

Projekt coli-conc

Zwischenbericht Projektzeit: 01.05.2018 - 30.04.2019

Reiner Diedrichs
Direktor der Verbundzentrale des GBV
Platz der Göttinger Sieben 1
37073 Göttingen
Tel.: 0551-39-31001
Fax: 0551-39-13980
E-Mail: diedrichs@gbv.de

Göttingen, 30.04.2019

Inhaltsverzeichnis

Tabellenverzeichnis	3
Abkürzungsverzeichnis	3
1 Allgemeine Angaben	5
2 Arbeits- und Ergebnisbericht	6
2.1 Ausgangslage und Zielsetzung	6
2.2 Projektpartner und Unterstützer	7
2.3 Erstes Projektjahr (12/2015 - 11/2016)	8
2.4 Zweites Projektjahr (12/2016 - 11/2017)	10
2.5 Drittes Projektjahr (05/2018 - 04/2019)	13
2.5.1 Arbeitspaket 2: Benutzeroberfläche	14
2.5.2 Arbeitspaket 3: Bereitstellung	15
2.5.3 Arbeitspaket 4: Mappingverfahren	15
2.5.4 Arbeitspaket 5: Qualitätssicherung	16
2.5.5 Arbeitspaket 6: Einbindung größerer Nutzergruppen	17
2.5.6 Arbeitspaket 7: Bestandsaufnahme weiterer KOS	17
2.5.7 Arbeitspaket 8: Technische Erweiterung	18
2.5.8 Arbeitspaket 10: Usability-Verbesserung	18
3 Öffentlichkeitsarbeit	19
3.1 Publikationen	19
3.1.1 Projektberichte/Präsentationen und Workshops	19
3.1.2 Fachartikel mit Peer-Review	19
3.1.3 Spezifikationen	20
3.1.4 Software	20
3.1.5 Daten	21
4 Zusammenfassung und Ausblick	21
5 Mitarbeiterinnen und Mitarbeiter	22
6 Unterschrift	22

Tabellenverzeichnis

Tab. 1: Allgemeine Angaben zum coli-conc-Projekt	5
Tab. 2: Aktueller Stand des Mittelabrufs aus der laufenden Bewilligung	6
Tab. 3: Stand der Arbeit erster Berichtszeitraum (01.12.2015 - 30.11.2016)	9
Tab. 4: Stand der Arbeit nach dem zweiten Projektjahr	11-12
Tab. 5: Aufgaben für die Fortsetzungsphase	13
Tab. 6: Fortsetzungsantrag - Zusätzlich genehmigte Arbeitspakete	14

Abkürzungsverzeichnis

AP Arbeitspaket

API Application Programming Interface

BARTOC Basel Register of Thesauri, Ontologies and Classifications

BK Basisklassifikation

DDC Dewey Dezimalklassifikation

DFG Deutsche Forschungsgemeinschaft

DNB Deutsche Nationalbibliothek

EDUG European DDC Users Group

GBV Gemeinsamer Bibliotheksverbund

GND Gemeinsame Normdatei

GVK Gemeinsamer Verbundkatalog

ISO International Organization for Standardization

JSKOS JavaScript Object Notation for Simple Knowledge Organization Systems

KK Konkordanz

KENOM Kooperative Erschließung und Nutzung der Objektdaten von Münzsammlungen

IxTheo Index Theologicus

GND4C GND für Kulturdaten

K10Plus Katalog für zehn Bundesländer, die Stiftung Preußischer Kulturbesitz und weitere Einrichtungen

KOS Knowledge Organization System / Wissensorganisationssystem

NKOS Networked Knowledge Organization Systems

OCLC Online Computer Library Center

PAIA Patrons Account Information API

RDA Research Data Alliance

RDF Resource Description Framework

RVK Regensburger Verbundklassifikation

SKOS Simple Knowledge Organization System

TIB Leibniz-Informationszentrum Technik und Naturwissenschaften Universitätsbibliothek
Hannover

UB Universitätsbibliothek

UDC Universelle Dezimalklassifikation

VZG Verbundzentrale des GBV

ZBW Deutsche Zentralbibliothek für Wirtschaftswissenschaften Kiel

1 Allgemeine Angaben

DFG-Geschäftszeichen: DI 1364/8-1

Antragsteller:

Reiner Diedrichs

Direktor der Verbundzentrale des GBV (VZG)

Platz der Göttinger Sieben 1

37073 Göttingen

Tel.: 0551 / 39-31001

Fax: 0551 / 39-13980

E-Mail: diedrichs@gbv.de

Thema des Projektes:	Mapping-Tool Cocoda - Entwicklung einer Infrastruktur für die semiautomatische Erstellung, den Austausch, und die Wartung von KOS und Konkordanzen zwischen bibliothekarischen KOS
Internetadresse des Vorhabens:	http://coli-conc.gbv.de/
Projektstart:	01.12.2015
Beantragte Förderzeit:	24 Monate + 24 Monate Fortsetzungsphase
Bewilligte Förderzeit:	24 Monate + 12 Monate Fortsetzungsphase
Erster Zwischenbericht:	01.12.2015 – 30.12.2016
Abschlussbericht erste Phase:	01.12.2015 – 30.11.2017
Aktueller Berichtszeitraum (erste Fortsetzungsphase):	01.05.2018 – 30.04.2019 (nach personalbedingter Pausierung des Projekts)
Geplante Abschlussphase:	24 Monate
Erstbewilligung der DFG:	Bewilligungsschreiben vom 05.08.2015
Derzeit laufende Bewilligung:	Bewilligungsschreiben vom 16.08.2017

Tab. 1: Allgemeine Angaben zum coli-conc-Projekt

2 Arbeits- und Ergebnisbericht

Der Bericht beschreibt den Stand der Arbeit und die erzielten Ergebnisse im dritten Projektjahr.

