

# **Basic Register of Thesauri, Ontologies & Classification (BARTOC)**

## **DFG-Projektantrag**

**im Bereich  
Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS)  
Förderprogramm Informationsinfrastrukturen für Forschungsdaten**

Februar 2023

(öffentliche, leicht gekürzte Version)

<https://doi.org/10.5281/zenodo.7673394>

# Inhaltsverzeichnis

<b>Beschreibung des Vorhabens</b>	<b>3</b>
1 Ausgangslage	4
1.1 Ausgangslage und eigene Vorarbeiten	4
1.2 Projektbezogene Publikationen	5
2 Ziele und Arbeitsprogramm	6
2.1 Voraussichtliche Gesamtdauer des Projekts	6
2.2 Ziele	6
2.3 Arbeitsprogramm und Umsetzung	7
AP 1 Metadaten	7
AP 2 Discovery	8
AP 3 Versionierung	8
AP 4 Redaktionsverfahren	9
AP 5 Integration	9
AP 6 Hosting	10
AP 7 Qualitätsanalyse	10
AP 8 Öffentlichkeitsarbeit	11
3 Literaturverzeichnis	12
4 Begleitinformationen zum Projektkontext	12
4.1 Allgemeine ethische Aspekte	12
4.2 Maßnahmen zur Erfüllung der Förderbedingungen und Umgang mit den Projektergebnissen	12
4.3 Erklärungen zur Erfüllung der Förderbedingungen	12
<b>Anlagen</b>	<b>13</b>
Bedarfsanalyse	13
Umfeldanalyse	14
Nachhaltigkeitskonzept	14

# Beschreibung des Vorhabens

Die Bedeutung und Anzahl von kontrollierten Vokabularen (Normdateien, Taxonomien, Ontologie...) nimmt im Zuge datengetriebener Wissenschaft zu. Damit wächst auch der Bedarf an Vokabular-Diensten und Informationen über Vokabulare für das Management von Forschungsdaten. Das vor zehn Jahren entstandene *Basic Register of Thesauri, Ontologies & Classification* (BARTOC) ist international das umfangreichste und universellste Vokabularverzeichnis. Ziel des auf drei Jahre angelegten Vorhabens ist die technische und inhaltliche Überarbeitung von BARTOC zum Ausbau zu einer fachübergreifenden Anlaufstelle für die semantische Erschließung. Konkret ist vorgesehen vor allem folgende in den letzten Jahren vernachlässigte Aufgaben umzusetzen:

- Aktualisierung und Erweiterung der Informationen zu Vokabularen in BARTOC
- Ausbau von Recherchemöglichkeiten nach und in Vokabularen durch etablierte Discovery-Verfahren
- Integration mit anderen Vokabular-Verzeichnissen und Forschungsdaten-Infrastrukturen
- Verbesserung der Qualität durch Redaktionsverfahren und Verfahren zur Qualitätsanalyse
- Unterstützung bei der Nutzung, Pflege und Verwaltung von Vokabularen

Mit dem geplanten Ausbau können sich Nutzende einfach über Vokabulare und fachspezifische Verzeichnisse informieren, passende Systeme für ihre Anwendungsfälle finden können und einheitlich auf Vokabulare zugreifen. Durch Integration mit Repositorien wird nicht nur die Sichtbarkeit von Vokabularen, sondern auch die von erschlossenen Publikationen und Forschungsdaten gesteigert. Forschungs- und Dokumentationseinrichtungen werden sowohl bei der Nutzung von Vokabularen als auch bei deren Pflege und Verwaltung unterstützt.

## Abstract

There has been a surge in the number of controlled vocabularies, such as authority files, taxonomies, ontology and their significance as a result of growth in data-driven and information science. This in turn boosts the need for vocabulary services as well as information about vocabularies for the management of research data. The *Basic Register of Thesauri, Ontologies & Classification* (BARTOC) created a decade ago is the most comprehensive and universal vocabulary index worldwide. There is however a need to expand BARTOC in the light of above developments. Taking note of this need the VZG has planned to revise, update and improve BARTOC both technically and content-wise so as to build BARTOC as a focal point for semantic indexing. The project aims to focus on the following tasks over a project period of 3 years:

- update and extend information on vocabularies in BARTOC
- further development of retrieval methods for and in vocabularies and vocabulary registries
- integration with other vocabulary repositories and research data infrastructures
- quality management by implementing review and quality analysis procedures
- support of use, maintenance and management of vocabularies

The planned improvement and further development of BARTOC will facilitate the users to discover easily the vocabularies and subject-specific directories, find suitable systems for their applications and access vocabularies uniformly. Integration with repositories not only increases the visibility of vocabularies, but also that of indexed publications and research data. Further, research and documentation institutions will be supported in the use of vocabularies as well as in their maintenance and administration.

