



e·infra
austria

AP 5 Metadaten

Sacherschließung von Forschungsdaten

Patrik Kennel

Universitäts- und Landesbibliothek Tirol

September 2019



Der Inhalt dieser Veröffentlichung steht unter einer Creative Commons Namensnennung 4.0 Lizenz (<http://creativecommons.org/licenses/by/4.0/>)

[DOI 10.25651/1.2019.0017](https://doi.org/10.25651/1.2019.0017)

Über e-Infrastructures Austria Plus

Das Projekt „e-Infrastructures Austria Plus“ (2017-2019) ist ein vom österreichischen Bundesministerium für Bildung, Wissenschaft und Forschung (<https://bildung.bmbwf.gv.at/>) gefördertes Projekt von neun österreichischen Universitäten. Ziel des Projekts ist der koordinierte Aufbau eines österreichischen Netzwerks zur Einrichtung und Weiterentwicklung gemeinsamer e-Infrastrukturen durch Bündelung von Ressourcen und vorhandenem Wissen.

Inhalt

Hintergrund	3
Einleitung	3
Sacherschließung in institutionellen Repositorien.....	4
Universität Heidelberg	4
Humboldt Universität zu Berlin.....	4
Eidgenössische Technische Hochschule Zürich (ETHZ).....	5
Universität Tübingen	5
Oxford University.....	5
University of Cambridge	6
Delft University of Technology /Eindhoven University of Technology/University of Twente/ Wageningen University and Research	6
Schlussfolgerungen	7
Evaluierung von Internet-Registern mit Kontrollierten Vokabularen.....	8
Schlussfolgerung	9

Hintergrund

Ziel des Arbeitspakets „Metadaten“ ist die Erarbeitung eines Metadatenframeworks für Repositorien in Österreich und von Guidelines, die mit FAIR-Richtlinien (FAIR Guiding Principles for scientific data management and stewardship) arbeiten. Die Suche von Forschungsdaten und Projektkooperationen können repositorienübergreifend nur funktionieren, wenn ein gemeinsamer Satz von Metadaten beim Einbringen der Daten erstellt wird und wenn die eingebrachten Daten mit Kontrolliertem Vokabular beschrieben werden. Das Projekt will diese beiden Voraussetzungen schaffen.

Einleitung

Forschungsdaten sind ein Abbild der jeweiligen Wissenschaften und ihrer Eigenheiten bezüglich Zusammenarbeit, Organisationsgrad, Publikationsgewohnheiten etc. Ebenso uneinheitlich präsentieren sich die vielen Metadatenstandards: einige stellen einen fachübergreifenden Anspruch, andere legen ihren Fokus auf eine oder wenige Wissenschaften.

Auch die Sacherschließung als Teil der Metadaten ist heterogen. Es gibt viele verschiedene Kontrollierte Vokabulare/Wissensorganisationssysteme (Schlagwortnormdateien, Thesauri, Klassifikationen etc.), darunter solche mit universellem Anspruch (z. B. Dewey Decimal Classification (DDC), Universal Decimal Classification (UDC), die Regensburger Verbundklassifikation (RVK) und die Gemeinsame Normdatei (GND) [beide vor allem im deutschsprachigen Europa]), andere wurden für die Erschließung von Literatur in einzelnen (Teil-) Wissenschaften geschaffen.

Dieser Bericht untersucht die Handhabung von Sacherschließung von Forschungsdaten an verschiedenen Institutionen, in der Regel Universitäten und Hochschulbibliotheken (d.h. in „institutionellen“ Repositorien). Die Auswahl der Institutionen und ihrer Repositorien erfolgte intuitiv nach Bedeutung. Die „spezifischen“ Repositorien, d.h. solche mit inhaltlichen Schwerpunkten, wurden nicht untersucht.