2.1 Ausgangslage und Zielsetzung

Das Projekt coli-conc wurde initiiert, um den Bedarf nach einer zentralen Plattform für die Erstellung von Mappings zwischen bibliothekarischen KOS und deren Sammlung zu decken und Standards und Verfahren zu schaffen, die einen Austausch, die Pflege und die Bereitstellung von Mappings vereinfachen. Um diese Idee zu verwirklichen, hat das coli-conc-Projekt sich für die DFG-Förderungszeit folgendes Ziel gesetzt:

Entwicklung einer Infrastruktur für die semi-automatische Erstellung, den Austausch und die Wartung von Konkordanzen mit Fokus auf DDC, RVK und BK. Die Maßnahmen umfassen dabei unter anderem die Entwicklung des Mapping-Tools Cocoda, die Weiterentwicklung und Umsetzung von automatischen Mapping-Verfahren und die Sammlung und Bereitstellung bereits vorhandener KOS und Konkordanzen.

Dieser Projektbericht beinhaltet die Zusammenfassung der insbesondere im dritten Projektjahr erzielten Ergebnisse des Projektes und gibt zugleich einen Ausblick auf die geplante weitere Fortsetzung des Projektes in einer abschließenden Förderphase.

2.2 Projektpartner und Unterstützer

Im dritten Projektjahr werden folgende Kooperationspartnerschaften fortgeführt bzw. neu eingegangen:

- Basel Register of Thesauri, Ontologies and Classifications (BARTOC)
- Dan Michael O. Heggø (Universitätsbibliothek Oslo) zur Entwicklung des Tools mc2skos
- Deutsche Nationalbibliothek (DNB)
- Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW)
- European DDC Users Group (EDUG)
- **GND4C (neu) – GND für Kulturdaten**
- Hochschule für Angewandte Wissenschaften Hamburg
- Hochschule Hannover
- Humboldt Universität zu Berlin Institut für Bibliotheks- und Informationswissenschaft (IBI)
- **KENOM (neu)**
- Leibniz Institute for Social Sciences (GESIS)
- Leibniz-Informationszentrum Technik und Naturwissenschaften
- Universitätsbibliothek Hannover (TIB)
- Russian State Library
- Online Computer Library Center (OCLC)
- Osma Suominen (Finnische Nationalbibliothek) zur Entwicklung des Tools Skosify
- Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (SBB-PK)
- Landesbibliothek Tirol, Österreich
- **SuUB Bremen (neu)**
- **Universitätsbibliothek Braunschweig (neu)**
- Universitätsbibliothek Technische Universität Berlin
- Universitätsbibliothek Freie Universität Berlin
- Universitätsbibliothek Konstanz
- Universitätsbibliothek Regensburg mit der RVK-Anwendergruppe
- Universitätsbibliothek St. Gallen, Schweiz
- **Universitätsbibliothek Tübingen (neu)**
- Wikimedia/Wikidata

2.3 Erstes Projektjahr (12/2015 - 11/2016)

Dieser Abschnitt fasst kurz den ersten Zwischenbericht zusammen. Im ersten Projektjahr lag der Fokus der Arbeiten auf AP 1 Bestandsaufnahme und AP 3 Bereitstellung für den Aufbau des Fundaments für die coli-conc-Infrastruktur und das Mapping-Tool Cocoda. Zu diesen Aufgaben gehörten u.a.:

- Durchführung einer Umfrage zur Ermittlung und Sammlung von den im deutschsprachigen Raum
 - aktiv genutzten KOS und deren Quellen
(<http://coli-conc.gbv.de/publications/Umfrageergebnisse.pdf>)
 - vorhandenen Mappings
 - erledigten und laufenden Mapping-Projekten und -Arbeiten
 - Mapping-Methoden und
 - Feststellung des Bedarfs an Anforderungen an und Eigenschaften von Normdatendiensten
- Erstellung von Konvertierungsskripten für verschiedene Formate ins JSKOS
- Entwicklung von Serverkomponenten für Speicherung und Bereitstellung von KOS und Mappings
- Umfassende Konvertierungsarbeiten der gesammelten
 - KOS-Metadaten
 - KOS-Daten, z.B. DDC, BK, RVK und
 - Mappingsin ein einheitliches Format JSKOS
- Weitere Anpassung des JSKOS-Formats
- Erstellung von KOS-Registry und Konkordanz-Registry mit einer Webanwendung für die Suche und das Herunterladen von KOS-Metadaten und Mappings
- Entwicklung von Schnittstellen, um einen einheitlichen Zugriff auf Inhalte der KOS DDC, RVK und BK zu gewährleisten
- Öffentlichkeitsarbeit (Vorträge) für die Einbindung größerer Nutzergruppen