---

# 1 Ausgangslage

## 1.1 Ausgangslage und eigene Vorarbeiten

Im Zuge datengetriebener Wissenschaft und Interoperabilität von Forschungsdaten wächst die Bedeutung semantischer Erschließung und Verknüpfung mittels kontrollierter Vokabulare, von Normdateien über Taxonomien bis zu Ontologien. Zur Übersicht und zum einheitlichen Zugriff dienen zunehmend Vokabular-Verzeichnisse und Terminology Services.<sup>1</sup> Das vor zehn Jahren entstandene *Basic Register of Thesauri, Ontologies & Classification* (BARTOC)<sup>2</sup> ist das umfangreichste und universellste Verzeichnis, da es sich nicht auf einzelne Fachgebiete, Vokabular-Typen oder Herausgeber beschränkt. Die als Open Data publizierten Inhalte von BARTOC werden von einem internationalen Redaktionsteam gepflegt<sup>3</sup> und umfassen Informationen zu rund 3.500 Vokabularen, 100 Vokabular-Verzeichnissen und zu Vokabular-Software mit Schwerpunkt auf Open Source Anwendungen.<sup>4</sup>

Der Grundfunktionen von BARTOC wurden 2020 durch Migration des Betriebs von der Universität Basel durch die VZG übernommen, wobei Ergebnisse des DFG-geförderten Projekt coli-conc bzw. Cocoda<sup>5</sup> nachgenutzt werden konnten. Die VZG hat die Datenbank in Eigenleistung ausgebaut und einige neue Funktionen umgesetzt, wie das direkte Browsing in ausgewählten Vokabularen.<sup>6</sup> Gleichwohl bleibt BARTOC, wie in der Bedarfsanalyse im Anhang dargelegt, unter seinen Möglichkeiten. Dies gilt insbesondere für Such- und Browsingmöglichkeiten, Integration mit anderen Verzeichnissen und Hilfestellung für Herausgeber von Vokabularen. Auch die assoziierten Projekte BARTOC Skosmos (Hosting von SKOS-Vokabularen)<sup>7</sup> und BARTOC FAST (Metasuche)<sup>8</sup> sind ohne zusätzliche Mittel nicht gesichert.

Vorarbeiten und Planungen existieren praktisch in allen geplanten Bereichen des im folgenden Kapitel dargelegten Vorhabens.<sup>9</sup> Die VZG verfügt über ausgezeichnete Kenntnisse im Bereich Metadaten und Discovery (AP 1 und 2) und den Betrieb von digitaler Infrastruktur für die Wissenschaft (AP 7). Zur weiteren Integration mit Forschungsdaten-Systemen (AP 5) gibt es insbesondere Vorarbeiten im Bereich Schnittstellen wie JSKOS-API,<sup>10</sup> BEACON,<sup>11</sup> SeeAlso (Voß 2008) und Occurrence-API,<sup>12</sup> durch die bereits einige Repositories und Kataloge an BARTOC angebunden sind. Ansätze zur Qualitätsanalyse von Vokabularen (AP 7) gibt es sowohl innerhalb<sup>13</sup> als auch außerhalb der VZG.<sup>14</sup>

---

<sup>1</sup> Siehe <https://bartoc.org/registries>

<sup>2</sup> Gegründet als *Basel Register of Thesauri, Ontologies & Classification*. Erreichbar unter <https://bartoc.org>

<sup>3</sup> Siehe <https://bartoc.org/contact>.

<sup>4</sup> Siehe <https://bartoc.org/software>.

<sup>5</sup> Siehe <https://coli-conc.gbv.de/> (GEPRIS-Projektnummer 276843344 und 455051200).

<sup>6</sup> Siehe <https://coli-conc.gbv.de/blog/2021/03/23/bartoc-vocabulary-browsing/>.

<sup>7</sup> Bisher <https://bartoc-skosmos.unibas.ch/>, geplant <https://skosmos.bartoc.org/>.

<sup>8</sup> Siehe Hindermann & Ledl (2020).

<sup>9</sup> Siehe Voß & Balakrishnan (2021), Voß, Balakrishnan, Peters (2021), <https://github.com/gbv/bartoc.org/issues> u.A.

<sup>10</sup> Neben der Referenzimplementierung unter <https://github.com/gbv/jskos-server> auch an anderer Stelle umgesetzt.

<sup>11</sup> Siehe <https://de.wikipedia.org/wiki/Wikipedia:BEACON>.

<sup>12</sup> Siehe <https://github.com/gbv/occurrences-api>.

<sup>13</sup> Siehe <https://format.gbv.de/validate> und <https://github.com/gbv/jskos-metrics>.

<sup>14</sup> Siehe Stock (2015) und die unter AP 7 genannten Systeme.

---

## 1.2 Projektbezogene Publikationen

### a) Arbeiten in Publikationsorganen mit einer wissenschaftlichen Qualitätssicherung

- Voß, J. et. al. (2017): *Erfassung von Wissensorganisationssystemen in BARTOC - Ergebnis eines Projektseminars an der Hochschule Hannover*. In: Informationspraxis, Bd. 3, Nr. 2. <https://doi.org/10.11588/ip.2017.2.40335>
- Voß, J., Ledl, A., & Balakrishnan, U. (2016): *Uniform description and access to Knowledge Organization Systems with BARTOC and JSKOS*. Terminology & Ontology: Theories and applications (TOTh), Chambéry. <https://doi.org/10.5281/zenodo.438019>
- Voß, J. et al. (2016): *Terminology Registries and Services*. Presented at Research Data Alliance Deutschland Treffen and SWIB 2016. <https://doi.org/10.5281/zenodo.166717>
- Ledl, A. & Voß, J. (2016): *Describing Knowledge Organization Systems in BARTOC and JSKOS*. In: Erdman Thomsen, Hanne; Pareja-Lora, Antonio; Madsen, Bodil Nistrup: *Term Bases And Linguistic Linked Open Data*. Frederiksberg: Copenhagen Business School, S. 168-178. <https://research.cbs.dk/en/publications/term-bases-and-linguistic-linked-open-data-proceedings-of-tke-201>
- Voß, J. (2008): *Get Tooled Up: SeeAlso: A Simple Linkserver Protocol*. In: Ariadne 57. <http://www.ariadne.ac.uk/issue/57/voss/>