Die Zusammentragung von Informationen erfolgte durch das Besuchen der entsprechenden Homepages und der Auswertung der Beschreibungen der Forschungsdatenmanagement-Services. Diese Informationen wurden mit Stichproben in den Metadatenätzen der entsprechenden Repositorien verglichen, da die Beschreibungen der Metadaten nicht immer klar sind, und bei fakultativen Anwendungen interessant ist, ob die Möglichkeit von den Forschenden auch genutzt wird. Die untersuchten Features waren:

- Gibt es die Möglichkeit einer Sacherschließung? (Def. Sacherschließung: thematische Erschließung als Teil der Metadaten mittels frei oder aus einem Kontrollierten Vokabular gewählten Schlagwörtern.)
- Falls ja, ist sie zwingend oder fakultativ?
- Finden freie Schlagwörter und/oder solche aus Kontrollierten Vokabularen Anwendung?
- Welche kontrollierten Vokabulare werden verwendet? Werden irgendwelche Kontrollierten Vokabulare vorgeschrieben, oder sind die Forscher in der Wahl der Kontrollierten Vokabulare frei?
- Wird ein übergeordneter Standard (z.B. eine Klassifikation) verwendet, wenn ja, welche; geschieht das obligatorisch oder fakultativ?
- Wie ist die Akzeptanz der Sacherschließung bei den Forschenden?
- Wohin geht die Tendenz bei der Sacherschließung von Metadaten? Was würden Sie anders machen?

Die Forschungsdatenmanagement-Dienste wurden auch per Email kontaktiert, einerseits, um Unklarheiten zu beseitigen, andererseits auch, um Begründungen für die entsprechende Ausgestaltung der Services und zugrundeliegender Regeln zu erfragen. Von neun kontaktierten Institutionen haben sieben geantwortet. Ich danke den Personen, die mir bereitwillig Auskunft gegeben haben, herzlich für Ihre Bemühungen!

Im zweiten Teil dieses Deliverable werden Internet-Register auf ihre Eignung untersucht, den Forschern einen Überblick über die zur Verfügung stehenden Ressourcen für die Sacherschließung zu geben (Kontrollierte Vokabulare als Klassifikationen, Thesauri, Normdateien etc.)

Sacherschließung in institutionellen Repositorien

Universität Heidelberg

Das Forschungsdatenrepositorium der Universität Heidelberg, HeiDATA,¹ bietet neben einer groben thematischen Kategorisierung (von der Software Dataverse Version 4 vorgegeben) zwei Felder, die zur inhaltlichen Erschließung genutzt werden können: die „Keywords“ sind frei gewählte Schlagwörter, die „Topic Classification“ kann mit Schlagwörtern aus Kontrollierten Vokabularen befüllt werden. Ein Zwang zur Verwendung dieser Felder besteht nicht, d. h. die inhaltliche Erschließung kann bis auf die grobe Kategorisierung nach Fachgebiet entfallen. Bis Version 3 von Dataverse wurde eine thematische Erschließung nach Dewey (DDC) vorgenommen, ab der aktuellen Version aber eingestellt, und die in der Software angelegte Fachklassifikation übernommen. Die Metadatenätze in HeiDATA werden auch in den Verbundkatalog (K10plus)² eingespielt, dort besteht für die Fachreferenten die Möglichkeit einer Beschlagwortung. Im Katalog gibt es für die Forschungsdaten unter „Dokumenttyp“ eine eigene Facette.

Humboldt Universität zu Berlin

Forschungsdaten werden an der Humboldt-Universität nicht in einem eigenen Forschungsdaten-Repositorium, sondern im allgemeinen Repositorium archiviert.³ Der verwendete Metadatenstandard ist Dublin Core. Die Erschließung mit Dewey (DDC) ist obligatorisch. Auch die Vergabe von freien Schlagwörtern ist Pflicht (Minimum vier deutsche und vier englische Begriffe), die Verwendung von Kontrollierten Vokabularen hingegen fakultativ. Die Datensätze sind auch in der Suchmaschine der Bibliothek eingespielt. In der Benutzersicht sind freie Stichwörter (Dublin-Core-Kategorie dc.subject) und Schlagwörter aus Kontrollierten Vokabularen (hier Dewey DDC: de.subject.dc) nicht unterscheidbar. Die Forschungsdaten-Homepage des edoc-Dienstes verlinkt auf ein Portal mit Kontrollierten Vokabularen: bartoc.org (siehe auch zweiter Teil des Deliverable). Im Verständnis der RDM-Dienstes der HU ist der Einsatz von Kontrollierten Vokabularen sinnvoll und Teil des „F“ in „FAIR“. Institutionelle Repositorien von klassischen Hochschulen mit einem umfassenden Fächerkanon würden sich mit der Betreuung der Metadaten angesichts der vielen verschiedenen Standards und Kontrollierten Vokabularen aber schwerer tun als fachlich spezialisierte Repositorien. Die Humboldt-Universität zu Berlin empfiehlt ihren Forschenden deshalb, Ihre Forschungsdaten in entsprechenden Repositorien zu hinterlegen. Die Forschungsdaten sind auch im Bibliothekskatalog nachgewiesen.⁴