Arbeitspaket	Monat	Status Projektphase (1. Jahr)
AP 1: Bestandsaufnahme	1. - 3.	Abgeschlossen
AP 2: Benutzeroberfläche	1. - 12.	Begonnen Entwicklung der Benutzeroberfläche <ul style="list-style-type: none"> • zur Navigation in KOS • Anlegen und Bearbeiten von Mappings • Mapping-Vorschläge & Bewertungen • Usability-Tests
AP 3: Bereitstellung	4. - 15.	Abgeschlossen <ul style="list-style-type: none"> • Entwicklung von Server-Komponenten zur Speicherung und Bereitstellung von KOS (KOS-DB) • Entwicklung von Server-Komponenten zur Verwaltung von Konkordanzen (KK-DB)
AP 4: Mapping-Verfahren	10. - 12.	Zum Teil abgeschlossen <ul style="list-style-type: none"> • Ermittlung und Verbesserung von vorhandenen Mapping-Verfahren für bibliothekarische Klassifikationssysteme
	13. - 18.	Ausstehend <ul style="list-style-type: none"> • Implementierung von Mapping-Verfahren/-Algorithmen
AP 5: Qualitätssicherung	19. - 24.	Ausstehend
AP 6: Einbindung größerer Nutzergruppen	5. - 9.	Begonnen <ul style="list-style-type: none"> • Dokumentation und Öffentlichkeitsarbeit zu KOS und Konkordanzen • Durchführung von Workshops mit Fachleuten und NutzerInnen
	22. - 24.	Ausstehend <ul style="list-style-type: none"> • Software-Dokumentation • Authentifizierung und Autorisierung

Tab. 3: Stand der Arbeit erster Berichtszeitraum (01.12.2015 - 30.11.2016)

2.4 Zweites Projektjahr (12/2016 - 11/2017)

Dieser Abschnitt fasst kurz den Abschlussbericht der ersten Projektphase zusammen. Der Schwerpunkt im zweiten Projektjahr lag auf AP 2 (Benutzeroberfläche) und ausstehenden Arbeiten:

- Aufbau einer Benutzeroberfläche für das Mapping-Tool „Cocoda“ mit der Auswahlmöglichkeit der KOS für Quell- und Zielsystem
- Anzeige der KOS-Daten von DDC, RVK, BK im Mapping-Tool
- Erste Implementierung des Workflows für Mapping-Vorschläge aus dem GVK-Katalog (Co-Occurrences) und Konkordanz-Registry
- Extraktion und Anreicherung von KOS-Metadaten in Zusammenarbeit mit BARTOC
- Erstellung eines Skripts zur BARTOC Metadaten-Konvertierung
- Entwicklung von zusätzlichen Skripten zur Konvertierung
- Weiterführung der Konvertierungsarbeiten, insbesondere Anpassung des Tools mc2skos für die RVK-XML Datei für ihre Konvertierung nach JSKOS
- Entwicklung von wdtaxonomy für Extraktion von Klassifikationsdaten und wdmapper in Zusammenarbeit mit der ZBW Kiel für Extraktion von Mappings aus Wikidata zur Anreicherung der Konkordanz-Registry und KOS-Registry
- Erstellung manueller Mappings zwischen der deutschen Ausgabe der BK und der niederländischen Ausgabe (NBC) zur Anreicherung des GVK und Konkordanz-Registry
- Öffentlichkeitsarbeit (Vorträge, Poster-Präsentationen)
- Anstoß zur Gründung einer coli-conc Expertengruppe für Qualitätssicherung
- Evaluation von Mapping- und Sacherschließungsprojekten mit dem Fokus auf automatischen und semi-automatischen Mapping- und Sacherschließungsverfahren, um deren Nutzbarkeit für das coli-conc-Projekt zu untersuchen.

Arbeitspaket	Monat	Status Projektphase (2. Jahr)
AP 2: Benutzeroberfläche	12. - 24.	<p>Abgeschlossen</p> <p>Cocoda-Oberfläche für:</p> <ul style="list-style-type: none"> ● Navigation, Anzeige von KOS-Daten (DDC, RVK und BK) ● Generierung von Mappings-Vorschlägen <ul style="list-style-type: none"> ○ aus Konkordanz-DB ○ von GVK-Titeldatensätzen <p>In Arbeit</p> <p>Ausbau der Oberfläche für die Anzeige von</p> <ul style="list-style-type: none"> ● weiteren Mapping-Vorschlägen ● GND-RVK-Mapping-Anzeige ● Mapping-Bewertungen/Beziehungstypen ● Funktionalitäten für <ul style="list-style-type: none"> ○ Bearbeiten und Speichern von Mappings ○ Export und Import von Mappings ○ Update-Verfahren ○ Benutzer-Authentifizierung ○ Online-Hilfe Dokumentation
AP 3: Bereitstellung	4. - 15.	<p>Abgeschlossen</p> <ul style="list-style-type: none"> ● Entwicklung von Server-Komponenten zur Speicherung und Bereitstellung von KOS (KOS-DB) ● Entwicklung von Server-Komponenten zur Verwaltung von Konkordanzen (KK-DB)
AP 4: Mapping-Verfahren	10. - 12.	<p>Teilweise abgeschlossen</p> <ul style="list-style-type: none"> ● Ermittlung und Verbesserung von vorhandenen Konkordanz-Verfahren für bibliothekarische Klassifikationssysteme
	13. - 18.	<p>Teilweise abgeschlossen</p> <ul style="list-style-type: none"> ● Mapping-Verfahren als Server-Komponenten zur Ermittlung und Bereitstellung von Mapping-Kandidaten