### b) Andere Veröffentlichungen

- Anderes et. al. (2022): *Terminology Services - Working Group Charter (NFDI Section Metadata)*. <https://doi.org/10.5281/zenodo.6759324>
- Voß, J. et. al. (2021): *Unified Access to controlled vocabularies with BARTOC*. Mini-ELAG 2021. <https://coli-conc.gbv.de/publications/2021-06-30-mini-elag-bartoc.pdf>
- Voß, J. (2021): *Vom Vokabularverzeichnis zum interdisziplinären KOS-Datendienst: 8 Jahre BARTOC*. 109. Bibliothekartag. <https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/17563>
- Voß, J. & Balakrishnan, U. (2021): *BARTOC: Looking both ways, then and now as it prepares for the future of Knowledge Organisation*. ISKO UK Research Observatory. <https://doi.org/10.5281/zenodo.4766970>
- Voß, J., Balakrishnan, U., Peters, S. (2021): *BARTOC & coli-conc: The story of a partnership and future plans*. EDUG 2021. <https://coli-conc.gbv.de/publications/2021-05-06-bartoc-coli-conc.pdf>

---

## 2 Ziele und Arbeitsprogramm

### 2.1 Voraussichtliche Gesamtdauer des Projekts

36 Monate.

### 2.2 Ziele

Das Gesamtziel des Projektes ist die Weiterentwicklung des Vokabular-Verzeichnis BARTOC als Informationsinfrastruktur für das Management von Forschungsdaten (FDM). Dabei soll der etablierte Betrieb als verlässlicher Dienst durch Ausbau der Funktionalität für weitere Fachcommunities ergänzt werden, so dass BARTOC Disziplin-übergreifend als Anlaufstelle für semantische Erschließung dient. Neben der Evaluation und Entwicklung zusätzlichen Funktionen soll auch Schulungs- und Trainingsmaterial zur Pflege und Nutzung von Vokabularen für das FDM bereitgestellt werden. Der Nutzen dieser Optimierung dient sowohl einzelnen Forscher\*innen bei der Auswahl und Anwendung von Vokabularen als auch übergeordneten Einrichtungen wie Forschungsdatenrepositories, Fachinformationsdiensten (FID) und Konsortien der Nationalen Forschungsdateninfrastruktur (NFDI) durch Integration von Vokabular-Diensten.

#### **Ziel 1: Ausbau von Inhalt und Recherchemöglichkeiten**

Dieses Ziel beinhaltet insbesondere:

- Aktualisierung und Erweiterung der Datenbasis von BARTOC zur Sicherstellung einer einheitlichen, aktuellen und vollständigen Erschließung der verzeichneten Vokabulare
- Gruppierung von Vokabularen nach fachbezogenen Anwendungsszenarien
- Evaluierung und Implementierung von Discovery-Verfahren (Ranking, Facettierung, Metasuche...)
- Verbesserung der Usability und Ausbau von Funktionen zur Visualisierung von Vokabularen

Maßnahmen zur Erreichung dieses Ziels sind vor allem in AP 1 und AP 2 vorgesehen.

#### **Ziel 2: Integration mit anderen Systeme zur Nutzung, Pflege und Verwaltung von Vokabularen**

Dieses Ziel beinhaltet insbesondere:

- Ausbau der Integration mit Vokabular-Verzeichnissen um weitere Softwaresysteme
- Automatischer Abgleich mit vorhandenen Verzeichnissen
- Direkter Zugriff auf Forschungsdaten und anderer mit Vokabularen erschlossener Publikationen
- Einbindung von Mappings zwischen Vokabularen zur Verbesserung des Retrievals

Maßnahmen zur Erreichung dieses Ziels sind vor allem in AP 5 vorgesehen.

#### **Ziel 3: Aufbau von zurückgestellten Funktionalitäten für das Management von Forschungsdaten**

Dieses Ziel beinhaltet insbesondere:

- Aktualisierung der von der Universität Basel übernommenen technischen Infrastruktur
- Einführung von Informationen zu Ausgaben und Versionen von Vokabularen
- Verbesserung der Möglichkeit und Usability zur Bearbeitung der Inhalte von BARTOC
- Einführung eines kontrollierten Redaktionsverfahren
- Bereitstellung von Angaben zur Nutzung und Qualität von Vokabulare

Maßnahmen zur Erreichung dieses Ziels sind vor allem in AP 3, AP 4 und AP 7 vorgesehen.

---

## **Ziel 4: Unterstützung bei Nutzung, Pflege und Verwaltung von Vokabularen**

Dieses Ziel beinhaltet insbesondere:

- Sammlung und Erstellung von Schulungs- und Trainingsmaterial
- Hinweise auf fachbezogene Hosting-Möglichkeiten und Hosting ausgewählter Vokabulare
- Empfehlungen zu Vokabular-Software und -Services
- Öffentlichkeitsarbeit mit Schwerpunkt auf Forschungsdatenmanagement

Maßnahmen zur Erreichung dieses Ziels sind vor allem in AP 6 und AP 8 vorgesehen.