¹ <https://heidata.uni-heidelberg.de/>

² <https://www.ub.uni-heidelberg.de/helios/kataloge/heidi.html>

³ <https://edoc.hu-berlin.de/>

⁴ https://hu-berlin.hosted.exlibrisgroup.com/primo-explore/search?vid=hub_ub&fromLogin=true&lang=de_DE

Eidgenössische Technische Hochschule Zürich (ETHZ)

Die ETH schreibt ihren Forschern bei der Archivierung der Forschungsdaten auf ihrem Repositoryum „Research Collection“⁵ (Metadatenstandard: Dublin Core) keine inhaltliche Erschließung vor. Laut Auskunft der ETH ist das Interesse seitens der Wissenschaftler gering. Die inhaltliche Erschließung sei bei einem breit angelegten Repositoryum nicht so wichtig wie bei einem fachspezifischen Repositoryum. Die Forschungsdaten würden eher über die darauf aufbauenden Publikationen gesucht. Die ETH bemüht sich deshalb um die Verlinkung zwischen wissenschaftlicher Publikation und zugrundeliegender Forschungsdaten.

Die Metadaten der Forschungsdaten scheinen nicht nur in der „Research Collection“ (kein reines Forschungsdaten-Repositoryum, sondern das allgemeine Repositoryum) auf, sondern auch im Bibliothekskatalog der ETH Zürich.⁶

Auf Nachfrage bestätigt die Mitarbeiterin der ETH Zürich, dass den Forschenden von ihrem Service in Beratungen und Workshops nahegelegt wird, fachspezifische Repositoryen für die Publikation ihrer Forschungsdaten zu favorisieren. In Ermangelung einer Forschungsdaten-Policy wird das aber nicht offiziell (z.B. auf der Homepage) kommuniziert.

Universität Tübingen

Im Forschungsdatenmanagement-Repositoryum FDAT⁷ der Universität Tübingen werden den Forschenden auf drei Ebenen jeweils verschiedene Sacherschließungselemente vorgeschrieben: auf der Projektebene ist eine Zuordnung zu einem Wissenschaftsgebiet Pflicht (der Begriff ist frei wählbar oder einer Wissenschaftssystematik zu entnehmen), auf Ordnerbene eine Inhaltsbeschreibung, auf Einzelressourcenebene Schlagwörter.

Die Kategorie „Keywords“ beinhaltet in der Regel freie Schlagwörter, kann aber auch für Schlagwörter aus Kontrollierten Vokabularen verwendet werden. Deren Verwendung wird empfohlen, ist aber aufgrund fehlender Standards in den einzelnen Disziplinen oft schwierig in Auswahl und Anwendung. Laut Information des FDM-Service der Universität Tübingen ist seitens der Forschenden eine Bereitschaft zur sorgfältigen Erschließung der Forschungsdaten erkennbar. Die Forschungsdaten werden nur im Repositoryum FDAT geführt. In den Katalog der Universitätsbibliothek findet kein Import der Metadatensätze statt, die Forschungsdaten würden von allerdings von den Suchmaschinen geharvested. Die Suche nach Forschungsdaten im Bibliothekskatalog sei wenig üblich.

Auch die Verwendung von Dewey (DDC) wurde evaluiert. Die Suche nach Forschungsdaten über Suchbegriffe aus einer Klassifikation wurde aber auch als nicht sehr wahrscheinlich eingeschätzt, die Suche mittels fachspezifischer Terminologie erschien plausibler. Deshalb und aus Gründen der Ressourcenknappheit wurde auf eine Erschließung mittels einer Klassifikation (wie Dewey) verzichtet.

