

Ergebnisse des Projekts coli-conc

Ein Rückblick auf mehr als 8 Jahre Sacherschließungs-Infrastruktur

Uma Balakrishnan, Jakob Voß

7. Workshop Computerunterstützte Inhaltserschließung

2023-11-14



<https://doi.org/10.5281/zenodo.10124793>



Projekt coli-conc

Entwicklung einer Infrastruktur für den Austausch, die Erstellung, und die Wartung von Konkordanzen zwischen bibliothekarischen Wissensorganisationssystemen

- Initiiert in 2012/2013 und durchgeführt durch die **Verbundzentrale des GBV (VZG)**
- **DFG-Förderung** ab 2015 bis 2023 in drei Phasen mit Pausen
(GEPRIS-ID 276843344 und 455051200)
- Bericht 2015–2019 unter <https://doi.org/10.5281/zenodo.2668204>
- Abschlussbericht 2021–2023 in Arbeit

Rückblick

Vorgeschichte

- Ab 2003 Projekte **Colibri/DDC** (Dr. Ulrike Reiner) aus Anlass der DDC-Einführung¹
- Ab 2012 Teilprojekt **coli-conc** aus Anlass der Erstellung von RVK-DDC-Mappings
 - Erstellung von Mappings mit mehreren Browser-Tabs und Excel-Tabellen
 - Sehnenscheidenentzündung
- Seit 2003² **Forschung und Entwicklung** zu Wissensmanagementsystemen (KOS)
 - Beschäftigung mit Bibliotheksdaten, Standards und Schnittstellen

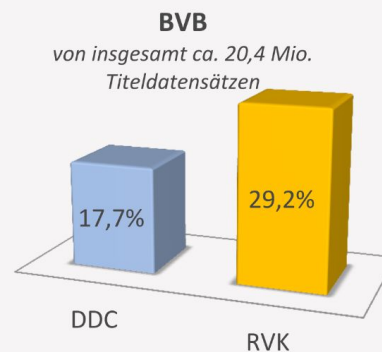
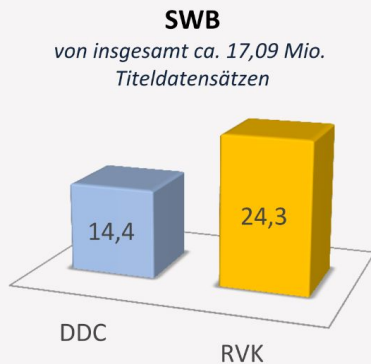
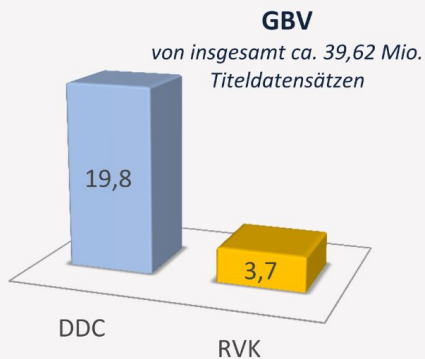
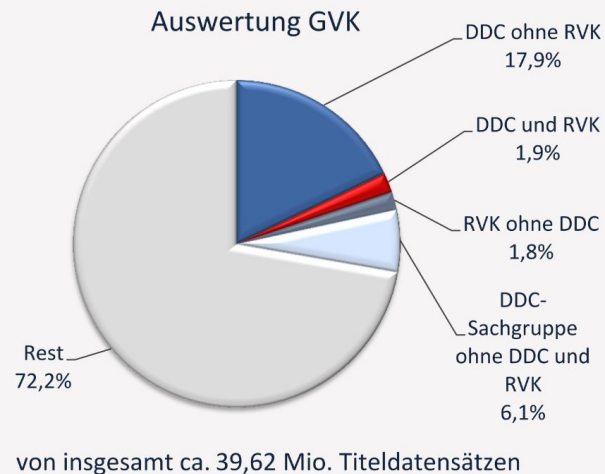
¹ <https://verbundwiki.gbv.de/download/attachments/884990/colibri01-04-03-11-without-appendix.pdf>