## **2.3 Arbeitsprogramm und Umsetzung**

### **AP 1 Metadaten**

In den letzten 10 Jahren sind in BARTOC Angaben zu rund 3.500 Vokabularen erfasst worden. Während laufend neue Vokabulare hinzukommen, haben sich Umfang und Tiefe der Erschließung teilweise uneinheitlich entwickelt. Zur Verbesserung der Datenqualität soll deshalb der gesamte Datenbestand einer Revision unterzogen und vereinheitlicht werden. Die Harmonisierung umfasst:

#### **Auswertung und Vervollständigung**

- Aktualisierung und Erweiterung veralteter Metadaten
- Überprüfung von und Abgleich mit Inhalten von Registries
- Statistische Auswertung fehlender Metadaten
- Recherche und Ergänzung fehlender Metadaten

Die Vervollständigung von Metadaten beinhaltet insbesondere einheitliche Angaben zur fachlichen Zuordnung (derzeit bei 85% aller Vokabulare vorhanden) und Lizenzinformationen (derzeit 26%). Zudem sollen Metadaten automatisch mit Angaben aus anderen Registries abgeglichen werden.

#### **Erweiterung um bisher nicht erfasste Metadaten**

- Angaben zur Nutzung von Vokabularen (Institutionen, Art der Nutzung, Umfang, Anwendungen)
- Literaturangaben (Handbücher, Leitfäden etc.)
- Gruppierung zusammengehöriger Vokabulare als Merklisten, Sucheinstieg und zur Kennzeichnung von Vokabularen, die für die Forschung besonders relevant sind
- Explizite Kennzeichnung von FAIR-Vocabularies (Cox et al., 2021)

Das Arbeitspaket umfasst Recherche und Katalogisierung sowie die Erweiterung des Datenmodells und der Erfassungsmasken.

---

## **AP 2 Discovery**

Die Suchfunktion von BARTOC ist derzeit sehr rudimentär, da bislang keine dezidierte Discovery-Oberfläche umgesetzt wurde. Auch die ursprünglich vorhandenen Möglichkeiten zur mehrsprachigen Suche konnten seit Übernahme der Datenbank durch die VZG nicht wieder aktiviert werden. Die Recherche-Funktion soll deshalb neu konzipiert und umgesetzt werden. Dies beinhaltet:

### **Evaluierung von Recherchemöglichkeiten durch mehrere Prototypen**

Grundsätzlich kommen mindestens vier Verfahren in Frage:

- Einfache Datenbank (bisherige Umsetzung)
- Zentraler Suchindex (Klassische Suchmaschinentechologie mit Ranking und Filtern)
- Wissensgraph (Graph-Datenbank)
- Metasuche über bekannte KOS-Registries (BARTOC FAST)

Die Recherche soll durch Umsetzung folgender Funktionen weiterentwickelt werden. Die Eignung der verschiedenen Verfahren wird mit Hilfe von Prototypen evaluiert:

- Ranking von Suchergebnissen
- Browsen und Filtern von Vokabularen
- Gruppieren von Vokabularen
- Visualisierung von Beziehungen zwischen Vokabularen
- Einbeziehung von Vokabular-Inhalten
- Mehrsprachige Suche

### **Umsetzung der neuen Recherche-Oberfläche**

Nach Abschluss der Evaluierung soll zur Verbesserung der Recherche begleitet durch Usability-Studien und in Rücksprache mit dem Redaktionsteam eine neue Suchoberfläche umgesetzt werden. Insbesondere sollen das Auffinden von thematisch passenden Vokabularen, Beziehungen zwischen Vokabularen und Informationen zur Vokabular-Qualität (AP 7) leichter ersichtlich werden.

## **AP 3 Versionierung**

Bisher werden Vokabulare in BARTOC nur als Werke erfasst (z.B. ein Eintrag für DDC in allen Sprachen und Ausgaben). Dies ermöglicht zwar das Auffinden von passenden Vokabularen, für ihre direkte Verwendung sind jedoch zusätzlich Informationen zu konkreten Ausgaben und Versionen notwendig. Insbesondere Angaben zu Umfang und Zugriffsmöglichkeiten sollten daher versioniert und mit Datum versehen werden. Das Arbeitspaket umfasst zwei Punkte:

### **Erweiterung des Datenmodells um Ausgaben und Versionen**

Das bislang flache Datenmodell muss um mindestens einen Entitätstyp erweitert werden. Zu untersuchen ist, welche Angaben auf welche Beschreibungsebene gehören. Dabei sollte zwischen Genauigkeit und Praktikabilität abgewogen werden. Als Grundlage für die Modellierung dient das IFLA Library Reference Model (LRM, siehe Žumer 2018).

### **Versionierung der Metadaten**

Die Datenbank soll so erweitert werden, dass wie bei einem Wiki alle Änderungen dauerhaft aufgezeichnet



---

werden. Diese Erweiterung dient der Verbesserung der Transparenz und ist Grundlage für eine Ausweitung des Redaktionsverfahrens (AP 4) und Qualitätsanalyse (AP 7).

