

# Datenorientierte Inhaltserschließung: Entwicklungen und Empfehlungen

Jakob Voß

Verbundzentrale des GBV (VZG)

<https://doi.org/10.5281/zenodo.8008918>

# Inhaltserschließung vs. Formalerschließung

## Inhaltserschließung

- Inhalt einer Publikation
- Themen, Zusammenfassungen, Fachgebiete...
- Reduktion von vollständigem Inhalt (Volltext)

## Formalerschließung

- Objektive Eigenschaften
- Titel, Medium, Umfang, URL, Autor\*innen...

# Datenorientierte Inhaltserschließung

Reduzierter Inhalt einer Publikation in **maschinenlesbarer Form**

- Hier keine Freitexte (Zusammenfassung, freie Schlagwörter...)  
Diese sind zwar auch hilfreich aber eher wie Volltexte zu behandeln

Inhaltserschließung durch ***kontrollierte Vokabulare, Klassifikationen, Thesauri...***

- Universalklassifikationen (BK, RVK, DDC...)
- Normdateien (GND, LCSH...)
- Fach und Lokalsystematiken (STW, NSK...)
- Ontologien für Forschungsdaten

# Vorteile Kontrollierter Inhaltserschließung

- Kontrolle von Homonymen & Synonymen
- Mehrsprachigkeit und Internationalität
- Hierarchien und Verknüpfungen
- Themenseiten, Knowledge Graphs...

11	<a href="#">1900-1999</a>
14	<a href="#">19th century</a>
10	<a href="#">2&amp;#8208;Childhood</a>
6337	<a href="#">2. Schuljahr</a>
1675	<a href="#">2. Weltkrieg</a>
35900	<a href="#">20. Jahrhundert</a>
34	<a href="#">20. Juli 1944</a>
8	<a href="#">20. Jahrhundert</a>
60	<a href="#">20er Jahre</a>
18	<a href="#">20th century</a>

<https://www.fachportal-paedagogik.de/literatur/schlagwortregister.html>

# Vokabulare und Mappings

# Voraussetzungen



- Jedes Vokabular muss in **BARTOC** eingetragen sein
- Einträge des Vokabulars haben **Identifizier**
- Identifizier haben eine Form als **URI** (können wir einrichten)
- Vokabular ist idealerweise **maschinenlesbar** als Open Data verfügbar

Beispiel: *Basisklassifikation (BK)*

- <http://bartoc.org/en/node/18785>
- **83.71** Handel
- <http://uri.gbv.de/terminology/bk/83.71>

# Dienst: Infos über Vokabulare



- **Recherche** von Vokabularen & Eintragung in BARTOC
  - Wer ist für jeweiliges Vokabular verantwortlich?
  - Wo wird das Vokabular genutzt?
  - Wie kann auf das Vokabular zugegriffen werden?
  - ...
  
- Zunehmend relevant im Bereich Forschungsdaten und Datenintegration
- Ausbau geplant im Rahmen eines beantragten DFG-Projekt  
<https://doi.org/10.5281/zenodo.7673393>

# Dienst: Konvertierung und Hosting von Vokabularen

- **Konvertierung** von Vokabularen in maschinenlesbare Form  
z.B. Excel/CSV → JSKOS/RDF
- **Hosting** der Vokabulare und Bereitstellung per APIs
- Bisher nur bei konkretem Bedarf
- VZG: Vokabular-Management per DANTE (bisher vor allem Museumsbereich)
- Besser: Hosting durch Vokabular-Herausgeber oder andere Terminology Services
  - Erfolgreiche Beispiele: UB Konstanz und UB Bielefeld
  - Terminology Software siehe <https://bartoc.org/software>
  - Terminology Services siehe <https://bartoc.org/registries>

# Anwendung: Cocoda

The screenshot displays the Cocoda Mapping Tool interface. The main window is titled "Mapping-Editor" and shows a mapping between "SR 760 Software-Unternehmen" (RVK) and "54.04 Ausbildung, Beruf, Organisationen" (BK). The interface is divided into several panels:

- Left Panel:** "RVK Regensburger Verbundklassifikation" with a search bar containing "Ausbildung". Below it is a list of categories, with "SR 760 Software-Unternehmen" selected. A "Hierarchie" section shows a tree structure of categories.
- Top Panel:** "BK Basisklassifikation" with a search bar containing "Software". Below it is a list of categories, with "54.04 Ausbildung, Beruf, Organisationen" selected. A "Hierarchie" section shows a tree structure of categories.
- Center Panel:** "Mapping-Editor" showing the mapping between the selected RVK and BK terms. It includes a "Konkordanz-Register" table and a "Vorschläge" section.
- Right Panel:** "BK Basisklassifikation" showing a list of categories, with "54.04 Ausbildung, Beruf, Organisationen" selected. A "Hierarchie" section shows a tree structure of categories.

