In: Challenges and Opportunities for Knowledge Organization in the Digital Age. Proceedings of the Fifteenth International ISKO Conference 9–11 July 2018 Porto. Hrsg. von Fernanda Ribeiro und Maria Elisa Cerveira. Baden-Baden: Ergon 2018. S. 693–701. <https://doi.org/10.5771/9783956504211-693>.
- Balakrishnan, Uma: Infrastruktur für Erstellung, Austausch und Pflege zwischen Wissensorganisationssystemen in Anwendung auf bibliothekarische KOS, Berliner Bibliothekswissenschaftliches Kolloquium 2017. <https://rs.cms.hu-berlin.de/bbk/pages/view.php?ref=177> (14.12.2020).
- Balakrishnan, Uma und Dagobert Soergel: Concept mapping through a hub: Coli-conc pilot study. ISKO-LC 2019, Brüssel 2019. <https://doi.org/10.5281/zenodo.3257136>.
- Balakrishnan, Uma, Dagobert Soergel und Olivia Helfer: Representing Concepts through Description Logic Expressions for Knowledge Organization System (KOS) Mapping. In: Proceedings of the Sixteenth International ISKO Conference, 2020 Aalborg, Denmark. Hrsg. von Marianne Lykke, Tanja Svarre, Mette Skov und Daniel Martínez-Ávila. Baden-Baden: Ergon 2020. S. 455–459. <https://doi.org/10.5771/9783956507762-455>.
- Bechhofer, Sean und Carole Goble: Thesaurus construction through knowledge representation. In: Data & Knowledge Engineering (2001) Bd. 37 H. 1. S. 25–45.
- Hanke, Mirko: Bibliothekarische Klassifikationssysteme im semantischen Web: Zu Chancen und Problemen von Linked-Data-Repräsentationen ausgewählter Klassifikationssysteme. In: Perspektive Bibliothek (2014) Bd. 3 Nr. 2. S. 91–119.
- Hermes, Hans-Joachim: Die Konkordanz von Klassifikationen – Hat sie eine Chance? In: Erschließen, Suchen, Finden. Vorträge aus den bibliothekarischen Arbeitsgruppen der 19. und 20. Jahrestagungen (Basel 1995/1996) der Gesellschaft für Klassifikation. Hrsg. v. Hans-Joachim Hermes und Hans-Joachim Wätjen. Oldenburg: Bis 1996. S. 83–101. <http://oops.uni-oldenburg.de/675/73/herwae96.pdf> (06.01.2021).
- Heiner-Freiling, Magdha: Einführung und Nutzung der Dewey Decimal Classification (DDC) im deutschen Sprachraum. Frankfurt a. M. u. Leipzig: Deutsche Bibliothek 2000. <https://www.yumpu.com/de/document/view/8659280/einfuehrung-und-nutzung-der-dewey-decimal-ddc-deutsch> (14.12.2020).
- International Organization for Standardization: ISO 25964-2:2013 Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies. <https://www.iso.org/standard/53658.html> (6.1.2021).
- Keil, Stefan: Terminologie Mapping: Grundlagen und aktuelle Normungsvorhaben. In: Information – Wissenschaft und Praxis (2012) Bd. 63 Nr. 1. S. 45–55. <https://doi.org/10.1515/iwp-2012-0004>.
- Koch, Traugott, Heike Neuroth und Michael Day: DDC mapping report. Renardus-projekt D7.4. 2001. <http://www.ukoln.ac.uk/metadata/renardus/wp7/ddc-mapping/d74-v01.pdf> (6.1.2021).

- Lauser, Boris, Gudrun Johannsen, Caterina Caracciolo, Johannes Keizer, Willem Robert van Hage und Philipp Mayr: Comparing human and automatic thesaurus mapping approaches in the agricultural domain. In: Proc. Int'l Conf. on Dublin Core and Metadata Applications 2008. <https://arxiv.org/abs/0808.2246> (14.12.2020).