## **AP 4 Redaktionsverfahren**

Bislang ist der schreibende Zugriff auf BARTOC so geregelt, dass alle Mitglieder des internationalen Redaktionsteam<sup>15</sup> alle Datensätze verändern können. Mit wachsender Zahl von Beteiligten ist jedoch ein genaueres Redaktionsverfahren notwendig. Neben Rechtemanagement soll das Redaktionsverfahren auch Funktionen zur Qualitätssicherung integrieren (siehe AP 7). Geplant sind Funktionen für:

- Benachrichtigung bei Änderungen an Datensätzen
- Admin-Accounts für Löschungen und Bearbeitung sensibler Felder
- Erweiterte Validierung während der Eingabe
- Vorschlagsfunktion für neue Vokabulare und Änderungen durch Dritte
- Änderung und Bestätigung von Änderungen durch Herausgeb\*innen von Vokabularen
- Prüfung/Bestätigung von automatischen Import-Routinen (siehe AP 5)
- Erweiterung der Dokumentation zur Bearbeitung von Inhalten (siehe AP 1)

## **AP 5 Integration**

Mit Übernahme von BARTOC konnten wesentlichen Funktionen in Normdaten-Anwendungen der VZG wie DANTE<sup>16</sup> und Cocoda integriert werden. Die Anbindung an Kataloge, Repositorien und Konkordanzen ist jedoch auf ausgewählte Systeme beschränkt. Durch Ausbau von Standards und Schnittstellen soll die Integration insbesondere auf Forschungsdaten-Repositorien ausgeweitet werden. Das Arbeitspaket umfasst:

### **Integration mit KOS-Registries**

- Umsetzung weiterer Wrapper zum direkten Zugriff auf Vokabulare, die in anderen Registries bereitgestellt werden. Bisher: Skosmos, SkoHub und zum Teil Wikibase, geplant: OntoPortal, Ontology Look-up Service (OLS), Wikibase u.A.
- Methoden zum (semi-)automatischen Abgleich mit Registries (siehe auch AP 1 und AP 4)
- Ermöglichung der Metasuche über verschiedene Registries (Ersatz für BARTOC FAST)
- Einbindung von Mappings zwischen Vokabularen

### **Integration mit Repositories**

- Deep-Links aus Vokabularen in BARTOC in digitale Sammlungen, Kataloge und Repositories
- Abfrage und Einbindung von Informationen zur Nutzungshäufigkeit von Vokabularen in Datenbanken und Repositories (u.A. via APIs und BEACON-Dateien)
- Suche und Zugriff auf Forschungsdaten und andere Publikationen aus Katalogen und Repositories mit Hilfe der Vokabulare, die in BARTOC erfasst sind.

---

<sup>15</sup> Siehe <https://bartoc.org/contact>.

<sup>16</sup> Siehe <https://api.dante.gbv.de/>.

---

## AP 6 Hosting

Grundsätzlich erfasst BARTOC nur Metadaten über KOS und deren Verfügbarkeit, während ihre Inhalte an anderer Stelle bereitgestellt werden müssen. Die Bandbreite der Bereitstellungsformen ist allerdings sehr breit: selbst wenn Vokabulare verfügbar sind, sind sie nicht unbedingt maschinenlesbar und im Rahmen der FAIR-Prinzipien nutzbar. BARTOC soll Produzenten von Vokabularen mit Anleitungen und Best-Practices helfen (siehe AP 1 und AP 8) und auf vorhandene Dienste zur Bereitstellung von Vokabularen verweisen. Nicht immer lässt sich jedoch zeitnah ein passender Dienst finden. Auch kommt es vor, dass externe Vokabular-Dienste eingestellt und kurzfristig neue Hosting-Möglichkeiten benötigt werden. Dies gilt auch für die rund 1.000 Vokabularen des von der Universität Basel gehosteten Dienst BARTOC-Skosmos,<sup>17</sup> dessen Betrieb und Pflege nicht gesichert ist. Da also Bedarf besteht und bereits jetzt in Ausnahmefällen Vokabulare direkt über BARTOC bereitgestellt werden, soll eine skalierbare Möglichkeit zum Hosting von Vokabularen geschaffen werden. Diese Maßnahmen dazu bestehen aus:

- Übersicht und Vergleich vorhandener Hosting-Möglichkeiten (siehe dazu auch AP 1)
- Evaluierung von Software zum Betrieb von Terminology-Services (Skosmos, OntoPortal etc.)
- Untersuchung von Möglichkeiten und Grenzen zur föderierten Bereitstellung von Vokabulardaten in Zusammenarbeit mit bestehenden Terminology Services
- Untersuchung der Möglichkeiten und Grenzen zur Archivierung von KOS-Daten in vorhandenen Forschungsdaten-Repositories
- Übernahme der wesentlichen Inhalte von BARTOC-Skosmos
- Erarbeitung eines kostenpflichtigen Dienstes zur Konvertierung und maschinellen Bereitstellung von Vokabularen

Eine vorläufige Analyse ergab dass bestehende Dienste<sup>18</sup> entweder für die detaillierte Verwaltung von Vokabularen mit Benutzermanagement konzipiert sind (und daher nicht ohne Betreuungsaufwand skalierbar) oder sich auf Vokabulare beschränken, die bereits sauber in den RDF-Formaten SKOS oder OWL vorliegen. Ein Hosting-Dienst von BARTOC könnte gut den Bedarf dazwischen decken.