The "Konkordanz-Register" table shows the following data:

RVK	BK	Mapping	Author	Date	Actions
SR 760 Software-Unternehmen	BK 54.04 Ausbildung, Beruf, Organisationen	David-Benja...	2022-05-25	+1   [ ]   [ ]	

The "Vorschläge" section shows the following data:

RVK	BK	Mapping	Count	Actions
SR 760 Software-Unternehmen	BK 85.06 Unternehmensführung		17	[ ]   [ ]
SR 760 Software-Unternehmen	BK 05.38 Neue elektronische Medien		13	[ ]   [ ]
SR 760 Software-Unternehmen	BK 54.01 Geschichte der Informatik		11	[ ]   [ ]

# Hauptfunktionen von Cocoda

- **Einbindung beliebiger KOS** mit einheitlichem Browsing und Suche
- **Deep-Links** zu Sacherschließung im K10plus-Katalog
- **Suche** und Auswertung vorhandener Mappings
- **Mapping-Vorschläge** mit verschiedenen Verfahren
- Vergabe von **Relationen**: = exact,  $\approx$  close, < narrower, > broader,  $\sim$  related
- **Speicherung und Änderung** eigener und gemeinsamer Mappings
- **Bewertung** von Mappings
- **Feedback Funktion für Korrekturen**

# Schnittstellen

## Zugriff auf Erschließungsdaten

# Vorhandene Schnittstellen für den K10plus-Katalog

<https://wiki.k10plus.de/display/K10PLUS/Schnittstellen>

- **SRU:** Suche per CQL-Suchsyntax
- **unAPI:** Abruf einzelner Datensätze

## Nachteile

- Rückgabe in Bibliotheksformaten PICA und MARC
- Sacherschließungsdaten über verschiedene Felder verteilt
- Unsaubere Daten

# JSKOS-Datenformat

- Vereinheitlichtes Format für kontrollierte Vokabulare (SKOS, OWL, GND, MeSH...)
- SKOS-Datenmodell + Erweiterungen
- JSON, auch auch RDF via JSON-LD

<https://gbv.github.io/jskos/>

## Beispiel

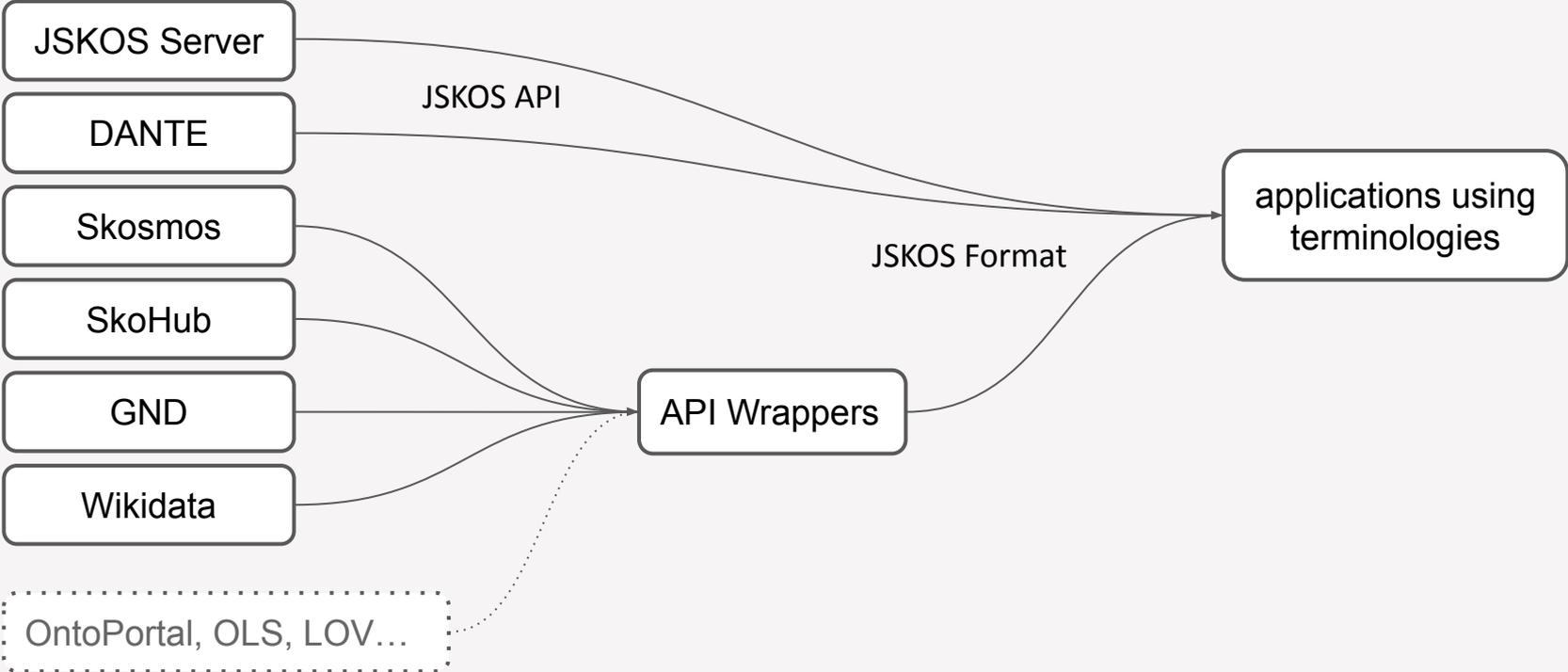
```
{
  "uri": "http://example.org/animals/",
  "notation": ["A"],
  "prefLabel": {
    "en": "animals",
    "de": "Tiere"
  }
}
```

