

Daten über Daten

Metadatenstandards und Hilfsmittel

Christiane Bayer (IT-Gruppe Geisteswissenschaften, LMU)

Vanessa Finger (Universitätsbibliothek der LMU)

Martin Spenger (Universitätsbibliothek der LMU)



Metadaten für Forschungsdaten

- Metadaten ermöglichen
 - Einbindung von Daten in Suchinfrastrukturen
 - selektives Auffinden von relevanten Datenquellen
 - maschinelle Erschließung von Datenbeständen
 - Referenzierung und Nachnutzung von Forschungsdaten
 - ...jedoch nur, wenn sie
 - auf Basis eines weit verbreiteten Metadatenschemas beschrieben werden
 - **konsistent** sind

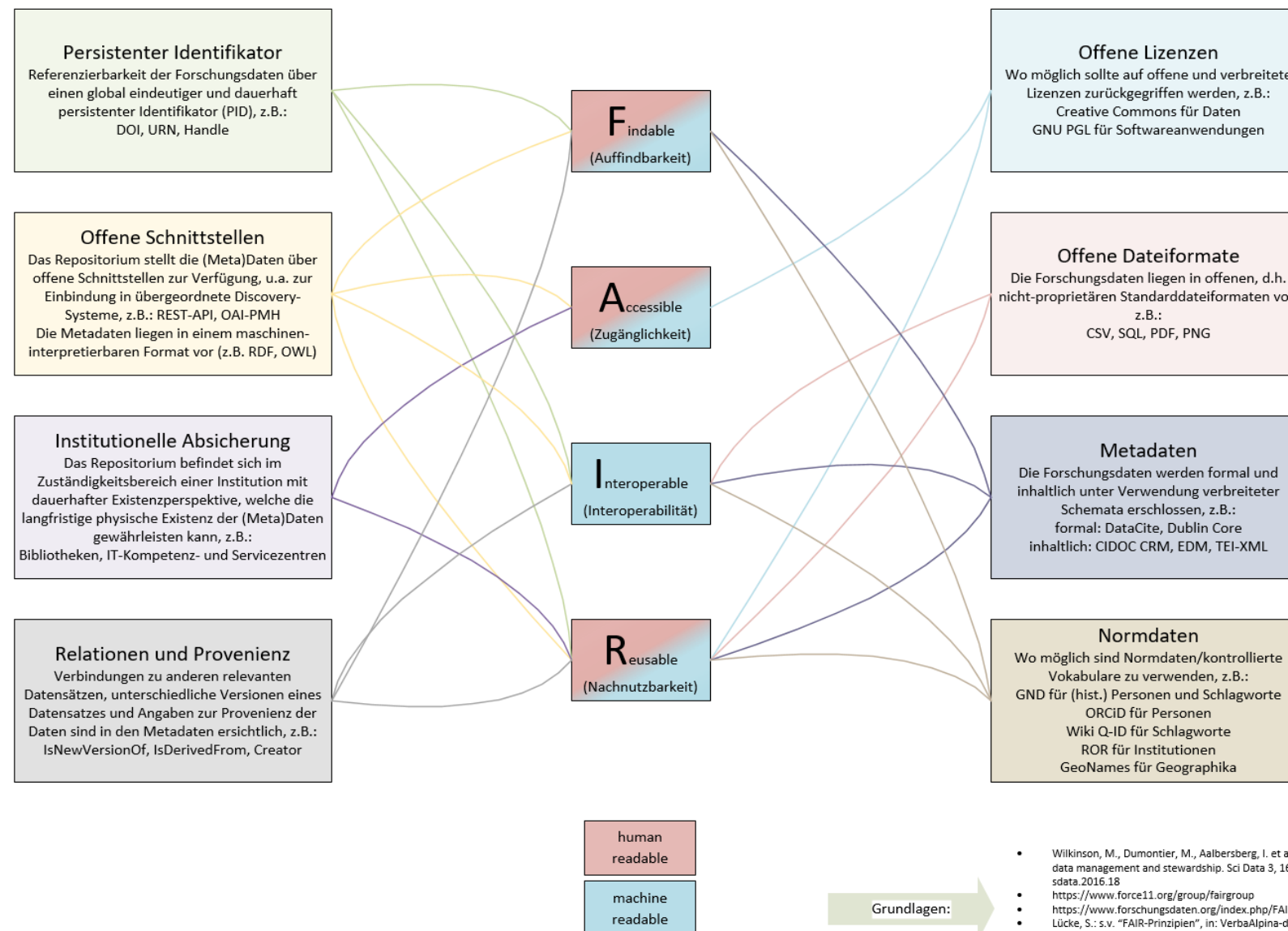
Arten von Metadaten

- Generische Metadaten
 - Formale Erschließung von Forschungsdaten
 - Enthalten bspw. Titel, Angaben zur Datenautorenschaft, zum zeitlichen Entstehungskontext etc.
 - Bekannte Vertreter: Das seit 1994 von der Dublin Core Metadata Initiative (DCMI) konzipierte [Dublin Core-Format](#) sowie das umfangreichere Metadatenschema [DataCite](#)

Arten von Metadaten

- **Inhaltserschließende Metadaten**
 - Enthalten Informationen, die durch die Analyse der Beziehungen innerhalb der beschreibenden Ressource gewonnen werden
 - Durch eine **umfassende Inhaltserschließung** werden dezidierte, fachspezifische Angaben durchsuchbar, wie bspw. wo/wann/wie etc. eine antike Stätte modelliert wurde
 - Bekannte Vertreter: [CIDOC CRM](#) für Begriffe des kulturellen Erbes oder auch das [Europeana Data Model \(EDM\)](#).