## AP 7 Qualitätsanalyse

Angaben zur Qualität von Vokabularen sind sowohl für Nutzende als auch für ihre Herausgeber hilfreich. Die Qualität ist allerdings sehr uneinheitlich und lässt sich nur schwer durch oberflächliche Betrachtung beurteilen. Deshalb sollen möglichst automatisch ermittelbare Angaben zur Qualität erfasst und bereitgestellt werden.

### Sammlung und Evaluation von Verfahren zur Qualitätsanalyse

Vorhandene Verfahren und Anwendungen zur Analyse von Vokabularen (Skosify, qSKOS, jskos-metrics...) sollen ermittelt und verglichen werden. Die Qualitätskriterien umfassen unter anderem:

- Einhaltung von Syntax und Dateiformaten
- Konsistenzregeln der verwendeten Modellierungssprachen (SKOS, OWL, OBO...)
- Vokabular-Statistiken zu Umfang, Komplexität etc.
- Vergleich mit angegebenen Vokabulartypen (z.B. Hierarchien in Klassifikationen im Gegensatz zu Thesauri und Normdateien)

---

<sup>17</sup> Siehe <https://bartoc-skosmos.unibas.ch/>.

<sup>18</sup> Siehe <https://bartoc.org/registries>.

- 
- Linguistische Verfahren zur Analyse von Benennungen
  - Einbeziehung der Nutzung von Vokabularen zur Aufdecken von Unausgewogenheiten
  - Nutzbarkeit der Vokabulardaten (FAIR-Data)

### **Entwicklung eines Dienstes zur Qualitätsanalyse von Vokabulardaten**

Die ermittelten Verfahren zur Validierung, Konsistenzprüfung und Bewertung sollen über einen Webdienst bereitgestellt werden. Die Anwendung sowohl BARTOC-Redakteur\*innen zur Beurteilung von Vokabularen als auch Autor\*innen von Vokabularen zur Überprüfung und Verbesserung helfen. Die Bewertung soll nicht nur technische, sondern auch inhaltliche Aspekte umfassen, beispielsweise bei der Identifizierung von veralteten und diskriminierenden Begriffen (Haugen & Billey 2020).

## **AP 8 Öffentlichkeitsarbeit**

In Fachkreisen ist BARTOC bekannt und geschätzt, die effektive Nutzung setzt allerdings ein Grundverständnis von Vokabularen voraus, das nicht von allgemeinen Wissenschaftler\*innen vorausgesetzt werden kann. Nach Ausbau der Discovery-Möglichkeiten (AP 2) und weiterer Funktionen soll daher gezielt die Dokumentation ausgebaut und BARTOC bekannt gemacht werden. Dazu sollen Forschungseinrichtungen und Initiativen (NDFI, RDA...) und Herausgeber\*innen angesprochen und eingebunden werden. Konkrete Ziele und Maßnahmen dieses AP sind:

- Erweiterung des Redaktionsteam
- Dokumentation von Standards und Best-Practice für gute Vokabulare (wie FAIR-Prinzipien und die in AP 7 ermittelten Qualitätskriterien)
- Beratung und Lehrmaterial für Herausgeber\*innen von Vokabularen
- Sammlung und Erstellung von Schulungs- und Trainingsmaßnahmen zum Einsatz von Vokabularen insbesondere zur Erschließung von Forschungsdaten
- Hilfe zur Anwendung und Integration von KOS in andere Systeme, um die in BARTOC identifizierten Vokabulare in eigenen Anwendungen zu nutzen

---

## 3 Literaturverzeichnis

Siehe auch die projektbezogenen Publikationen oben in Kapitel 1.2.

- Cox et al. (2021): *Ten simple rules for making a vocabulary FAIR*. In: PLoS Comput Biol 17(6). <https://doi.org/10.1371/journal.pcbi.1009041>
- Haugen M. & Billey, A. (2020): *Building a More Diverse and Inclusive Cataloging Cooperative*. In: *Cataloging & Classification Quarterly* 58(3-4). <https://doi.org/10.1080/01639374.2020.1717709>
- Hindermann, M. & Ledl, A. (2020): *BARTOC FAST: A Federated Asynchronous Search Tool for Remote Vocabulary Access*. In: *Knowledge Organization at the Interface: Proceedings of the Sixteenth International ISKO Conference*, Seite 200-206. <https://doi.org/10.5771/9783956507762-200>
- Stock, W.: (2015): *Informetric Analyses of Knowledge Organization Systems (KOSs)*. In: C. R. Sugimoto (Ed.): *Theories of Informetrics and Scholarly Communication*. De Gruyter. <https://arxiv.org/abs/1505.03671>
- Žumer, Maja. 2018: *IFLA Library Reference Model (IFLA LRM): Harmonisation of the FRBR Family*. In Hjørland, Birger and Gnoli, Claudio eds., *ISKO Encyclopedia of Knowledge Organization*, <http://http://www.isko.org/cyclo/lrm>

## 4 Begleitinformationen zum Projektkontext

### 4.1 Allgemeine ethische Aspekte

Ethisch relevante negative Auswirkungen sind nicht zu erwarten. Im Gegenteil ist davon auszugehen, dass die Verbesserung von Informationen von und zu Vokabularen dazu beitragen wird, diskriminierende Bestandteile und Katalogisierungspraktiken aufzudecken (vgl. Haugen & Billey 2020 und AP 7).