# Unsere Grundlage: JSKOS-Format und JSKOS-API

- Zugriff auf Vokabulare und Mappings
- Gleiche API für mehrere Datenbanken
  - **BARTOC**: Informationen über Vokabulare  
<https://bartoc.org/api/>
  - **DANTE**: Ausgewählte Vokabulare für Digitale Sammlungen  
<https://api.dante.gbv.de/>
  - **coli-conc** : Ausgewählte Vokabulare und Mappings  
<https://coli-conc.gbv.de/api/>
  - ...
- Einbindung von Vokabularen und Mappings in weitere Anwendungen (DA-3, WinIBW...)
- API-Wrapper für weitere Datenquellen

- **A** Allgemeines
- **B** Theologie und Religionswissenschaften
- ✓ **CA - CK** Philosophie
  - CA 1000** Zeitschriften und Reihenwerke
- **CB** Allgemeines
- ✓ **CC** Systematische Philosophie
  - CC 1000** Versuche zur Grundlegung der Philosophie, Gesamtsysteme
  - CC 1100** Untersuchungen zur Philosophie
  - CC 1200** Geschichte und Untersuchungen philosophischer Methoden
- ✓ **CC 2000 - CC 2600** Logik
  - CC 2000** Geschichte der Logik
  - CC 2200** Werke aus dem Gedankenkreis der traditionellen Logik
  - CC 2400** Einführungen, Handbücher, Sammelbände und Bibliografien

# Federated terminology architecture used by VZG



# Neu: Occurrences-API

- <https://coli-conc.gbv.de/occurrences/>
- **Occurrences**
  - Anzahl der mit einer Klasse erschlossenen Titel
  - Link auf Titellisten im K10plus-OPAC
  - <https://coli-conc.gbv.de/occurrences/api/?member=http:%2F%2Frvk.uni-regensburg.de%2Fnt%2FQR%2520544>
- **Co-Occurrences**
  - Häufigste gemeinsam vergebenen Notationen

## CO Co-Occurrences

RVK	QR 544 Nahrungsmittelleinzelhandel	BK	85.40 Marketing	35
RVK	QR 544 Nahrungsmittelleinzelhandel	BK	83.71 Handel	33
RVK	QR 544 Nahrungsmittelleinzelhandel	BK	83.66 Agrarwirtschaft	11

# Neu: Mapping-Inference API

- Ermittlung herleitbarer Sacherschließung auf Grundlage von Mappings (**coli-rich**)

## Beispiel

- **RVK QR 544 Nahrungsmittelleinzelhandel** ⇒ **BK 83.71 Handel**  
Weil **RVK QR 540 - QR 544 Handelsgewerbe** > broader match **BK 83.71 Handel**  
<https://coli-conc.gbv.de/api/mappings/infer?from=http:%2F%2Frvk.uni-regensburg.de%2Fnt%2FQR%2520544&fromScheme=http://bartoc.org/en/node/533&toScheme=http://bartoc.org/en/node/18785>

## Mapping-Info

Quellvokabular: RVK Regensburger Verbundklassifikation

Quellkonzept: QR 540 - QR 544 Handelsgewerbe

Zielvokabular: BK Basisklassifikation

Zielkonzept: 83.71 Handel

Mappingtyp: spezifischere Bedeutung (<)

Erstellt von: Melanie Janßen

Erstellt: 2023-05-05

Teil von Konkordanz: [Q Wirtschaftswissenschaften](#) von Melanie Janßen (in [Konkordanzregister öffnen](#))

Identifizier: <https://coli-conc.gbv.de/api/mappings/42dd4f21-a6a2-4701-87df-19bba58d2018> (URI)

urn:jskos:mapping:content:c93a1ac5c1369c4f5066526f12b58a0af5c3337f

urn:jskos:mapping:members:358cc00d97cd8ee00068cd31c923758437ad7e3c

Anreicherung im K10plus <https://opac.k10plus.de/DB=2.299/CMD?>

suchen: <ACT=SRCHA&IKT=8659&TRM=https+coli+conc+gbv+de+api+mappings+42dd4f21+a6a2+4701+87df+19bba58d2018>

Download: [.csv](#) [.tsv](#) [.json](#)

Feedback: [Dieses Mapping in Cocoda bewerten](#) oder [Feedback per E-Mail senden](#) (coli-conc@gbv.de)

# Web-Ansicht (JSKOS Viewer)

## Kontext-Info

[In Cocoda öffnen](#)

### Quelle (RVK)

QR 540 - QR 544 Handelsgewerbe

┆ Q Wirtschaftswissenschaften

┆ QR Gewerbepolitik. Einzelne Wirtschaftszweige

┆ QR 500 - QR 870 Industrie, Bergbau, Handel, Dienstleistungen, Handwerk

↳ QR 540 Allgemeines

↳ QR 541 Geschichte

↳ QR 544 Nahrungsmittel Einzelhandel

[Info](#) [Bezeichnungen](#) [Editorial](#)