Oxford University

Das Oxford University Research Archive ORA⁸ ist ein allgemeines Repositoryum, es enthält nicht nur Forschungsdaten. Die Metadaten der Forschungsdatensets weisen „Keywords“ als freie Schlagwörter auf, sie können aber auch einem Kontrollierten Vokabular entnommen werden. Die Erfahrungen des ORA-Teams mit der Sacherschließung durch Forschende sind nicht sehr positiv.

⁵ <https://www.research-collection.ethz.ch/>

⁶ <http://www.library.ethz.ch/de/>

⁷ <https://fdat.escience.uni-tuebingen.de/portal/#/start>

⁸ <https://ora.ox.ac.uk/>

Die Wissenschaftler würden die Publikation der Forschungsdaten oft nur machen, weil sie aufgrund der Bedingungen der Drittmittelgeber dazu gezwungen seien. Deshalb würden viele Forschende eine über die Kern-Metadaten hinausgehende Erschließung als lästig empfinden, und gegebenenfalls auf andere Repositorien ausweichen.

Zusätzlich zu den „Keywords“ gibt es noch die „Subjects“, die aus den FAST-Normdateien der OCLC genommen werden (für die Forschenden aus einem Drop-Down-Menu zu wählen). Auch sie sind fakultativ.

ORA hat somit keine Mindeststandards für die Sacherschließung von Forschungsdaten. Viele Datensätze sind entsprechend gar nicht erschlossen. Die Metadaten sind auch im Bibliothekskatalog nachgewiesen.⁹

University of Cambridge

Apollo ist das allgemeine Repository der University of Cambridge, in dem auch Forschungsdaten publiziert und archiviert werden.¹⁰ Die Forschungsdaten können mit einer Facette „Dataset“ gefiltert werden.

Einen Zwang zur Verwendung von Schlagwörtern (auch freien) gibt es nicht. Das Wissenschaftskommunikationsteam von Apollo meint, eine eingehende Beschreibung der Forschungsdaten durch die Forschenden würde genug geeignete Stichwörter hervorbringen, welche die Funktion von freien Schlagwörtern übernehmen. Hier ist anzumerken, dass selbstverständlich die in den Metadaten (Dublin Core) stehenden Beschreibungen mit einer Suche abgedeckt werden, aber nicht die teilweise ausführlicheren Beschreibungen in README-Dateien, welche den Forschungsdaten beigelegt wurden.

Der Wissenschaftskommunikations-Dienst der Universität Cambridge ist auch der Ansicht, dass es schwierig sein kann, für neuartige Forschungsthemen überhaupt geeignete kontrollierte Vokabulare zu finden. Das spreche gegen eine Pflicht zur vertieften Erschließung.

Die Metadaten der Forschungsdaten sind auch über den Bibliothekskatalog suchbar.¹¹

Delft University of Technology /Eindhoven University of Technology/University of Twente/Wageningen University and Research

Die vier Hochschulen betreiben ein gemeinsames Repository: 4TU.Centre for Research Data.¹² Durch die naturwissenschaftlich-technische Ausrichtung der teilnehmenden Hochschulen weist auch das Repository entsprechende Schwerpunkte auf. Die Data Stewards der vier Hochschulen empfehlen zwar die Verwendung fachspezifischer Repositorien, durch die Ausrichtung des 4TU.Centre for Research Data auf Naturwissenschaften und Technik ist 4TU.Centre for Research Data aber selbst ein zwar institutionelles, aber auch fachspezifisches Repository mit entsprechend vertiefter Fachexpertise. Die Sacherschließung wird deshalb auch nicht nur den Forschenden überlassen. Sie findet auf zwei Ebenen statt: die Kategorie „Keywords“ kann sowohl für frei wählbare Schlagwörter, als auch für solche aus kontrollierten Vokabularen verwendet werden. Die freien Schlagwörter kommen in der Regel von den Forschenden selbst, bei den kontrollierten Vokabularen helfen meist die Data Stewards. Forschende hätten nicht die Zeit und die Kenntnis für eine adäquate Beschlagwortung.