### 4.2 Maßnahmen zur Erfüllung der Förderbedingungen und Umgang mit den Projektergebnissen

BARTOC wird bereits jetzt dauerhaft als Basisdienst von der VZG betrieben. Die Erweiterung von BARTOC als Informationsinfrastruktur stützt sich im wesentlichen auf Techniken und Systeme, die auch an anderer Stelle in der VZG relevant sind, so dass ein nachhaltiger Betrieb durch die VZG sichergestellt ist (siehe Nachhaltigkeitskonzept im Anhang). Die Inhalte der Datenbank von BARTOC werden laufend als Open Data und die Software zum Betrieb als Open Source bereitgestellt.

### 4.3 Erklärungen zur Erfüllung der Förderbedingungen

Hiermit bestätigen wir, dass die aus dem Projekt resultierenden Publikationen ebenso wie einschlägige Dokumentationen im Open Access verfügbar gemacht und Dritten zur umfassenden Nachnutzung bereitstehen werden und dass der Quellcode der im Projekt entwickelten Software nach etablierten Standards dokumentiert, mit einer Open Source Lizenz versehen und für die Nachnutzung durch Dritte kostenlos verfügbar gemacht wird.

---

# Anlagen

## Bedarfsanalyse

Der Bedarf an maschinenlesbaren kontrollierten Vokabularen für die Wissenschaft hat in den letzten Jahren beständig zugenommen. Grund dafür ist ihre zentrale Rolle für die Zusammenführung und Nachnutzung von Daten, von einfachen Listen bis zu komplexen Wissensgraphen. Aktuelle Beispiele für übergreifende Initiativen zu Vokabularen in der Wissenschaft sind die Vocabulary Services Interest Group der Research Data Alliance (RDA VSSIG),<sup>19</sup> die CODATA Working Group on FAIR Vocabularies<sup>20</sup> und die NFDI Sektion Sektion (Meta)daten, Terminologien, Provenienz.<sup>21</sup> Erst kürzlich hat letztere den Bedarf an Standards und Übersichten von Vokabular-Diensten noch einmal festgestellt (Anders et al. 2022). Mit wachsender Zahl von Vokabularen gibt es zum Überblick und Zugriff auf Ontologien und Terminologien immer mehr Vokabular-Dienste und -Verzeichnisse – BARTOC verzeichnet inzwischen über 100. Ausgerichtet an den spezifischen Bedürfnissen einzelner Fachcommunities erfüllen die einzelnen Verzeichnisse eine wichtige Rolle, können aber hinsichtlich Umfang und Nachhaltigkeit nicht die Bedarfe aller Disziplinen decken.

Dank der breit abgedeckten Fachbereiche wird BARTOC als übergreifender Einstiegspunkt bei Produzenten und Nutzenden von Vokabularen sehr geschätzt (siehe dazu auch die Letter of Support im Anhang). Wie die Analyse von Zugriffszahlen, Gespräche mit Forschungseinrichtungen, dem internationalen Redaktionsteam von BARTOC und eine für die FAG Erschließung und Informationsvermittlung (FAG-EI)<sup>22</sup> durchgeführte Umfrage zeigen, nimmt die Nutzung & das Interesse ständig zu. Gleichzeitig werden Schwächen und Lücken deutlich, da mit der Übernahme der Datenbank von der Universität Basel nur die Basisfunktionen abgedeckt werden konnten. Insbesondere werden die eingeschränkte Such- und Browsingmöglichkeit sowie Möglichkeiten zum Hosting von Vokabularen genannt.

BARTOC erhält regelmäßig Anfragen, wie Vokabulare in maschinenlesbare Form überführt und verfügbar gemacht werden können. Die technischen Möglichkeiten der Produzenten von Vokabularen sind sehr unterschiedlich und es ist fast immer intellektuelle Datenanalyse und -bereinigung notwendig. Dies zeigt, dass BARTOC bereits jetzt eine Anlaufstelle für die Pflege von Vokabularen ist und dass Dokumentation und Beratung ausgebaut werden muss. Grundsätzlich sollen solche Anfragen an passende fachspezifische Vokabular-Dienste weitergeleitet werden. Dies ist jedoch nicht immer möglich und es kommt auch vor, dass bestehende Dienste eingestellt und kurzfristig neue Hosting-Möglichkeiten gefunden werden müssen. Dies betrifft auch den bisher nicht von der VZG übernommenen Dienst BARTOC-Skosmos (mit mehr als 1.000 SKOS-Vokabularen), dessen Weiterbetrieb durch die Universität Basel nicht gesichert ist. Aus diesen Gründen ist es notwendig, Möglichkeiten zu schaffen, Vokabulare auch direkt bei BARTOC zu hosten. Klassische Forschungsdaten-Repositories reichen dafür nicht aus, da nicht nur statische Dateien, sondern auch Schnittstellen und Benutzeroberflächen für Vokabulare angeboten werden müssen.