² <http://eprints.rclis.org/8308/>

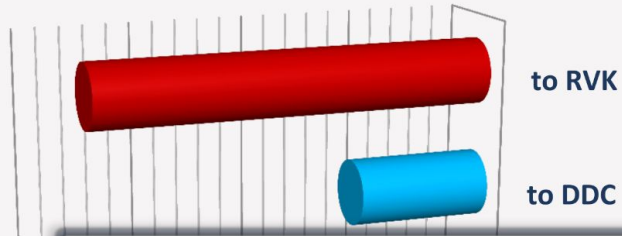
Ausgangslage

- Wenige Vokabulare (KOS) in **maschinenlesbarer Form**
 - Keine Standard-Datenformate
 - Interne Identifier ohne Kontext
 - Verantwortlichkeiten und Datenqualität unklar
- Wenige **Konkordanzen & fehlende Dokumentation** zur Erstellung von Mappings
- Notwendige **Expertenkenntnisse** zu KOS & Mappings
- Keine ausreichenden **Werkzeuge**

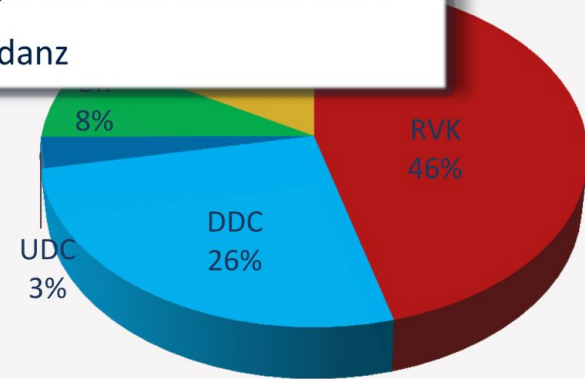
Statistik der DDC- und RVK Notationen im Verbundkatalog 2013



Umfrage 2011



Entwicklungsstand einer DDC-RVK-Konkordanz
Anwendungsbereiche und die Gründe für die Nutzung der DDC
Methode & Probleme bei Erstellung einer DDC-X Konkordanz
Interesse an einer DDC-RVK-Konkordanz



Ziele & Maßnahmen

Ziel 1: Effektivere Erstellung und Pflege von Konkordanzen zwischen bibliothekarischen KOS

- Entwicklung eines Konkordanztools
- Bereitstellung von Wissensorganisationssystemen
- Weiterentwicklung und Umsetzung von automatischen Konkordanz-Verfahren

Ziel 2: Verbesserung der Qualität der Konkordanzen

- Entwicklung und Umsetzung von Maßnahmen zur Qualitätsbewertung
- Einbindung von weiteren Nutzergruppe

Ziel 3: Förderung des Austauschs und der Nutzung von Konkordanzen und Wissensorganisationssystemen

- Sammlung vorhandener Konkordanzen und Wissensorganisationssystemen
- Bereitstellung von Konkordanzen und Wissensorganisationssystemen
- Dokumentation und Öffentlichkeitsarbeit

Ergebnisse

Sammlung von KOS-Daten und Mappings ^{AP1}

- KOS-Registry (inzwischen BARTOC)
- Mapping-Registry mit rund 500.000 gesammelten und neu erstellten Mappings
- Alle Daten über offene APIs abrufbar

Infrastruktur für KOS-Daten ^{AP3}

- JSKOS-Datenformat
- JSKOS-Server
- Wrapper zu bestehenden APIs (Skosmos, Skohub, Iobid...)
- Zahlreiche Open-Source-Komponenten

Das Mapping-Tool Cocoda AP2

The screenshot displays the Cocoda Mapping Tool interface, which is used for mapping between different classification systems. The main window is titled "Mapping-Editor" and shows a mapping between two concepts:

- RVK (Regensburger Verbundklassifikation):** WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin
- DDC (Dewey-Dezimalklassifikation):** 612.112 Leukozyten (Weiße Blutkörperchen)

The interface includes several panels:

- Left Panel (RVK):** Shows the RVK classification tree with a search bar containing "Erythrozyten". The selected concept is "WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin".
- Right Panel (DDC):** Shows the DDC classification tree with a search bar containing "Blutkörperchen". The selected concept is "612.111 Erythrozyten (Rote Blutkörperchen)".
- Mapping-Editor:** Displays the mapping between the two concepts. It shows the RVK concept on the left, the DDC concept on the right, and a "Kontordanz" (concordance) value of "keine Konkordanz".
- Bottom Panel (Mappings):** A table showing the mapping between RVK and DDC concepts. The table has columns for RVK, DDC, and other metadata.
- Bottom Panel (Vorschläge):** A section for suggestions, showing a list of RVK concepts that are related to the selected DDC concept.
- Right Panel (DDC):** Shows the DDC classification tree with a search bar containing "Blutkörperchen". The selected concept is "612.111 Erythrozyten (Rote Blutkörperchen)".

RVK	DDC	Author	Date	Concordance
WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	612.111 Erythrozyten (Rote Blutkörperchen)	Dennis Tobola	2019-03-19	±0
WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	612.112 Leukozyten (Weiße Blutkörperchen)	Stefan Peters	2019-06-05	±0
WW 8840 Allgemeines	612.111 Erythrozyten (Rote Blutkörperchen)	GESIS		±0
WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	Q211709 Blutkörperchen	Dennis Tobola	2019-03-19	+1

Unterstützung von Bibliotheksklassifikationen ^{AP14}

- Fast **alle übergreifenden Klassifikationen** werden unterstützt, insgesamt 77 Systeme
- Ausnahmen von Systemen, die im K10plus vorkommen:
 - Deutsch: ASB, Sfb, KAB
 - International: LCC, NLM, LCSH
- Lokale Systematiken nach Bedarf

Mapping-Registry in Cocoda ^{AP1-3}

RVK Regensburger Verbundklassifikation (0) PUBLICDOMAIN x

Q Tippen zum Suchen...

> **W** Biologie
⋮
↑ **WW 8720 - WW 9239** Blut und Blutbestandteile
* **WW 8840 - WW 8879** Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin +

Konzept Meta Links

Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin bevorzugt; de →
(Blutgruppen s. XD 3200) redaktionelle Anmerkung; de

✚ **WW 8840** Allgemeines
✚ **WW 8841 - WW 8844** Einzelne biologische Disziplinen
✚ **WW 8845 - WW 8879** Systematische Spezifizierung

</> ⚙

Mapping-Editor +

⋮

Konkordanzen Suche Navigator

Mappings C L W

C Konkordanz-Register ⓘ ✎ </>

RVK	WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	>	DDC	612.112 Leukozyten (Weiße Blutkörperchen)	±0 ⓘ ✎ ⓘ
RVK	WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	≈	DDC	612.111 Erythrozyten (Rote Blutkörperchen)	±0 ⓘ ✎ ⓘ
RVK	WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	=	WD	Q211709 Blutkörperchen	+1 ⓘ ✎ ⓘ

Einbindung in andere Systeme ^{AP14}

- DA-3, WinIBW
- Digitale Bibliotheken & Portale

041A/00 **f9106076612f8**USA ; ID: gnd/4078704-7
041A/01 **f9104152206f8**Politik ; ID: gnd/4046514-7
041A/02 **f9104470348f8**Kultur ; ID: gnd/4125698-0
041A/03 **fz**Geschichte
041A/09 **fADE**-101
044Z/00 **fihisfa**USA
044Z/00 **fihisfa**Geschichte
044Z/00 **fihisfa**Politik
045B/00 **fa**Enl 1

Passende Mappings

Folgende passenden Mappings sind in Cocoda vorhanden




RVK NO 3100 - NO 3640 --> BK 15.87 [von Anne Schuchardt, spezifischere Bedeutung, 2015] [von Melanie Janßen, 2022-02-28]
15.87 [von Melanie Janßen, 2022-02-21]
15 --> BK 74.25 [von Anne Schuchardt, exakte Übereinstimmung, 2016]
H sh85034755 [von DNB, hohe Übereinstimmung, 2016]
H sh85026423 [von DNB, hohe Übereinstimmung, 2016]
H sh85104466 [von DNB, hohe Übereinstimmung, 2016]
078704-7 [von UB Regensburg, 2015]