### Ziel (BK)

83.71 Handel

┆ 7-8 Sozialwissenschaften

┆ 83 Volkswirtschaft

┆ 83.00 Volkswirtschaft: Allgemeines

┆ 83.60 Wirtschaftssektoren: Allgemeines

┆ 83.68 Dienstleistungen: Allgemeines

[Info](#) [Bezeichnungen](#)

<http://uri.gbv.de/terminology/bk/83.71> (URI)

# In Arbeit: Subjects API

- **Abfrage der Sacherschließung** zu einem Titel
- **Änderung der Sacherschließung** per API und Webanwendung (ohne WinIBW!)

## Vorteile

- Keine PICA/MARC-Kenntnisse notwendig, nur inhaltliches Fachwissen
- Nicht nur für K10plus, auch Repositories, Forschungsdaten, andere Verbünde...

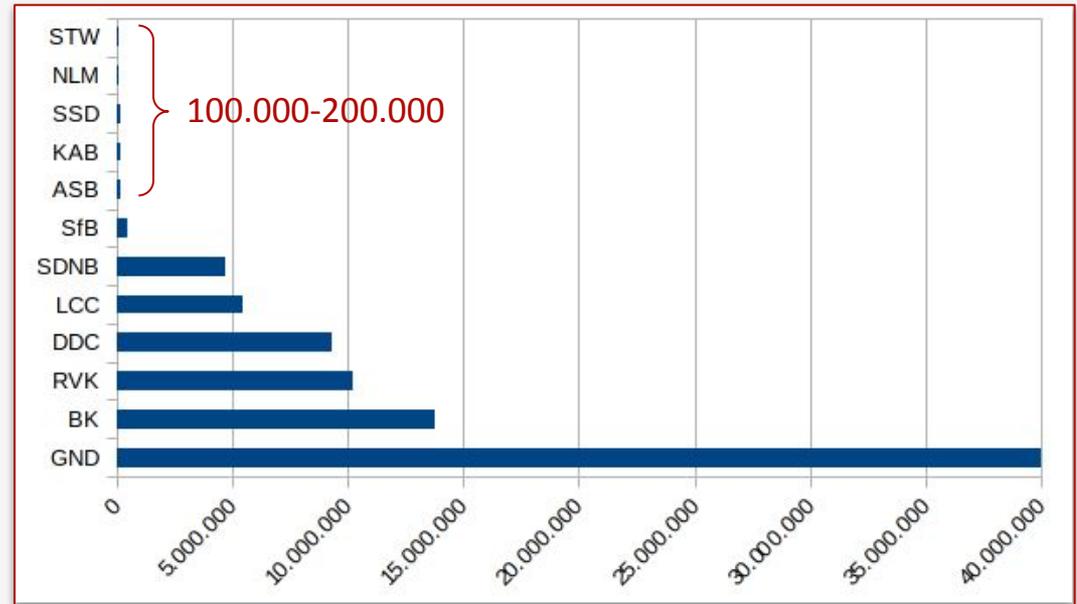
# Analyse und Korrekturen vorhandener Erschließung

# Bereitstellung von Datenabzügen

- **Sacherschließungsfelder** aus dem K10plus
- Regelmäßige Publikation als Open Data unter <https://doi.org/10.5281/zenodo.7016625>
  - Tabellarisches Format (TSV-Datei)
  - **RDF-Format** (NTripel)
- **Datenbereinigung**
  - Nur syntaktisch korrekte Identifier (Beispiel BK: [0-9][0-9]\.[0-9][0-9])
  - Geplant: Nur tatsächlich vorhandene Normdatensätze/Klassen
  - Konvertierung von Identifier in URIs (83.71 Handel => <http://uri.gbv.de/terminology/bk/83.71>)
- Datenbasis für per Occurrences API und Subjects API

# Unterstützte Vokabulare

- Voraussetzungen wie beschrieben (BARTOC & URIs): bislang zwölf Vokabulare
- 32% der K10plus-Titel haben eine dieser Erschließungen
- $\emptyset$  3,5 Einträge pro Titel



# Statistiken und Qualitätsmanagement

- **QA Catalogue** im Testbetrieb für den gesamten K10plus
- Ausschnitte für einzelne Bibliotheken möglich
- Fehlerlisten (PPN) zur Erstellung von Korrekturlisten

## Vollständigkeit

Alle Felder und Unterfelder mit der jeweiligen Anzahl und dem Anteil ihres Vorkommens

Bibliothek:

Dokumenttyp: [all](#) · [Druckschriften \(einschließlich Bildbänden\)](#) · [Elektronische Ressource im Fernzugriff](#) · [Tonträger, Videodatenträger, Bildliche Darstellungen](#) · [Elektronische Ressource auf Datenträger](#) · [Mikroform](#) · [Medienkombination](#) · [Handschriftliches Material](#) · [Objekt](#) · [Blindenschriftträger und andere taktile Materialien](#)