⁹ http://solo.bodleian.ox.ac.uk/primo-explore/search?vid=SOLO&lang=en_US&sortby=rank

¹⁰ <https://www.repository.cam.ac.uk/>

¹¹ http://idiscover.lib.cam.ac.uk/primo-explore/search?vid=44CAM_PROD&lang=en_US

¹² <https://researchdata.4tu.nl/en/use-4turesearchdata/find-research-data/>

Für eine Zuweisung an ein Fachgebiet verwendet 4TU.Centre for Research Data die Australian and New Zealand Standard Research Classification (ANZSRC). Diese hat sich in einer Evaluierung gegen Dewey (DDC) und die Aufstellung in re3data durchgesetzt, weil die ANZSRC die technischen Disziplinen besser abdeckt. Außerdem wollte man mit einem Schema arbeiten, das zwei Ebenen besitzt.

Ein willkommener Effekt einer flächendeckenden Sacherschließung bzw. Zuweisung an Wissenschaftsgebiete ist die Möglichkeit für den Benutzer, über die Homepage im Repositorium nach Fachgebieten zu browsen. Das Abbilden der Forschungsschwerpunkte auf der Homepage trägt zur Identity der vier Hochschulen im Hinblick auf deren stakeholder bei.

Schlussfolgerungen

- Die inhaltliche Erschließung der Forschungsdaten soll hervorheben, dass Forschungsdaten nicht ein Abfallprodukt der Forschung sind. Die methodischen Vorüberlegungen und der Prozess der Datenerhebung stellen eine grundlegende Phase des Forschungsprozederes dar. Die Qualität und Bedeutung der Forschungsdaten wird durch die Publikation und ein angemessenes Set von Metadaten inklusive Sacherschließung unterstrichen.
- Viele FDM-Services raten ihren Forschenden explizit dazu, die Forschungsdaten auf fachspezifischen Repositorien zu archivieren, und ihr eigenes, institutionelles Repositorium nur zu benützen, wenn es keine spezifischen Repositorien für ihr Fach gibt, oder das institutionelle Repositorium als parallele Publikation/Zweitarchiv genutzt werden soll. Die einschlägigen Repositorien würden mit den spezifischen Datenformaten besser umgehen können, und auch die inhaltliche Erschließung im Zweifelsfall besser leisten können, als thematisch breiter aufgestellte institutionelle Repositorien.
- Die meisten Institutionen lassen die Metadaten der Forschungsdaten in den Bibliothekskatalog einfließen. Es gibt (z.B. Universität Tübingen) aber solche, die sich bewusst dagegen entschieden haben. Die Suche nach Metadaten im Bibliothekskatalog sei nicht der übliche Weg: bei der Frage nach Forschungsdaten-Metadaten in Bibliothekskatalogen spielt aber wohl auch die Sichtbarkeit der Forschung eine große Rolle, vor allem, wenn die Forschungsdaten (wie von vielen FDM-Diensten geraten) extern auf fachspezifischen Repositorien publiziert und archiviert werden.
- F in FAIR für Findable (Hinweis des FDM-Dienstes der Humboldt Universität zu Berlin). Die Publikation/Archivierung von Forschungsdaten dient zu mehr als nur dem Nachvollzug von Experimenten und der Falsifikation von Hypothesen. Die Publikation von Forschungsdaten dient auch deren Nachnutzung. Eine Suchbarkeit nach thematischen Kriterien ist für die interdisziplinäre Forschung wichtig.
- Eine vertiefte inhaltliche Erschließung ist prinzipiell wünschenswert, aber eine Frage der personellen Ressourcen. Eine über freie Schlagwörter hinausgehende Erschließung können Forschende in der Regel nicht leisten. Für die Auswahl und den Umgang mit Kontrollierten Vokabularen braucht es Kenntnisse. Forscher haben durch ihre sporadische Beschäftigung mit dem Thema der Erschließung meistens wenig Übersicht über das Angebot und oft ungenügend Routine in der Anwendung. Die Forschenden brauchen deshalb zumindest Hilfestellung; eine gänzliche Übernahme der Erschließung scheint bei entsprechend zur Verfügung stehenden personellen Ressourcen sinnvoll zu sein. Hochschulbibliotheken mit einem gut ausgebauten Fachreferenten-Dienst könnten diesen als Basis für eine entsprechende Dienstleistung nehmen. Allerdings wird AI zukünftig wohl einiges an Sacherschließungsarbeit leisten können (man denke nur an die Fortschritte bei der Bilderkennung).
- Im Anschluss an diese zwei Punkte stellt sich auch die Frage nach der Voreingenommenheit der Forschenden als sacherschließende Instanz. Eine außenstehende Person mit einem Blick