OK Abbrechen

Vorschläge Status | Rohdaten | Einstellungen #

Filtern Aktualisieren Erweitern

BK

44.41 Pharmazeutische Biologie @ddc-narrow/cc, @rvk-narrow/cc   

44.73 Geomedizin @rvk-related/cc   

Quelle: ▪ via RVK VW 6100 «Allgemeines» (Mappingtyp: verwandt) (Melanie Janßen, 2022) [\[coli-conc\]](#)

Einbindung in andere Systeme AP14

- BARTOC

Dewey Decimal Classification

About Access Content Identifiers

Access [license required](#) | [freely available](#)

License <http://creativecommons.org/licenses/by-nc-nd/3.0/> | <http://rightsstatements.org/vocab/InC/1.0/>

Format [CD-ROM](#) | [Floppy-Disc](#) | [Microform](#) | [Online](#) | [Printed](#)

Publisher [Online Computer Library Center \(OCLC\)](#)

Address 6565 Kilgour Place
Dublin
OH
43017
United States

Listed in [coli-conc KOS Registry](#)
[DANS KOSo Observatory](#)

Services (API) <https://coli-conc.gbv.de/api/> (JSKOS API) [Cocoda Mapping Tool](#)

Dewey Decimal Classification

edit

About Access Content Identifiers

Languages [ar](#) | [nb](#) | [zh](#) | [fr](#) | [de](#) | [el](#) | [he](#) | [is](#) | [it](#) | [ru](#) | [es](#) | [sv](#) | [vi](#)

Search and browsing in vocabularies registered in BARTOC is experimental. See [cocoda-sdk](#) for technical background.

Search in vocabulary

9 Geschichte & Geografie x

Top concepts

- ↑ 9 Geschichte & Geografie
- ↑ 930-990 Geschichte einzelner Kontinente, Länder, Ortschaften; außerirdischer Welten
- ↑ 940-990 Geschichte einzelner Kontinente, Länder, Ortschaften der Neuzeit; Geschichte außerirdischer Welten
- ↑ 94 Geschichte Europas

943 Deutschland und benachbarte mitteleuropäische Länder

<http://dewey.info/class/943/e23/>











Deutschland und benachbarte mitteleuropäische Länder^{de}

📅 2000-02-02 📅 2018-03-09

- ↳ 943.0001-943.0008 Standardschlüssel für Mitteleuropa
- ↳ 943.0009 Archäologie; Zeitabschnitte; Gebiete, Regionen, Orte im Allgemeinen; Sammelbiografien von Personen aus Mitteleuropa
- ↳ 943.001-943.009 Standardschlüssel für Deutschland
- ↳ 943.01-943.08 Zeitabschnitte der Geschichte Deutschlands
- ↳ 943.1 Nordostdeutschland
- ↳ 943.2 Sachsen und Thüringen
- ↳ 943.3-943.5 Orte und Gebiete von Bayern, Südwestdeutschland, Nordwestdeutschland
- ↳ 943.6 Österreich und Liechtenstein
- ↳ 943.7 Tschechien und Slowakei
- ↳ 943.8 Polen
- ↳ 943.9 Ungarn

Sacherschließungsdaten aus dem K10plus

- Occurrences per subjects API
<https://coli-conc.gbv.de/subjects/>
- Datenbanabzüge der bereinigten Sacherschließungsdaten als TSV & RDF
<https://doi.org/10.5281/zenodo.7016625>