Datensätze: **71,757,790**

## Feldgruppen

Felder	Anzahl	%
0... PICA+ bibliographic description	71,757,790	100.00
unknown unknown origin	3,840,699	5.35

## Felder

	Datensätze		Vorkommen			
	Anzahl	%	Anzahl	von bis	ø	σ
<b>PICA+ bibliographic description</b>						
001A=0200 — Kennung und Datum der Ersterfassung	71,757,790	100.00	71,757,790	1		
001B=0210 — Kennung und Datum der letzten Änderung	71,757,790	100.00	71,758,521	1	2	1,000,00
001Q=000Q — VAL-Warnungen	439	0.00	444	1	3	1,010,01

# Anreicherung

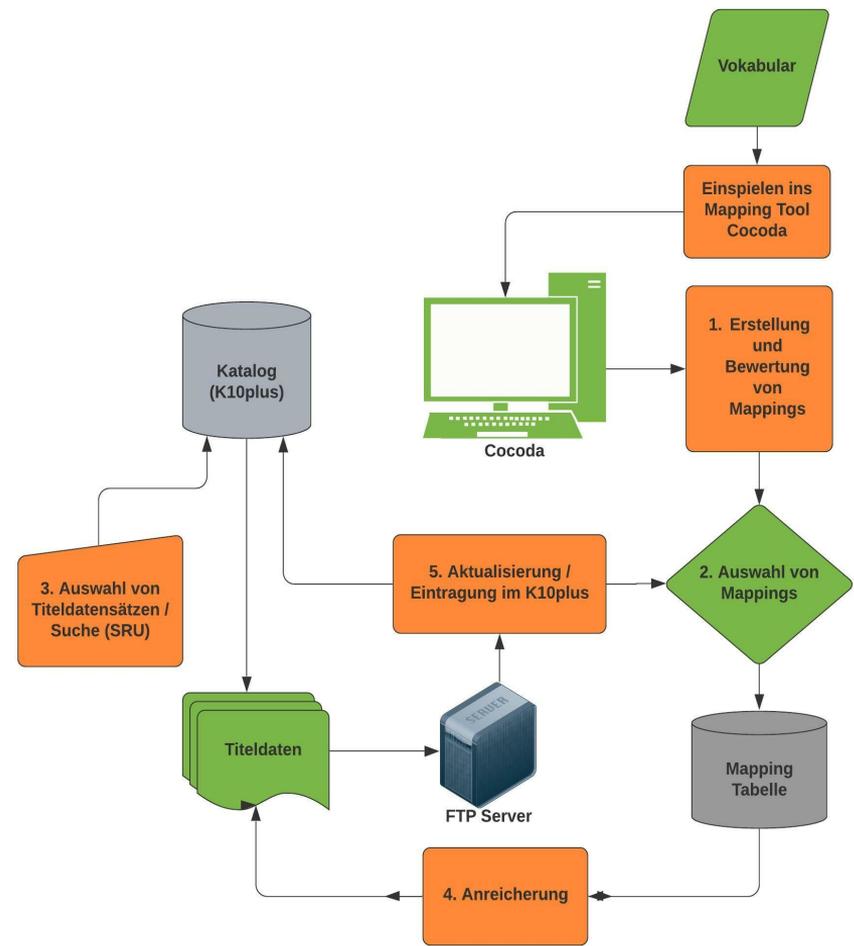
## Ergänzung vorhandener Erschließung



# coli-rich Anreicherung

- Erstellung von Mappings in Cocoda
- Überprüfung der Vollständigkeit
- Abzug von Mappings und Titeldaten
- Aktualisierung des K10plus auf Grundlage der Mappings

Über 300.000 Titeldaten wurden mit einer BK-Notation angereichert auf Grundlage von RVK-BK und NSK-BK Mappings.  
Nächste Anreicherung: 1 Mio. Titeldaten