Metadaten für Forschungsdaten



Die **FAIR**-Prinzipien
als Dreh- und
Angelpunkt im FDM

Quelle: eigene Darstellung

- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- <https://www.force11.org/group/fairgroup>
- https://www.forschungsdaten.org/index.php/FAIR_data_principles
- Lücke, S.: s.v. "FAIR-Prinzipien", in: VerbaAlpina-de 18/2, Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D182%26letter%3DF%23128
- <https://blog.fid-romanistik.de/ag-digitale-romanistik/>

Entscheidung für Metadatenschema

Warum DataCite?

- Persistente Identifikation über DOI
- Weit verbreitet, über Disziplingrenzen hinweg
- Weiterentwicklung mit Feedback aus Community (PID-Forum, Roadmap, ...)
- Interoperabilität mit anderen Modellen
- Relativ ausdifferenziertes Schema mit begrenzter Möglichkeit zur inhaltlichen Beschreibung

DataCite

- Internationales Konsortium und eingetragener Verein.
- Die Bereitstellung des DataCite Metadata Schema ist eine der Kernaufgaben von DataCite und der Metadata Working Group.
 - Vision: “Connecting research, identifying knowledge.”
 - Mission: “DataCite’s mission is to be the world's leading provider of persistent identifiers for research. Through our portfolio of services, we provide the means to create, find, cite, connect, and use research. We seek to create value and develop community-driven, innovative, open, integrated, useable, and sustainable services for research.”

- Dienste:
 - Vergabestelle für Digital Object Identifier (DOI) = Persistente Identifikatoren, die eine eindeutige und URL-unabhängige Zuordnung einer digitalen Ressource erlauben.
 - Bereitstellung von forschungsunterstützenden Services.
 - Beteiligung an internationalen Projekten (FAIRsFAIR, FREYA).
 - **Herausgabe und Weiterentwicklung des gleichnamigen Metadatenschemas.**

DataCite Metadata Schema

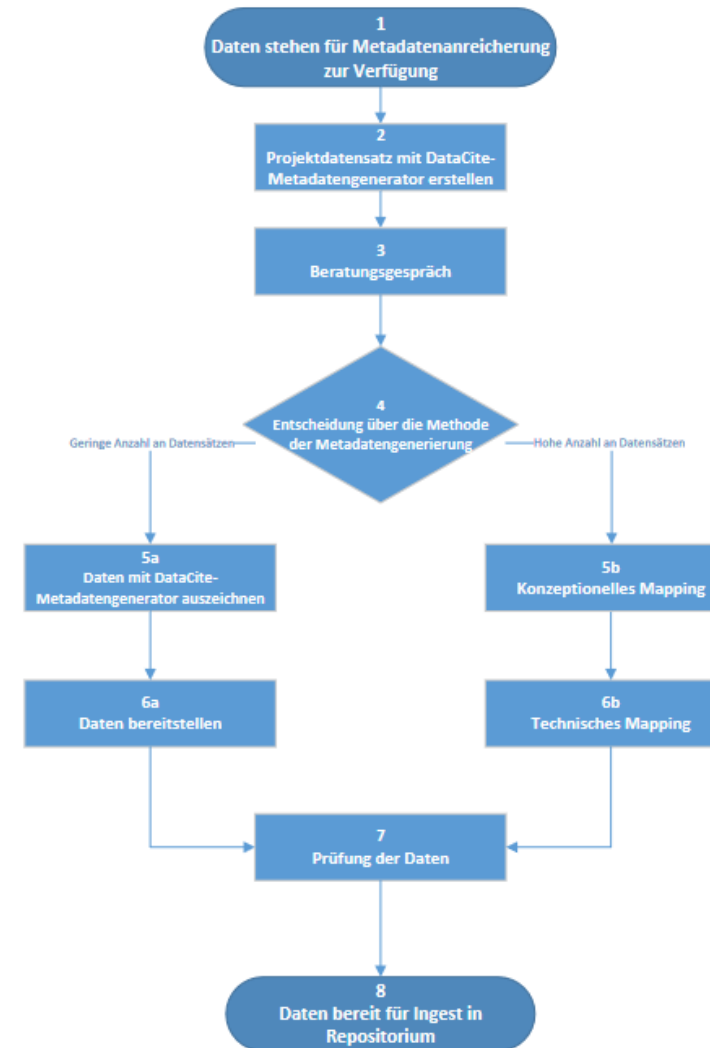
- Frei zugänglich: <https://schema.datacite.org/>
- Aktuelle Version 4.4. (veröffentlicht am 30.03.2021)
- Wird von Metadata Working Group gepflegt und weiterentwickelt
- Umfangreiche Dokumentation (82 Seiten):

https://schema.datacite.org/meta/kernel-4.4/doc/DataCite-MetadataKernel_v4.4.pdf

DataCite Metadata Schema

TITEL	KONTAKT	DATUM	SEITE
Teilworkflow Metadatenanreicherung	ITG/UB	14.09.2021	1 von 1

- Aufgrund der Komplexität des Themas wurde im Rahmen der Servicestelle FDM-DH ein Teilworkflow erstellt.