CO Co-Occurrences		« < 1 2 3 4 > »		5 von 20 </>	
DDC	943 Deutschland und benachbarte mitteleuropäische Länder	SDNB	943 Geschichte Deutschlands	9271	 
DDC	943 Deutschland und benachbarte mitteleuropäische Länder	BK	15.43 Deutsche Geschichte des 19. Jahrhunderts bis zur Gegenwart	3618	 
DDC	943 Deutschland und benachbarte mitteleuropäische Länder	BK	15.38 Europäische Geschichte nach 1945	1631	 
DDC	943 Deutschland und benachbarte mitteleuropäische Länder	GND	Deutschland	1508	 
DDC	943 Deutschland und benachbarte mitteleuropäische Länder	BK	15.48 Sachsen, Sachsen-Anhalt, Thüringen	979	 

Workflows zur Qualitätssicherung ^{AP16}

- Annotation von Mappings auf verschiedene Weise
- Flexible Workflows möglich, aber kaum genutzt

The screenshot shows a web interface titled 'Konkordanz-Register'. It displays a table with four rows of mapping data. A modal window is open over the second row, showing a confirmation step for a change made by 'David-Benjamin Rohrer' on '2023-05-26'. The modal includes a '+1' score, thumbs up/down icons, and a 'Bestätigen' button.

Code	Label	Relation	Value	Category	Author	Date	Score	Actions
RVK	YC 6200 - YC 6299 Stoffwechsel- und Ernährungskrankheiten, allgemein	~	BK	44.76 Ernährungsstörung Mangelkrankheiten			+1	2023-05-26 David-Benjamin Rohrer
RVK	ZP 3400 - ZP 3499 Wasserkraft	<	BK	52.56 Regenerative Energieformen, alternative Energieformen			+1	
DDC	823 Englische Erzählprosa	<	BK	18.03 Englische Sprache und Literatur	JakobVoss	2023-05-09	±0	
RVK	QR 870 Container, Kombinierter Verkehr, Multimodaler Verkehr	<	BK	83.72 Verkehrswirtschaft	Melanie J...	2023-05-05	±0	

Mehrstufige Mapping-Vorschläge ^{AP17}

Vorschläge				CO	CR	CRe	CC
CO Co-Occurrences				« < 1 <u>2</u> 3 4 5 > »		→ 2 von 10 </>	
RVK	WW 8844 Homo	DDC	612.111 Erythrozyten (Rote Blutkörperchen)	16			
BK	44.86 Hämatologie	DDC	612.111 Erythrozyten (Rote Blutkörperchen)	13			
CR coli-conc Vorschläge				« < 1 <u>2</u> 3 > »		2 von 5 </>	
RVK	WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	DDC	612.111 Erythrozyten (Rote Blutkörperchen)				
RVK	WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	DDC	612.112 Leukozyten (Weiße Blutkörperchen)				
CC ccmapper Mapping-Vorschläge				« < 1 2 <u>3</u> 4 ... > »		→ 2 von 20 </>	
RVK	WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	DDC	612.1121 Biochemie				
RVK	WW 8840 - WW 8879 Blutkörperchen (Erythrozyt, Leukozyt), Hämoglobin	DDC	571.96 Immunität				

Workflows zur Qualitätssicherung ^{AP16}

- **Leitfäden** zur Erstellung von Mappings mit Beispielen
 - RVK nach BK
 - RVK nach DDC
 - Wikidata nach BK und GND
- <https://coli-conc.gbv.de/publications/#mapping-guidelines>

DDC-Zerlegung mit coli-ana ^{AP13}

- Webanwendung und API
- Einbindung in Cocoda und DA-3

Info) DDC	#
Zurück	
615.3210916 [coli-ana]	
6----- Technik, Medizin, angewandte Wissenschaften (600)	
61----- Medizin & Gesundheit (610)	
615----- Pharmakologie und Therapeutik (615)	
615.3----- Einzelne Arzneistoffe und Arzneigruppen (615.2-615.3)	
615.3----- Organische Arzneistoffe (615.3)	
615.32----- Aus Pflanzen und Mikroorganismen gewonnene Substanzen (615.32)	
615.321----- Aus Pflanzen gewonnene Substanzen (615.321)	
---.---0--- Hilfstafel 1. Standardschlüssel (T1--0)	
---.---09-- Geschichte, geografische Behandlung, Biografien (T1--09)	
---.---091- Gebiete, Regionen, Orte im Allgemeinen (T1--091)	
---.-----6 Zonale, naturräumliche, sozioökonomische Regionen (T2--11-17)	
---.-----6 Luft und Gewässer (T2--16)	

DDC Analysis

641.50902 analyze Language: **Deutsch, Norsk**



Examples: 700.23, 700.90440747471, 666.444, 555.55, 641.50902, 341.4202664304409443835, 869.1309150981

641.50902 {} Q ...