<p>AP 5: Qualitätsbewertung</p>	<p>19. - 24.</p>	<p>In Arbeit</p> <ul style="list-style-type: none"> ● Bestandsaufnahme vorhandener Verfahren zur Bewertung von Konkordanzen <p>Ausstehend</p> <ul style="list-style-type: none"> ● Entwicklung einer Server-Komponente zur Ermittlung und Bereitstellung von Bewertungen (KK-Measure)
<p>AP 6: Einbindung größerer Nutzergruppen</p>	<p>5. - 9.</p>	<p>Begonnen</p> <ul style="list-style-type: none"> ● Dokumentation und Öffentlichkeitsarbeit zu KOS und Konkordanzen <p>In Arbeit</p> <ul style="list-style-type: none"> ● Konzeption von Authentifizierungs- und Autorisierungsverfahren <p>Ausstehend</p> <ul style="list-style-type: none"> ● Implementierung von Authentifizierungs- und Autorisierungsverfahren
	<p>22. - 24.</p>	<p>Teilweise abgeschlossen</p> <ul style="list-style-type: none"> ● Publikation und Dokumentation der Software aus AP 2 und AP 3 <p>Ausstehend</p> <ul style="list-style-type: none"> ● Durchführung von Workshops für NutzerInnen

Tab. 4: Stand der Arbeit nach dem zweiten Projektjahr

2.5 Drittes Projektjahr (05/2018 - 04/2019)

Für das dritte Projektjahr ergab sich personalbedingt eine Verschiebung der Fortführung um fünf Monate (siehe Antrag auf Laufzeit-/ Fristverlängerung vom Mai 2018). Das dritte Projektjahr beinhaltet die Fortführung einiger Aufgaben und Tätigkeiten aus den Arbeitspaketen der ersten beiden Projektjahre (siehe Tab. 5) sowie zusätzlichen Arbeitspaketen aus dem Fortsetzungsantrag (siehe Tab. 6).

Arbeitspakete	
AP 2	Benutzeroberfläche: <ul style="list-style-type: none"> ● Fertigstellen der Anzeige für KOS und Mapping-Vorschläge (konkret: Anzeige der Mappings aus dem Zielsystem) ● Abschließen der Komponenten für Speicherung, Bearbeitung von Mappings ● Usability-Tests
AP 3	Bereitstellung: <ul style="list-style-type: none"> ● Weiterentwicklung der Server-Komponenten zur Realisierung der Anzeige der in AP2 genannten Punkte
AP 4	Mapping-Verfahren: <ul style="list-style-type: none"> ● Automatisches Mapping der Begriffssysteme
AP 5	Qualitätsbewertung: <ul style="list-style-type: none"> ● Entwicklung einer Server-Komponente zur Ermittlung und Bereitstellung von Bewertungen (KK-Measure)
AP 6	Einbindung größerer Nutzergruppen: <ul style="list-style-type: none"> ● Publikation und Dokumentation der Software aus AP 2 und AP 3 ● Durchführung von Workshops mit NutzerInnen

Tab. 5: Aufgaben für die Fortsetzungsphase

Arbeitspakete	
AP 7	Bestandsaufnahme weiterer KOS: <ul style="list-style-type: none"> • KOS-Verzeichnisse und -Dienste von Kultureinrichtungen
AP 8	Technische Erweiterung des Mapping-Tools: <ul style="list-style-type: none"> • Erweiterte Import- und Export-Funktion der KOS-Daten, Mappings • Nutzungsstatistiken • Aufbau einer Plattform für den Austausch zwischen NutzerInnen und ExpertInnen
AP 10	Usability-Verbesserung: <ul style="list-style-type: none"> • Online-Hilfe und Benutzerdokumentation • Multilinguale Anzeige (Englisch, Deutsch) • Usability-Test der Benutzeroberfläche

Tab. 6: Fortsetzungsantrag - Zusätzlich genehmigte Arbeitspakete

2.5.1 Arbeitspaket 2: Benutzeroberfläche

Wegen personalbedingtem Ausfall bei der Softwareentwicklung konnte AP 2 nicht wie geplant erledigt werden. Mit Einstellung eines neuen Entwicklers wurde entschieden, die zuvor in Java umgesetzte Benutzeroberfläche lediglich als weiteren Prototypen zu verwenden und Cocoda als Webanwendung in JavaScript neu zu implementieren. Deshalb lag im dritten Projektjahr der Schwerpunkt auf einer Neuprogrammierung und Fertigstellung der Benutzeroberfläche von Cocoda. Diese Entscheidung führte zu einer deutlich schnelleren Entwicklung, sodass trotz der Verzögerung alle wesentlichen Funktionen im Rahmen der Förderphase umgesetzt werden konnten, darunter:

- Anzeige vollständiger KOS-Daten für inzwischen mehr als 100 KOS
- Browsing und Auswahl von Klassen bzw. Konzepten in KOS
- Verbesserte Suchfunktion in KOS
- Kontextlinks in externe Systeme (RVK Online, CCMapper, Wikipedia, ...)
- Einbindung der Konkordanz-Registry mit Schreibzugriff
- Einbindung von Wikidata-Mappings
- Mapping-Editor für Bearbeiten, Speichern und Übernahme von Mapping-Kandidaten
- Benutzerauthentifizierung mittels Single-Sign-On
- Einfache Review- und Feedback-Funktion
- Einbindung von externen Empfehlungsdiensten (Recommendation Services), insbesondere CCMapper für RVK-DDC und Reconciliation Services für Mappings auf GND und Wikidata
- Verbesserung der Webanwendungen von KOS- und Konkordanz-Registry

Besonderes Augenmerk wurde bei der Neuentwicklung auf Usability gelegt (siehe AP 10). Als letzte größere Funktion steht eine Überarbeitung der Suche in allen vorhandenen Mappings aus, um eine bessere Übersicht zu bekommen, in welchen Bereichen Mappings vorhanden sind und wo Konkordanzen ergänzt bzw. überarbeitet werden müssen.