# coli-rich Anreicherung im K10plus

<b>K10plusPPN:</b>	1756577099 
<b>Titel:</b>	<b>Diagnostic Pathology: Forensic Autopsy</b> / Mary Ann Sens, Rhome Hughes
<b>Autorin/Autor:</b>	Sens, Mary Ann [Verfasserin/Verfasser]
<b>Beteiligt:</b>	Hughes, Rhome [Mitwirkende/Mitwirkender]
<b>Erschienen:</b>	Philadelphia : Elsevier, 2021
<b>Umfang:</b>	xxi, 571, xxxv pages : [zahlreiche farbige Illustrationen] ; 28 cm
<b>Sprache(n):</b>	Englisch
<b>Anmerkung:</b>	Zugang zu "enhanced digital version" mit Code.
<b>Bibliogr. Zusammenhang:</b>	Erscheint auch als: (Online-Ausgabe)
<b>ISBN:</b>	978-0-323-75617-4
<b>Sonstige Nummern:</b>	OCOLC: 1277170607 
<b>RVK-Notation:</b>	<a href="#">XL 1600</a> <a href="#">INFO</a>   <a href="#">XL 2101</a> <a href="#">INFO</a>   <a href="#">XL 2103</a> <a href="#">INFO</a> 
<b>Sachgebiete:</b>	<b>Basisklassifikation: 44.72 (Rechtsmedizin)</b>
<b>Sonstige Schlagwörter:</b>	MeSH: <a href="#">Forensic</a> ; <a href="#">Autopsy</a>
<b>Inhaltliche Zusammenfassung:</b>	Intro -- Title Page -- Dedication -- Introduction -- 1. So you want to be a forensic pathologist -- 2. The forensic pathologist -- 3. The forensic pathologist's role in the legal system -- 4. The forensic pathologist's role in the forensic laboratory -- 5. The forensic pathologist's role in the forensic laboratory -- 6. Lightning and environmental deaths -- 7. Every contact leaves a trace -- 8. The forensic pathologist's role in the forensic laboratory -- 9. The forensic pathologist's role in the forensic laboratory -- 10. The forensic pathologist's role in the forensic laboratory -- 11. The forensic pathologist's role in the forensic laboratory -- 12. The perfect murder ... -- 13. Animals and death -- 14. Life from the dead -- 15. The forensic pathologist's role in the forensic laboratory -- 16. The forensic pathologist's role in the forensic laboratory -- 17. The forensic pathologist's role in the forensic laboratory -- 18. On death and dying -- Acknowledgements -- Glossary --

# coli-rich Anreicherung im PICA-Datensatz

```
001@ f0711,2006
001A f03806:29-04-21
001B f0BS2:21-07-22ft15:17:10.000
001D f03806:29-04-21
001U f0utf8
001X f00
002@ f0Aau
002C faTextfbtxt
002D faohne Hilfsmittel zu benutzenfbn
002E faBandfbnc
003@ f01756577099
0030 faOCoLCf01277170607
004A f0978-0-323-75617-4
007G fiKXPf01756577099
010@ faeng
010E ferda
011@ fa2021
021A faDiagnostic Pathology: Forensic AutopsyfhMary Ann Sens, Rhome Hughes
028A fdMary AnnfaSensfbVerfasserInf4aut
028C fdRhomefaHughesfbMitwirkenderf4ctb
033A fpPhiladelphiafnElsevier
034D faxxi, 571, xxxv pages
034I fa28 cm
034M fa[zahlreiche farbige Illustrationen]
037A faZugang zu "enhanced digital version" mit Code.
039D fiErscheint auch alsfnOnline-AusgabefCISBNf6978-0-323-75618-1
044C fn650fsfaForensicfxPathology
044C fn650fsfaAutopsy
045@/01 f9106409476f844.72§jRechtsmedizinfAcoli-conc RVK->BKfAhttps://coli-conc.gbv.de/api/mappings/d415aba4-14c2-4a9c-822a-1a589787545d
045R f91271582376f8XL 1600: X-Y§jMedizin / XL§jRechtsmedizin / XL 1400-XL 6099§jAllgemeine Rechtsmedizin / XL 1600-XL 2199§jAllgemeine r
1699§jAllgemeine rechtsmedizinische Diagnostik und Methodik (z.B. Tod und Leichenveränderungen) / AllgemeinesfADE-14/sred
045R f9127708291xf8XL 2101: X-Y§jMedizin / XL§jRechtsmedizin / XL 1400-XL 6099§jAllgemeine Rechtsmedizin / XL 1600-XL 2199§jAllgemeine r
2199§jTodesfeststellung, Todeszeitbestimmung / XL 2101-XL 2128§jA. Formalschlüssel / Hand- und Lehrbücher, einführende Werke
045R f91277082944f8XL 2103: X-Y§jMedizin / XL§jRechtsmedizin / XL 1400-XL 6099§jAllgemeine Rechtsmedizin / XL 1600-XL 2199§jAllgemeine r
2199§jTodesfeststellung, Todeszeitbestimmung / XL 2101-XL 2128§jA. Formalschlüssel / Kompendien, Repetitorien, Leitfäden, Kurzdarstellun
047A fa***ACHTUNG***Beim Kauf dieser Ausgabe wird ein Online-Zugang für Privatpersonen angeboten. Dieser darf nur von Einzelpersonen mit
```

# In Arbeit: coli-rich Webanwendung

**coli-rich-web**← zurück zur coli-conc Webseite    ISIL: 12345▼    EN  
DE

Semi-automatische Eintragung von Sacherschließungsdaten in PICA-Datenbanken.

Mit folgendem Formular können Titeldaten aus PICA-Datenbanken abgerufen werden, um diese mit Sacherschließungsdaten (auf Basis von coli-conc Mappings) anzureichern.

Datenbankauswahl K10plus-Verbundkatalog ▼

Titel laden

Titelangaben

PPN

Titel: Estimating Ore Grade Using Evolutionary Machine Learning Modelsinfo / Ehteram, Mohammad. - 1st ed. 2023. - Singapore : Springer Nature Singapore, 2023

Sacherschließung DDC: 549  
bicsc: RBGG ; bisacsh: SCIO48000  
LOC-SH: Mineralogy. ; Machine learning. ; Artificial intelligence--Data processing. ; Artificial intelligence.