- 6----- Technik, Medizin, angewandte Wissenschaften (600) Q ...
- ↳ 64----- Hauswirtschaft & Familie (640) Q ...
- ↳ 641----- Essen und Trinken (641) Q ...
- ↳ 641.5---- Kochen (641.5) Q ...
- ↳ 641.509-- Geschichte, geografische Behandlung, Biografien (641.509) Q ...
- .---0--- Hilfstafel 1. Standardschlüssel (T1--0) Q ...
- ↳ ----.---09-- Geschichte, geografische Behandlung, Biografien (T1--09) Q ...
- ↳ ----.---0902 Zeitabschnitte (T1--0901-T1--0905) Q ...
- ↳ ----.---0902 *6.-15. Jahrhundert, 500-1499 (T1--0902) Q ...

Eintragung von Anreicherung (coli-rich) ^{AP14}

- Modul & Verfahren für automatische und semi-automatische Anreicherung coli-rich
- **Anreicherung von über 1 Mio. Titeldatensätzen mit BK Mappings**

K10plusPPN:	1756577099 
Titel:	Diagnostic pathology : forensic autopsy / Mary Ann Sens, MD, PhD, Rhome Hughes, MD, MS
Autorin/Autor:	Sens, Mary Ann [Verfasserin/Verfasser]
Beteiligt:	Hughes, Rhome [Verfasserin/Verfasser]
Erschienen:	Philadelphia, PA : Elsevier, [2021] [© 2021]
Umfang:	xxi, 571, xxxv Seiten : Illustrationen, Diagramme ; 28 cm
Sprache(n):	Englisch
Weitere Titel:	Forensic autopsy
Anmerkung:	Literaturangaben. - Auf dem Umschlag: Enhanced digital version included.
Bibliogr. Zusammenhang:	Erscheint auch als: Diagnostic Pathology / Mary Ann Sens (Online-Ausgabe)
ISBN:	978-0-323-75617-4
Sonstige Nummern:	OCOLC: 1277170607 
RVK-Notation:	XL 1600 info XL 2101 info XL 2103 info → Ähnliche Literatur
Sachgebiete:	Basisklassifikation: 44.72 (Rechtsmedizin)
Sonstige Schlagwörter:	MeSH: Forensic Pathology ; Autopsy
Inhaltliche Zusammenfassung:	Intro -- Title Page -- Dedication -- Introduction -- 1. So you want to be a forensic pathologist? -- 2. My last living patient -- 3. A deaths -- 7. Every contact leaves a trace -- 8. Risking life for death -- 9. Ways of dying, part I -- 10. Ways of dying, part II -- 11. 15. On equality in death -- 16. The Lindy Effect and forensics -- 17. Ten lessons for the living from the dead -- 18. On death an
044C	\$N650\$S\$aForensic Pathology
044C	\$N650\$S\$aAutopsy
045C	\$aW 825
045Q/01	\$9106409476\$844.72 (Rechtsmedizin)\$Acoli-conc RVK->BK\$Ahttps://coli-conc.gbv.de/api/mappings/d415aba4-14c2-4a9c-822a-1a589787545d
045R	\$91271582376\$8XL 1600 : X-Y =Medizin/XL =Rechtsmedizin/XL 1400-XL 6099 =Allgemeine Rechtsmedizin/XL 1600-XL 2199 =Allgemeine recht

Mehrsprachigkeit und Usability ^{AP15 & AP17}

- Benutzeroberfläche in Deutsch und Englisch (weitere nach Bedarf)
- Kontextsensitive Hilfe, Screencasts, Dokumentation...
- Vollständige Unterstützung mehrsprachiger Vokabulare