2.5.2 Arbeitspaket 3: Bereitstellung

Auf Wunsch von neuen Projektpartnern (KENOM, SuUB Bremen, UB Braunschweig) wurden weitere Haus-Systematiken konvertiert und in Cocoda zum Mapping aufgenommen. Daneben stand unter anderem eine Neu-Konvertierung der RVK an, um RVK-Schlüssel und zusammengesetzte Klassen erkennen zu können. Durch gesonderte Behandlung dieser Klassen im Suchindex konnte die Such-Performance deutlich gesteigert werden.

Die Bereitstellung von KOS-Daten als Grundlage zur Erstellung und Beurteilung von Mappings schritt auch im dritten Projektjahr weiter voran, sodass inzwischen mehr als hundert verschiedene KOS angeboten werden können.

Die Konkordanz-Registry wurde zusätzlich mit ca. 250.000 Mappings angereichert und beinhaltet derzeit ca. 385.000 Mappings. Alle Konkordanzen und Mappings können nun aus der Konkordanz-Registry in verschiedenen Formaten (JSKOS, CSV) zur Weiternutzung heruntergeladen und einzeln über eine API abgefragt werden. Dazu werden die Notationen vorhandener KOS beim Export mit den jeweiligen Benennungen angereichert. Die Bereitstellung von Wikidata-Mappings, welche nicht in der Konkordanz-Registry enthalten sind, umfasst derzeit weitere 360.000 Mappings. Mit Zusammenführung dieser und weiterer Mappings ist davon auszugehen, dass in einer weiteren Förderphase die Anzahl bereitgestellter Mappings eine Millionen übersteigen wird.

2.5.3 Arbeitspaket 4: Mappingverfahren

Ein Aufgabenschwerpunkt im zweiten Projektjahr war die Evaluierung von Mapping-Projekten mit automatischen und semi-automatischen Verfahren. Das Ziel dabei war, das Cocoda Mapping-Vorschlagssystem anzureichern und zu verbessern. Die Ergebnisse dieser Evaluation ergaben, dass es sinnvoll wäre, das Mapping-Tool CCMapper, das vor kurzem für UB Oslo für Humord-DDC Mappings entwickelt wurde, auch für Cocoda für die Generierung von RVK-DDC Mapping-Vorschläge anzupassen und einzusetzen. Um dies zu realisieren, wurde im Herbst 2018 mit der Firma Pansoft, die als einziger Anbieter weitere Anwendungen für die maschinelle Verarbeitung der DDC entwickelt (WebDewey, Dewey Search, Dewey MARCXML, ...), ein Vertrag abgeschlossen und das Mappingverfahren des CCMappers für diesen Zweck angepasst.

Insgesamt sind in Cocoda nun vier Mapping-Empfehlungs-Dienste eingebunden, die für unterschiedliche KOS implementiert sind.

- coli-conc Services (CR)
- Katalog-Co-Occurrences (CO) aus dem GVK
- CCMapper (CC) für RVK-DDC
- Reconciliation Services (GR/WR) für GND und Wikidata

Der coli-conc Mapping-Recommendation-Service erzeugt Vorschläge auf der Grundlage eines String-Matching-Algorithmus für ein Mapping von Begriffen ausgewählter Quell- und Zielsysteme.

Die Katalog-Co-Occurrences ermitteln Mapping-Vorschläge aus dem GVK auf Basis der Anzahl von Vorkommen der Sacherschließung im Katalog. Mit Zusammenlegung der Verbundkataloge des GBV und des SWB zum K10Plus-Katalog steht allerdings eine Neu-Implementierung dieses Dienstes an.

CCMapper generiert RVK-DDC-Mappings basierend auf mehreren Algorithmen, wie dies ursprünglich im Projekt coli-conc geplant wurde (siehe Bericht vom November 2017).

Die Reconciliation Services bieten Mapping-Vorschläge für GND und Wikidata auf Basis von OpenRefine Reconciliation Service APIs von lobid bzw. Wikimedia.¹

2.5.4 Arbeitspaket 5: Qualitätssicherung

Bei der Umsetzung von Maßnahmen zur Qualitätssicherung stellte sich heraus, dass diese sich weniger als abschließbares Arbeitspaket, sondern besser begleitend mit der Entwicklung des Mapping-Tools umsetzen lassen. Viele Aspekte kamen erst gegen Ende des Berichtszeitraums mit zunehmender Nutzung des Tools und beim Import weiterer KOS und Konkordanzen auf. Insbesondere kommt es bei nahezu jeder Datenlieferung zu unvorhergesehenen Fehlern und/oder Besonderheiten. Um diese schneller erkennen zu können, wurde die eingehende Validierung von JSKOS-Daten ausgebaut und mit der Erstellung von KOS- und Mapping-Statistiken begonnen. In Cocoda wurde eine Review-Funktion umgesetzt und im Rahmen von KENOM eingesetzt, um Mappings zu bewerten. Da die Anforderungen an Review-Verfahren je nach KOS und Anwenderkreis unterschiedlich sind, steht allerdings noch ein Feedback von ExpertInnen aus, um die Review-Funktionen nach ihren Wünschen zu erweitern. Grundsätzlich kann eine hohe Qualität nur durch schnelle Feedback-Zyklen erreicht werden; als weitere Maßnahmen haben wir deshalb mit der Einbindung von Funktionen aus Cocoda in andere Anwendungen wie das Katalogisierungswerkzeug WinIBW und dem Sacherschließungstool “Digitaler Assistent” begonnen.