Zur allgemeinen Verbesserung von BARTOC werden Vorschläge von Nutzer\*innen, Kolleg\*innen und dem Redaktionsteam in einem öffentlichen Issue-Tracker gesammelt.<sup>23</sup> Da Entwicklung bislang ausschließlich mit Eigenmitteln erfolgte, haben sich zahlreiche Funktionen angesammelt, die bisher nicht implementiert werden konnten. Zur Priorisierung wird derzeit eine erste Evaluation der wichtigsten Funktionen vorhandenen Vokabularverzeichnisse durchgeführt.

---

<sup>19</sup> <https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html>

<sup>20</sup> <https://codata.org/initiatives/decadal-programme2/fair-vocabularies/iussp-codata-working-group-on-fair-vocabularies/>

<sup>21</sup> <https://www.nfdi.de/section-meta/>

<sup>22</sup> <https://verbundwiki.gbv.de/display/FAGEI>

<sup>23</sup> <https://github.com/gbv/bartoc.org/issues>

---

## Umfeldanalyse

Als gemeinsames Datenmodell für maschinenlesbare Vokabulare hat sich grundsätzlich RDF durchgesetzt. Dabei kann unterschieden werden zwischen formalen Ontologien in OWL und anderen RDF-Vokabularen auf der einen Seite und einfacheren Terminologien in SKOS auf der anderen Seite. Der Unterschied zwischen beiden Arten von Vokabularen korreliert mit der Unterscheidung zwischen TBox und ABox in Wissensgraphen. Eine strikte Unterscheidung ist aber je nach Anwendungsgebiet und Community nicht immer möglich. Das im Rahmen des Projekt *coli-conc* entwickelte Datenformat JSKOS fasst Eigenschaften aller Arten von Vokabularen zusammen und berücksichtigt zusätzlich erweiterte Metadaten und Mappings. Durch Konvertierung von und nach RDF ist die Interoperabilität mit anderen Vokabular-Formaten sichergestellt. Somit bietet es eine gute Grundlage für die erweiterte Beschreibung von Vokabularen.

Die Formen und Orte der Bereitstellung von Vokabularen sind aufgrund der hohen Zahl von Organisationen und Initiativen relativ heterogen. BARTOC enthält derzeit Informationen zu 110 Vokabular-Verzeichnissen (Registries). Eine erste systematische Auswertung fand 2016 statt. Derzeit stellen 69 Registries neben Informationen über Vokabulare auch die Daten der Vokabulare selbst bereit, in mindestens 38 Fällen über eine API. Erfasst wird auch die Software mit der Vokabulare bereitgestellt werden. Verbreitete Systeme sind *OntoPortal* und *Ontology Lookup Service* für für Ontologien sowie *Skosmos*, *Skohub* und *JSKOS Server* für Vokabulare auf Basis von SKOS. Die Registries sind teilweise in Föderationen organisiert (*OntoPortal Alliance* und *Ontology Lookup Service Community*). Die VZG pflegt im Rahmen von BARTOC eine aktuelle Übersicht zu Vokabular-Software mit rund zwei Dutzend aktiven Open Source Systemen, als Grundlage für eine genauere Evaluation. Ein Teil dieser Software (*Skosify*, *qSKOS/SKOS Testing Tool* u.A.) lässt sich auch für die geplante Qualitätsanalyse einsetzen.

Eine systematische Übersicht von Informationsmaterial zur Erstellung und Pflege von Vokabularen, insbesondere zur Erschließung von Forschungsdaten, liegt noch nicht vor. In der Regel beschränkt sich die Literatur auf informationstheoretische Aspekte, auf bestimmte Arten von Vokabularen (Ontologien, Taxonomien, Normdateien...) oder auf ausgewählte technische Aspekte.

## Nachhaltigkeitskonzept

BARTOC wird im Rahmen seiner Basisdienste seit zwei knapp zwei Jahren von der VZG in Eigenleistung betrieben und gepflegt. Die Datenbank basiert inzwischen auf Techniken, die auch an anderer Stelle innerhalb der Einrichtung eingesetzt werden und ist in zunehmendem Maße mit lokalen Diensten wie DANTE und Cocoda und mit der regulären Sacherschließungsarbeit integriert. Der dauerhafte Betrieb von BARTOC ist daher bereits jetzt sichergestellt. Die im Rahmen dieses Projekt zu entwickelnden Erweiterungen werden in die vorhandenen Inhalte und Software von BARTOC einfließen, so dass der zusätzliche Aufwand für den technischen Betrieb nach Ablauf der Projektphase überschaubar ist. Für einzelne Funktionen wie Hosting und Beratung, deren Aufwand mit Anzahl der zu betreuenden Anwender\*innen steigt, ist ein kostenpflichtiger Zusatzdienst vorgesehen. Für den dauerhaften Betrieb der Infrastruktur sind dessen Einnahmen jedoch nicht notwendig. Die Anbindung der Ergebnisse an die Forschungspraxis ist durch die Tätigkeit der VZG als Dienstleister für Wissenschaftliche Bibliotheken und die Mitgliedschaft im NDFI e.V. sichergestellt. Die Zusammenarbeit mit weiteren Fachcommunities soll im Rahmen des Projekts ausgebaut werden, sobald die von wissenschaftlicher Seite geforderten Funktionen wie Retrieval, Integration, Versionierung und Redaktionsprozesse umgesetzt sind.