auf eine interdisziplinäre Nachnutzung der Forschungsdaten beschlagwortet eventuell unter anderen Gesichtspunkten als der Forschende, der die Daten zu einem bestimmten Zweck erhoben, und einen entsprechend fokussierten Blick auf seine Datensammlung hat.

- Mindeststandards in der Sacherschließung: sinnvoll erscheint eine obligatorische freie Beschlagwortung durch die Forschenden (evt. mit einer Mindestanzahl von Begriffen in vorgeschriebenen Sprachen, darunter selbstverständlich Englisch). Darüber hinaus können Forschende kaum zu einer Verwendung von Kontrollierten Vokabularen gezwungen werden (siehe oben). Auch die Auswahl der fachspezifischen Vokabulare sollte in Abstimmung zwischen Forschenden und FDM-Spezialisten geschehen.
- Die Verwendung eines global akzeptierten Klassifikationssystems (meist Dewey [DDC]) wird kontrovers beurteilt. Neben der Suchfunktion (deren Wert bei Dewey [DDC] und anderen Klassifikationen teilweise in Frage gestellt wird) bietet die Zuweisung an (Teil-) Wissenschaften die Möglichkeit, die Schwerpunkt der Forschung auf einer Homepage offensichtlich zu machen und so Interessierten einen Überblick zu ermöglichen.
- Ein Detail: zwei (fakultative) Sacherschließungsfelder, die einfach zu handhaben sind, und wertvolle Suchkriterien darstellen dürften, sind Angaben von Zeit und Ort.

Evaluierung von Internet-Registern mit Kontrollierten Vokabularen

Nicht alle hier besprochenen Internet-Register listen tatsächlich Kontrollierte Vokabulare auf. Ich erwähne sie aber trotzdem, da einige von ihnen bedeutende Sammlungen von Metadatenstandards sind, bei denen man auch Kontrollierte Vokabulare vermuten könnte (einige Informations-Seiten von RDM-Services sind diesbezüglich missverständlich).

- <http://www.bartoc.org/de>

Das Basel Register of Thesauri, Ontologies and Classifications (BARTOC), seit 2013 aufgebaut und unterhalten von der UB Basel, listet 2988 (Stand 03.07.2019) Wissenschaftsorganisationssysteme/ Kontrollierte Vokabulare auf. Die Suche geht über alle Felder. Die angezeigten Treffer werden nach Dewey (DDC) (Klassifikation), EuroVoc (Klassifikation), nach KOS Types (Typen von Wissensorganisationssystemen wie Ontologien, Thesauri etc.), Sprache, Land und Format (z.B. gedruckt, PDF, MARC21 etc.) facettiert. Mit denselben Kategorien lässt sich auch der Gesamtbestand filtern (Browse).

Die Datensätze sind übersichtlich gehalten, und bestehen aus dem Titel, einer Beschreibung des Vokabulars (meist Abstract in Englisch), Link, verantwortlicher Institution (mit Nachweis im Virtual Authority File VIAF), Verlinkung zum entsprechenden Artikel auf Wikipedia falls vorhanden, und den oben genannten Klassifikationsbegriffen. Die populärsten (top-rated) Wissensorganisationssysteme werden zusätzlich mit der Integrative Levels Classification (ILC) erschlossen. Die ILC ist ein Klassifikationssystem, das nicht auf Hierarchien, sondern auf frei miteinander kombinierbaren Konzepten aufbaut.¹³

Darüber hinaus werden 89 Internet-Seiten (Registries) aufgeführt, die ähnlich wie BARTOC eine Übersicht über Wissenschaftsorganisationssysteme bieten. Diese sind aber in ihrer großen Mehrheit fokussierter (z. B. auf spezifische Wissenschaften oder Anwendungen) als BARTOC.