- Mayr, Philipp und Viven Petras: Building a terminology network for search: the KoMoHe project 2008. In: Proc. Int'l Conf. on Dublin Core and Metadata Applications 2008. <https://arxiv.org/abs/0808.0518> (14.12.2020).
- Mayr, Philipp: Information Retrieval-Mehrwertdienste für Digitale Bibliotheken: Crosskonkordanzen und Bradfordizing. Bonn: GESIS 2010. <https://doi.org/10.21241/ssoar.26122>.
- OCLC: Annual Report to VIAF Council. August 2014. <https://www.oclc.org/content/dam/oclc/viaf/OCLC-2014-VIAF-Annual-Report-to-VIAF-Council.pdf> (14.12.2020).
- Pfeffer, Magnus: Automatic creation of mappings between classification systems. Workshop on Classification and Subject Indexing in Library and Information Science (LIS 2013), Luxemburg (Vortrag). <https://nbn-resolving.org/urn:nbn:de:swb:90-357776> (14.12.2020).
- Reiner, Ulrike: Automatische DDC-Klassifizierung bibliografischer Titeldatensätze der Deutschen Nationalbibliografie. In: Dialog mit Bibliotheken (2010) H. 1. S. 23–29. <https://nbn-resolving.org/urn:nbn:de:101-2011012860> (14.12.2020).
- Reiner, Ulrike: Bewertung von automatisch DDC-klassifizierten Titeldatensätzen der Deutschen Nationalbibliothek (DNB). In: VZG-Colibri-Bericht 1/2008, VZG. <https://coli-conc.gbv.de/publications/colibri05.pdf> (8.1.2021).
- Reiner, Ulrike: DDC-based Search in the Data of the German National Bibliography. In: New Perspectives on Subject Indexing and Classification. Essays in Honour of Magda Heiner-Freiling. Red. von Kristina Knull-Schlomann. Leipzig und Frankfurt a. M.: Dt. Nationalbibliothek 2008. S. 121–129.
- Soergel, Dagobert: Conceptual foundations for semantic mapping and semantic search. In: Concepts in context. Proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization, July 19th–20th, 2010. Hrsg. von Felix Boteram, Winfried Gödert und Jessica Hubrich. Würzburg: Ergon 2011. S. 13–35.
- Voß, Jakob: Linking Folksonomies to Knowledge Organization Systems. In: Metadata and Semantics Research. MTSR 2012. Hrsg. von Juan Manuel Doderó, Manuel Palomo-Duarte und Pythagoras Karampiperis. Berlin, Heidelberg: Springer 2012. S. 89–97. [https://doi.org/10.1007/978-3-642-35233-1\\_9](https://doi.org/10.1007/978-3-642-35233-1_9) (14.12.2020).
- Voß, Jakob: Tagging, Folksonomy & Co – Renaissance of Manual Indexing? In: 10th International Symposium for Information Science. Hrsg. von Achim Osswald, Maximilian Stempfhuber und Christian Wolff. Konstanz 2007. S. 243–254.
- Voß, Jakob: JSKOS data format for KOS. <https://gbv.github.io/jskos/jskos.html> (14.12.2020).
- Voß, Jakob: Classification of Knowledge Organization Systems with Wikidata. In: Proceedings of the 15th European Networked Knowledge Organization Systems Workshop (NKOS 2016). Hrsg. von Philipp Mayr, Douglas Tudhope, Koraljka Golub, Christian Wartena und Ernesto William De Luca. CEUR 2016. S. 15–22. <http://ceur-ws.org/Vol-1676/paper2.pdf> (14.12.2020).
- Voß, Jakob, Andreas Ledl und Uma Balakrishnan: Uniform description and access to Knowledge Organization Systems with BARTOC and JSKOS. Terminology & Ontology: Theories and applications (TOTh), Chambéry, 9–10 June 2016. <https://doi.org/10.5281/zenodo.438019>.

Vrandečić, Denny und Markus Krötzsch: Wikidata: A Free Collaborative Knowledgebase. In: Communications of the ACM (2014), Bd. 57 Nr. 10. S. 78–85. <https://doi.org/10.1145/2629489>.