Mögliche Anreicherung

Vokabular	Notation	Quelle	<input type="checkbox"/>
BK	38.30 Mineralogie	DDC: 549 Mineralogie = BK: 38:30 Mineralogie (Autor, Datum) <a href="#">Details</a> DDC: 549 Mineralogie -> BK: 38:30 Mineralogie (Autor, Datum) <a href="#">Details</a>	<input checked="" type="checkbox"/>

Eintragung erfolgreich! ✓

coli-rich-web ist Teil des coli-conc Projekts. Der Quellcode und die technische Dokumentation werden über [GitHub](#) bereitgestellt. Dieses Tool ist noch in der Entwicklungsphase.

coli-conc ist ein Projekt der [Verbundzentrale des GBV \(VZG\)](#). Es wird gefördert von der Deutschen Forschungsgemeinschaft (DFG) 2015-2019 und 2021-2023.

Impressum | Barrierefreiheit | Datenschutz

# coli-conc Mappings im DA-3

## Mappings im Konkordanz-Register

The screenshot shows the Mapping-Editor interface. At the top, there are two dashed boxes for dropping concepts, separated by an arrow. Below this is a toolbar with various icons. The main area is divided into three tabs: 'Konkordanzen', 'Suche', and 'Navigator'. The 'Konkordanzen' tab is active, displaying a table of mappings. The table has columns for RVK, VW, BK, and other details. Two mappings are visible:

RVK	VW	BK	Other	Date	Actions	
RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	<	BK 44.41 Pharmazeutische Biologie	David-Benja...	2022-11-15	+10 [trash] [info]
RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	~	BK 44.73 Geomedizin	Melanie Jan...	2022-04-22	±0 [trash] [info]

Below the table, there are sections for 'Lokal' and 'Wikidata-Mappings', both showing 'keine Ergebnisse </>'.

## Anzeige im DA-3 als Vorschläge

The screenshot shows the DA-3 interface with a 'Vorschläge' (Suggestions) section. It includes a header with 'Status | Rohdaten | Einstellungen #' and a sub-header with 'Filtern Aktualisieren Erweitern'. The suggestions list includes:

- BK
- 44.41 Pharmazeutische Biologie @ddc-narrow/cc, @rvk-narrow/cc [info] [copy] [add]
- 44.73 Geomedizin @rvk-related/cc [info] [copy] [add]

The source is noted as 'Quelle: - via RVK VW 6100 «Allgemeines» (Mappingtyp: verwandt) (Melanie Janßen, 2022)'. A blue link '[coli-conc]' is visible at the bottom of the suggestions list.

Link zu JSKOS Viewer,  
dort Link auf Cocoda

# Bearbeitung von Mappings in Cocoda

## Neues Mapping erstellen zum Ergänzen

RVK Mapping-Editor BK

VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen < 44.39 Toxikologie

Konkordanz-Register: nicht gespeichert V Chemie und Pharmazie (David-Benjamin Rc)

## Übernahme in den DA-3

Konkordanzen		Suche	Navigator	
<b>Mappings</b>				
Konkordanz-Register				
RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	<	BK 44.41 Pharmazeutische Biologie	David-Benja... 2022-11-15
RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	<	BK 44.39 Toxikologie	David-Benja... 2023-05-17
RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	~	BK 44.73 Geomedizin	Melanie Jan... 2022-04-22

Vorschläge Status | Rohdaten | Einstellungen #

Filtern Aktualisieren Erweitern

BK

44.39 Toxikologie @rvk-narrow/cc

Quelle: via RVK VW 6100 «Allgemeines» (Mappingtyp: eng) (David-Benjamin Rohrer, 2023)

[coli-conc]

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

44.73 Geomedizin @rvk-related/cc

# Anreicherung paralleler Ausgaben

- Vokabulare: GND, STW, BK, RVK, DDC, freie SW und sonstige Notationen
- Aktuelle Anzahl der angereicherten Titelsätze: **469.385 Titel** (Stand: 28.04.2023)
- In Onlineausgaben (Monografien) mit Verlinkung zu anderen Ausgaben (Routine)
- Neu: In anderen Medienarten, im Test wurden ca. 130.000 Titel angereichert

Lauf vom 08.03.2023	
Ergänzung mit	Anzahl Titel
DDC (5010)	4.021
RVK (5090)	7.290
BK (5301)	13.212
Notationen (5060)	364
STW (5200ff.)	3.237
freie SW (5520)	1.574
GND (5100ff., 5550ff., 5580ff.)	18.742
<b>Summe Titel brutto</b>	<b>48.440</b>
<b>Summe Titel Netto</b>	<b>23.944</b>



# Ausblick

## weitere Möglichkeiten und Pläne

# Geplant: Weitere Anreicherungen

- Optimierung bestehende Verfahren (parallele Ausgaben und Exemplare)
- Auswertung der ISBN-Nummern paralleler Ausgaben auch ohne Verlinkung
- Auswertung der Culturegraph-Werkbündel
- Regelmäßige Hebungen von lokalen Sacherschließungselementen für interessierte Bibliotheken