- <https://fairsharing.org/>

¹³ <http://www.iskoi.org/ilc/>

FAIRsharing ist ein Portal des e-Research-Centers an der Uni Oxford (gegründet 2011 auf der Basis eines Vorgängerprojektes). Es bietet neben Metadaten-Standards auch Kontrollierte Vokabulare (Terminology Artifact: 728). Die Suche geht über alle Felder. Die Treffer werden facettiert nach Typ (z.B. Terminology Artifact), Domain, Subjects (Wissenschaften), Taxonomies, Land u.a. Veraltete Standards werden ebenfalls (als solche) angezeigt. In den Datensätzen wird auch Literatur zum Vokabular angegeben. Die Sprachen der Vokabulare werden nicht angeführt, mindestens eine Version muss aber auf Englisch vorliegen.

Eine weitere, interessante Facettierungsmöglichkeit ist der „Recommended“-Tag. Mit diesem Attribut können Kontrollierte Vokabulare als von Institutionen oder wissenschaftlichen Publikationen empfohlen gekennzeichnet werden. Zur Orientierung für Wissenschaftler und Mitarbeiter von FDM-Diensten wäre es ein wertvolles Feature. Leider wird das Attribut selten vergeben.

- <http://www.dcc.ac.uk/resources/metadata-standards>

Die Homepage des Digital Curation Centers (DCC) verzeichnet keine Vokabulare.

- <http://rd-alliance.github.io/metadata-directory/subjects/>

Die Homepage zählt Metadaten-Standards auf, auch sogenannte Tools, aber Vokabulare werden nicht aufgelistet, sondern allenfalls als „related vocabulaires“ in der Beschreibung der Standards genannt. Die Verlinkungen sind veraltet, wenn sie überhaupt aufgeführt sind. Als Vorschlag einer Sammlung von Vokabularen für die Sacherschließung wenig geeignet.

- <https://lov.linkeddata.es/dataset/lov>

Dieses Internet-Register zählt 671 linked open data vocabularies auf. Die Suche geht nur über den Titel, nicht über die Beschreibung. Damit ist die Suchbarkeit von Vokabularen nicht gut gewährleistet. Es gibt nur eine alphabetische Auflistung nach Akronymen, aber keine Kategorisierung nach Disziplinen oder Wissenschaften. Die gängigen Kontrollierten Vokabulare (z.B. RSWK, DDC Dewey) sind nicht enthalten. Auf linked open data vocabularies spezialisiert, und zur Suche eher ungeeignet.

- https://en.wikipedia.org/wiki/Outline_of_academic_disciplines

Die Übersicht in Wikipedia listet Wissenschaften und Teilwissenschaften auf. Der Grad der Auffächerung ist nicht ganz ausgewogen, man vergleiche die wenige Teilbereiche umfassenden „History“ und „Language and Literature“ mit den umfangreichen „Sociology“ oder „Computer Science“. Deshalb ist diese Aufstellung in ihrer Gesamtheit eher nicht verwendbar. Eine Visualisierung als privates Projekt gibt es auch.¹⁴

Schlussfolgerung

Als Übersicht über Wissensorganisationssysteme, die Forschenden als Informationsquelle für die adäquate Sacherschließung ihrer Forschungsdaten zur Verfügung gestellt werden sollen, kommen nur BARTOC und FAIRsharing in Frage. BARTOC hat den größeren Umfang und ist einfacher zu handhaben (das ist mein subjektiver Eindruck). FAIRsharing hat das attraktive Feature des „Recommended“-Tags, das aber selten genutzt wird. Prinzipiell können aber beide den Forschenden auf einer FDM-Homepage vorgeschlagen werden.

¹⁴ http://www.thingsmadethinkable.com/item/fields_of_knowledge.php