¹ Siehe <http://blog.lobid.org/2018/08/27/openrefine.html> für die GND und <https://tools.wmflabs.org/openrefine-wikidata/> für Wikidata.

2.5.5 Arbeitspaket 6: Einbindung größerer Nutzergruppen

Es wurde ein Authentifizierungsverfahren mittels Single-Sign-On implementiert, bei dem NutzerInnen sich mit einem bestehenden Account (z.B. GitHub, ORCID, LDAP) einloggen können, sodass keine personenbezogenen Login-Daten oder Passwörter gespeichert werden müssen. Weitere Login-Provider können durch ein modulares System einfach hinzugefügt werden. Der Authentifizierungsserver ist von der JSKOS-Datenbank separiert und es werden JSON Web Tokens (JWT) genutzt, um API-Anfragen zu authentifizieren. Diese können ohne Kommunikation mit dem Authentifizierungsserver verifiziert werden. Als Weiterentwicklung ist ein Gruppen- und Rechtesystem geplant.

Der erste Workshop für Cocoda findet im Juni diesen Jahres statt. Die Zielgruppe ist FAG-EI und K10Plus SE. Im Rahmen des Workshops sollten Workflows für Qualitätssicherung und Anbindung an den Routinebetrieb überlegt bzw. mit den ExpertInnen zusammen ausgearbeitet werden. Weitere Workshops sind im August und November in der Staats- und Universitätsbibliothek zu Berlin und der UB Regensburg in Planung.

Zur direkten Nutzung und Verbesserung von Konkordanzen aus Katalogisierungswerkzeugen (WinIBW und Digitaler Assistent) stehen Schnittstellen bereit. Die vollständige Einbindung verzögert sich jedoch durch die Einführung des K10Plus-Katalogs auf die geplante Fortsetzungsphase.

2.5.6 Arbeitspaket 7: Bestandsaufnahme weiterer KOS

Die ursprünglich im ersten Folgeantrag geplante Ausweitung auf KOS, KOS-Verzeichnissen und KOS-Diensten von und für Kultureinrichtungen außerhalb des Bibliothekswesens wurde auf Anraten der Gutachter und aufgrund der kürzeren Förderphase zunächst zurückgestellt. Während der Projektlaufzeit wurde allerdings deutlich, dass ein Bedarf nach Unterstützung musealer KOS besteht. Auf konkrete Nachfragen haben wir deshalb einige KOS aus dem Kulturbereich übernommen und begonnen, Cocoda zum Mapping dieser Vokabulare anzupassen. Insbesondere wird Cocoda bereits seit Herbst 2018 für die Kooperative Erschließung und Nutzung der Objektdaten von Münzsammlungen (KENOM) eingesetzt, um eigene Vokabulare auf etablierte KOS zu mappen. Dafür wurden mit unserem Mappings-Tool bis jetzt mehr als 26.000 Mappings intellektuell erstellt. Darüber hinaus gibt es Gespräche mit dem DFG-Projekt GND4C zum Mapping verschiedener Vokabulare auf die GND.

2.5.7 Arbeitspaket 8: Technische Erweiterung

Die Umsetzung dieses AP stand wie AP 2 im Zeichen der Neu-Implementierung. Einige Komponenten konnten weiter verwendet werden, im Wesentlichen stellte es sich jedoch als effektiver heraus, die gesamte technische Infrastruktur unter Beibehaltung der bereits festgelegten Datenformate (JSKOS) und Schnittstellen (JSKOS-API, OAuth, ...) auf eine moderne Grundlage zu stellen. Insbesondere wurden eine Server-Komponente zur Verwaltung von JSKOS-Daten in einer Datenbank sowie ein Login-Server zur Umsetzung von Single-Sign-On umgesetzt.

Die im Fortsetzungsantrag geplante, aber von den Gutachten zurückgestellte Einbindung der Zerlegung von synthetischen DDC-Notationen (VZG-Projekt coli-ana) wurde aufgrund konkreter Nutzer-Anfragen (iXTheo) zumindest vorbereitet und kann in einer weiteren Fortsetzungsphase endgültig umgesetzt werden. Dies gilt ebenso für Komponenten zur Erstellung von KOS- und Mapping-Statistiken.

2.5.8 Arbeitspaket 10: Usability-Verbesserung

Im Rahmen der Entwicklung des Mapping-Tools wurde eine Online-Hilfe erstellt, welche sowohl eine kurze Anleitung für NutzerInnen des Tools anbietet, als auch als Dokumentation für die Entwicklung des Tools dient. Des Weiteren wurden an vielen Stellen im Tool kleine Tooltips hinzugefügt, die Erklärungen zu bestimmten Funktionen anbieten. Während der Entwicklung flossen zahlreiche Fragen und Hinweise von ersten AnwenderInnen in die Verbesserung der Usability ein.

Neben der Online-Hilfe gibt es ein anwendungsbezogenes Handbuch speziell zum Mapping von Vokabularen von KENOM. Außerdem erstellen wir regelmäßig Screencasts, um neue Funktionen von Cocoda anhand konkreter Beispiele zu demonstrieren.