# Aktuell: Automatische Erschließung

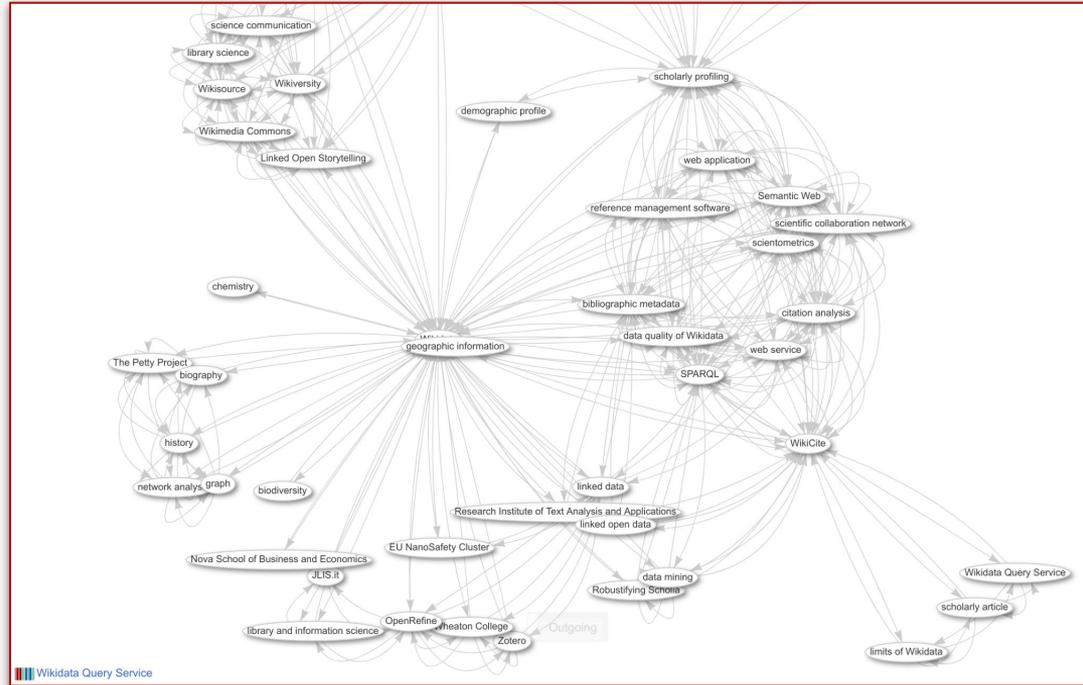
- Heuristische Algorithmen zur Ermittlung wahrscheinlicher Sacherschließung
  - Basierend auf vorhandenen Daten (Titel, Autor:innen, Volltexte...)
  - Tendenziell (aber nicht grundsätzlich) schlechter als intellektuelle Erschließung
  - Momentan spannende Entwicklung (LLM, KI...)
- 
- Umsetzung durch Bibliotheken je nach Bedarf und Ressourcen (z.B. DNB & ZBW)
  - Eintragung der Ergebnisse im K10plus
    - PICA-Unterfelder \$k und \$v: Verfahren, Konfidenzwert und Datum der Generierung
    - Weitergabe in MARC 883 Metadata Provenance

# Warum macht die VZG keine automatische Erschließung?

- Begrenzte Ressourcen
  - Erschließung ist (neben Sammeln und zugänglich Machen) eine der wesentlichen Aufgabe von Bibliotheken
  - Kontrollierte vs. probabilistische Anreicherung
  - Heterogene Metadaten erfordern Bestandsabhängige Verfahren
- 
- Ausnahme: Vorschlagsfunktion für die intellektuelle Erschließung
  - Ausnahme: KI-Verfahren im Rahmen des Retrieval (in Entwicklung)

# Forschung: K10plus Knowledge Graph

- Triple-Store aller Informationen in RDF
- Ermöglicht Analysen wie z.B. “Welche Bibliotheken haben Titel aus den 1920er Jahren zu einem Fachgebiet?”



# Zusammenfassung, Kommentare, Empfehlungen

# Die VZG bietet datenorientierte Sacherschließung

- Maschinenlesbare Vokabulare (BARTOC, URIs, JSKOS...)
- Web-Schnittstellen (Vokabulare, Mappings, [Co-]Occurrences, Subjects...)
- Web-Anwendungen (Vokabular-Browser, Cocoda, DA-3, coli-ana, QA Catalogue...)
- Datenbankabzüge bereinigter Sacherschließung (TSV, JSON & RDF)
- Statistiken & Anreicherungen der K10plus-Datenbank  
auf Grundlage vorhandener Sacherschließung
- Infrastruktur für (aber keine eigene) automatische Erschließung
- Zusammenführung der Daten auch Knowledge Graph in RDF

# Empfehlungen

- Verwendung gemeinsamer kontrollierter Vokabulare
- Vokabulare maschinenlesbar bereitstellen
- JSKOS-API zum einheitlichen Zugriff und Einbindung in Anwendungen
- Konkordanzen zum Abgleich von Vokabularen
- Analyse der vorhandenen Daten zur Datenqualität und Anreicherung

# Fragen & Diskussion!