<p>▲3185 Handelsbeziehungen</p> <p>✕ 1567 Handelsabkommen ★ (+)</p> <p>Konzept Übersetzungen Meta Links</p> <p>Handelsabkommen <small>bevorzugt; de</small></p> <p>trade agreement <small>bevorzugt; en</small></p> <p>Handelsverhandlung <small>alternativ; de</small></p> <p>trade treaty <small>alternativ; en</small></p> <p>trade negotiations <small>alternativ; en</small></p>	<p>▲3185 Handelsbeziehungen</p> <p>✕ 1567 Handelsabkommen ★ (+)</p> <p>Konzept Übersetzungen Meta Links</p> <p>handelsavtal <small>bevorzugt; sv</small></p> <p>prekybos susitarimas <small>bevorzugt; lt</small></p> <p>accordo commerciale <small>bevorzugt; it</small></p> <p>trgovinski sporazum <small>bevorzugt; hr, sl</small></p> <p>obchodná dohoda <small>bevorzugt; sk</small></p> <p>εμπορική συμφωνία <small>bevorzugt; el</small></p>
--	--

Nutzereinbindung & Crowdsourcing ^{AP6 & AP18}

- Erstellung und Bewertung von Mappings steht direkt allen offen
- Single-Sign-On via ORCID, GitHub u.A.
- Zahlreiche Vorträge und Workshops
- Meinungen über die erreichbare Qualität gehen auseinander
- Keine dedizierte Werbung über Fachkreise hinaus

Wie geht es weiter?

- Stabile Infrastruktur & Betrieb sichergestellt
- Kostenfreie Nutzung der gesamten Infrastruktur
- Schulung und Unterstützung bei der Erstellung von Mappings
- Bereitstellung von Daten und Statistiken
- Regelmäßige automatische Anreicherung des K10plus
- Lieferung von Mappings an externe Dienste
- Regelmäßige Aktualisierung der KOS
- Redaktionelle Arbeit

Anmerkung: Für die kontinuierliche Erstellung und Qualitätssicherung von Mappings zwischen RVK-BK, DDC-BK und DDC-RVK sind dauerhaft Ressourcen notwendig

Ausblick

- Förderantrag zum Ausbau von BARTOC und KOS-Infrastruktur für Forschungsdaten
- Unterstützung von KOS-Management
- Webanwendung coli-rich für semi-automatische Anreicherung

Lessons Learned

- **Hofstadters Gesetz:** Es dauert immer länger als gedacht, selbst unter Berücksichtigung von Hofstadters Gesetz.
- **Digitalisierung von Sacherschließung** ist noch immer nicht abgeschlossen.
- **Konkordanzen** sind gerne gesehen, haben aber selten Priorität.
- **Vokabulare und Werkzeuge** sind immer nur Angebote.

Statistik der DDC- und RVK Notationen im Verbundkatalog 2013

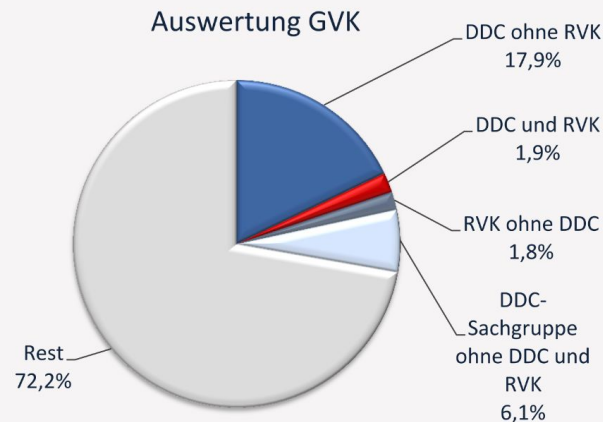
Statistik 9/2023 für den K10plus

(ca. 73 Mio Titeldatensätze mit Bestand):

DDC: 14.7%

RVK: 9.9%

DDC/RVK/BK/GND/STW...: 31.5%



von insgesamt ca. 39,62 Mio. Titeldatensätzen