Cocoda wird bislang in deutscher und englischer Sprache angeboten. Dabei werden auch KOS unterstützt, die beide Sprachen anbieten, bei denen entsprechend die Benennung der vom NutzerInnen ausgewählten Sprache angezeigt wird. Weitere Sprachen können nachträglich hinzugefügt werden, dazu sind allerdings noch weitere Tests notwendig. Zur Übersetzung der in Cocoda verwendeten Texte wurde ein eigenes Tool entwickelt. Speziell zur Verwendung von CCMapper für das Mapping von RVK- nach DDC-Klassen wurde ein erstes deutschsprachiges Handbuch erstellt.

3 Öffentlichkeitsarbeit

3.1 Publikationen

3.1.1 Projektberichte/Präsentationen und Workshops

Teilergebnisse des Projekts wurden zeitnah in Form von meist englischsprachigen Berichten auf der Projekthomepage und in einem Dokumentenserver veröffentlicht. Im Fortsetzungsjahr wurden bislang folgende Berichte veröffentlicht:

- Vortrag auf dem Bibliothekskongress: Nutzung und Erstellung von Normdaten-Mappings. Presentation at [7. Bibliothekskongress](#), March 20th, 2019 in Leipzig. <https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/16404>
- Werkstattbericht coli-conc: Stand des Mapping-Tools Cocoda. Presentation at [RVK Anwendertreffen](#), November 7th, 2018 in Regensburg. <https://doi.org/10.5281/zenodo.1477899>
- coli-conc report 11: [Cocoda Technical Architecture](#) (May 2018) <https://doi.org/10.5281/zenodo.1256498>
- coli-conc report 12: [Übersicht von GND-Konkordanzen](#) (November 2018) <https://doi.org/10.5281/zenodo.1689997>
- Vortrag auf ELAG (Mai 2019)
- Vortrag auf European DDC User Group (EDUG) (Mai 2019)
- coli-conc Workshop ist geplant, Ort: Verbundzentrale des GBV, Göttingen (Juni 2019)

Darüber hinaus erstellen wir seit Oktober 2018 regelmäßig Screencasts zur Demonstration von neuen Funktionen des Mapping-Tools.

3.1.2 Fachartikel mit Peer-Review

Im Berichtszeitraum erschien ein weiterer Fachartikel unter Peer-Review:

Balakrishnan, Uma; Voß, Jakob; Soergel, Dagobert: Towards integrated systems for KOS management, mapping, and access. Coli-conc and its collaborative computer-assisted KOS mapping tool Cocoda. In: Ribeiro, Fernanda; Cerveira, Maria Elisa (eds.): Challenges and Opportunities for Knowledge Organization in the Digital Age. Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto. pp. 693-701

Weitere Publikationen sind auf der coli-conc Webseite zu finden:

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

3.1.3 Spezifikationen

Das für coli-conc zentrale Datenformat JSKOS wurde im Laufe des Berichtszeitraums erweitert. Insbesondere gibt es nun JSON Schemas zur besseren Validierung von JSKOS-Daten sowie Unterstützung von Anmerkungen (Annotations) für das Review von Mappings auf Grundlage des W3C Web Annotation Data Model.

3.1.4 Software

Die im Rahmen des Projekts entwickelte Software ist vollständig als Open Source Repositories auf GitHub verfügbar. Eine vollständige Übersicht befindet sich auf der Projekt-Homepage unter <http://coli-conc.gbv.de/publications/software/>. Im vergangenen Projektjahr wurden im Wesentlichen folgende Komponenten entwickelt:

- Neuentwicklung der Cocoda-Weboberfläche in JavaScript mit dem Framework Vue.js (<https://github.com/gbv/cocoda>)
- Eine auf Node.js basierende API für den Zugriff auf JSKOS-Daten (<https://github.com/gbv/jskos-server>)
- Eine JavaScript-Library zum Arbeiten mit JSKOS-Daten, welche in vielen anderen Softwarekomponenten genutzt wird (<https://github.com/gbv/jskos-tools>)
- Eine JavaScript-Library zum Validieren von JSKOS-Daten (<https://github.com/gbv/jskos-validate>)
- Eine Server-Komponente zur Bereitstellung von Katalog-Co-Occurrences (<https://github.com/gbv/occurrences-api>)
- Ein Login-Server zur Bereitstellung von Authentifizierung (<https://github.com/gbv/login-server>)
- Eine JavaScript-Client-Library zur Nutzung des Login-Servers (<https://github.com/gbv/login-client>)
- Skripte zur Konvertierung und Nachbearbeitung von KOS-Daten in JSKOS (<https://github.com/gbv/jskos-data>)
- Skripte zur statistischen Analyse von JSKOS-Daten (<https://github.com/gbv/jskos-metrics>)
- Daten und Skripte zur Sammlung und Zusammenführung von KOS-Daten (<https://github.com/gbv/kos-registry>)
- Sammlung von Konkordanzen und Mappings sowie Konvertierungsskripte (<https://github.com/gbv/cocoda-mappings>)
- Eine Serverkomponente zum Zugriff auf Wikidata im JSKOS-Format (<https://github.com/gbv/wikidata-jskos>)
- Kommandozeilentools zur Verarbeitung von JSKOS-Daten (<https://github.com/gbv/jskos-cli>)
- Ein Webtool, welches Erstellung und Bearbeitung von Interface-Texten für Cocoda in verschiedenen Sprachen vereinfacht (<https://github.com/gbv/cocoda-locale-editor>)

Einige der Softwarekomponenten wurden auch als unabhängig nutzbare Softwarebibliotheken im Softwarerepository npm veröffentlicht.

3.1.5 Daten

Alle im Rahmen des Projekts gesammelten Daten werden unter CC0 oder einer anderen Datenlizenz zur Verfügung gestellt. Eine vollständige Liste befindet sich auf der Projekthomepage unter <http://coli-conc.gbv.de/publications/data/>.

Die Daten der RVK wurden neu konvertiert und werden nun auch für frühere Datenlieferungen in verschiedenen Formaten und Auszügen unter <https://coli-conc.gbv.de/rvk/> gesondert aufbereitet bereitgestellt.

4 Zusammenfassung und Ausblick

Das Projekt coli-conc hat bis zum Abschluss des vorliegenden Berichtszeitraums erfolgreich eine Infrastruktur entwickelt, die die semi-automatische Erstellung von Konkordanzen zwischen bibliothekarischen Wissensorganisationssystemen (KOS), deren Speicherung und den Austausch in einem einheitlichen Format unterstützt. Somit ist erstmals eine aktuelle Übersicht von bisher nur verstreut und uneinheitlich vorhandenen Konkordanzen verfügbar. Mit dem Mapping-Tool Cocoda gibt es zum ersten Mal eine frei verfügbare, offene Webanwendung für alle Arten von KOS und ihre Konkordanzen. Bislang können mehr als 100 KOS einheitlich durchsucht werden, um basierend auf automatischen Mapping-Vorschlägen intellektuelle Konkordanzen zu erstellen und zu bearbeiten. Insgesamt enthält die Konkordanz-Registry inzwischen rund 385.000 Mappings. Die Mappings können sowohl lokal als auch in einer zentralen Konkordanz-Registry gespeichert und bearbeitet werden und stehen über coli-conc zur freien Nachnutzung zur Verfügung. Dies gilt auch für die Software selbst, die sowohl als Ganzes als auch in einzelnen Komponenten unabhängig eingesetzt werden kann. Das Mapping-Tool steht unter <https://coli-conc.gbv.de/cocoda/app/> zur Verfügung und wird bereits produktiv eingesetzt.

In dem aktuellen Projektjahr lag der Fokus auf der Beendigung der Entwicklung der technischen Grundinfrastruktur von coli-conc und auf der Neu-Implementierung der Webanwendung Cocoda in JavaScript. Trotz anfänglicher Verzögerungen durch die Neubesetzung der Haupt-Entwicklerstelle zahlte sich dieses Vorgehen durch das insgesamt performantere Ergebnis und inzwischen deutlich schnellere Entwicklungszyklen aus. Trotz der Verzögerung konnten somit alle wesentlichen Funktionen im Rahmen der Förderphase umgesetzt werden. Die noch offenen Punkte betreffen vor allem Funktionen, die aufgrund der Kürzung der zweiten Förderphase zurückgestellt wurden. Weitere interne Anpassungen stehen noch aufgrund der Umstellung des Verbundkatalogs GVK zum Katalog K10plus an.

Speziell zum Mapping von RVK auf DDC wurde die Firma Pansoft beauftragt mit ihrem Tool CCMapper Mapping-Vorschläge zu erzeugen, die in Cocoda eingebunden werden.

Das durchweg positive Feedback unser Projektpartner und TestanwenderInnen zeigt den Bedarf an den Ergebnissen des Projekts, führte aber auch zu weiteren Anforderungen (Unterstützung weiterer KOS, Anbindung an weitere externe Systeme, Statistikfunktionen, Prä- und Postkombinierte Konzepte und Konzept-Vorschläge, weitere Algorithmen für Mapping-Vorschläge auch für nicht bibliothekarische KOS, Multilingualität, ...). Es wurde auch deutlich, dass die Mehrheit der potentiellen AnwenderInnen aus Bibliothekseinrichtungen interessiert auf eine endgültige Version des Mapping-Tools warten und eher persönliche Einführungen in Form von Workshops bevorzugen. Aus diesem Grund planen wir noch in diesem Jahr mehrere ganztägige Workshops vor allem für Sacherschließung-ExpertInnen. Dabei sollen auch Erweiterungen von Review-Verfahren zur Pflege und Qualitätssicherung von Mappings diskutiert werden. Es ist davon auszugehen, dass sich aus den Workshops weitere Anpassungen und Anforderungen an Cocoda ergeben.

Grundsätzlich werden die im Rahmen des Projekts entwickelte Infrastruktur und das Mapping-Tool Cocoda von der VZG und von einigen engagierten AnwenderInnen produktiv eingesetzt. Der dauerhafte Betrieb ist somit bereits jetzt sichergestellt. Zur vollständigen Entfaltung des Potentials und der Erschließung weiterer Nutzergruppen sind jedoch noch Erweiterungen notwendig, die im Rahmen einer abschließenden Förderphase bearbeitet werden sollen (siehe Fortsetzungsantrag).

5 Mitarbeiterinnen und Mitarbeiter

- Dipl. Kaufm. Reiner Diedrichs, Direktor der VZG, Projektleitung
- M. Sc. LIS Umamaheswari Balakrishnan, Projekt coli-conc, Projektkoordination
- Dr. Jakob Voß, Forschung- und Entwicklung, Techn. Koordination & Dokumentation
- BSc. Stefan Peters, Projekt coli-conc, Softwareentwicklung
- Dr.-Ing. Ulrike Reiner, Forschung- und Entwicklung, Projekt Colibri/DDC, Fachliche Beratung

6 Unterschrift

Göttingen, der 30.04.2019

Reiner Diedrichs,
Direktor der Verbundzentrale des GBV (VZG)