DataCite Metadata Generator

DataCite Metadata Generator - Kernel 4.3

Mandatory Elements

DOI:

[DOI]

Title(s):

[TITLE]

[LANG] [titleType]

Creator(s):

[CREATOR NAME] [nameType]

[GIVEN NAME] (optional) [FAMILY NAME] (optional)

[NAME IDENTIFIER] [nameIdentifierScheme]

[IDENTIFIER SCHEME URI]

[CREATOR AFFILIATION] [LANG]

Publisher:

[PUBLISHER]

[LANG]

Publication Year:

[YYYY]

Resource Type:

[RESOURCE TYPE] [resourceTypeGeneral]

<https://dhvlab.gwi.uni-muenchen.de/datacite-generator/>

UB-LMU / datacite-metadata-generator

<> Code

Issues

Pull requests 2

Actions

Projects

Security

Insights

master

1 branch

0 tags

Go to file

Code

ub-lmu-1 Merge pull request #6 from osmancahir/master

aa38dc8 on 4 Dec 2020 21 commits

LICENSE.md	original version	15 months ago
README.md	Update README.md	11 months ago
index.html	remove accidentally added files & fix dateinformation as tag attribute	2 months ago

README.md

datacite-metadata-generator

This is a single HTML form which can be used to generate DataCite Metadata Kernel 4.3 XML.

It is the updated version of [Marcin Paluch's](#) metadata generator. Metadata is generated by populating text boxes and selecting values from drop-downs. The results can then be saved to a file. The main goal of this form was to create a tool which does not require any installation, extra libraries, or an application server, just download and go.

This update was created by [Osman Cakir](#) from the IT-Group for the Humanities (ITG) at Ludwig-Maximilians-Universität München. The work is part of the project "eHumanities – interdisziplinär" and is mentioned in the [DataCite Best Practice Guide](#).

<https://github.com/UB-LMU/datacite-metadata-generator>

DataCite Metadata Generator

■ DataCite-XML-Datensatz:

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <resource
3    xmlns="http://datacite.org/schema/kernel-4"
4    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
5    xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4/metadata.xsd">
6    <identifier identifierType="DOI">n.v.</identifier>
7    <creators>
8      <creator>
9        <creatorName nameType="Personal">Focht, Josef</creatorName>
10       <givenName>Josef</givenName>
11       <familyName>Focht</familyName>
12       <nameIdentifier nameIdentifierScheme="GND" schemeURI="http://d-nb.info/gnd/">115059865</nameIdentifier>
13       <nameIdentifier nameIdentifierScheme="ORCID" schemeURI="https://orcid.org/">0000-0001-6053-7932</nameIdentifier>
14       <affiliation xml:lang="de">Musikinstrumentenmuseum der Universität Leipzig</affiliation>
15     </creator>
16   </creators>
17   <titles>
18     <title xml:lang="de">Bayerisches Musiker-Lexikon Online (BMLÖ)</title>
19     <title xml:lang="en" titleType="TranslatedTitle">Digital Encyclopedia of Bavarian Musicians</title>
20   </titles>
21   <publisher xml:lang="de">Universitätsbibliothek der Ludwig-Maximilians-Universität München</publisher>
22   <publicationYear>2019</publicationYear>
23   <subjects>
24     <subject xml:lang="en" schemeURI="http://dewey.info/" subjectScheme="dewey">780 Music</subject>
25     <subject xml:lang="en" schemeURI="http://dewey.info/" subjectScheme="dewey">908 History with respect to groups of people</subject>
26     <subject xml:lang="de" schemeURI="http://d-nb.info/gnd/" subjectScheme="GND" valueURI="7503845-6">Musikgeschichte</subject>
27     <subject xml:lang="en" schemeURI="https://www.wikidata.org/wiki/" subjectScheme="wikidata" valueURI="Q846047">history of music</subject>
28     <subject xml:lang="de" subjectScheme="GND" schemeURI="http://d-nb.info/gnd/" valueURI="4170825-8">Musikgeschichte (Fach)</subject>
29     <subject xml:lang="en" schemeURI="https://www.wikidata.org/wiki/" subjectScheme="wikidata" valueURI="Q10590700">music history</subject>
30     <subject xml:lang="de" subjectScheme="GND" schemeURI="http://d-nb.info/gnd/" valueURI="4132300-2">Biographieforschung</subject>
31     <subject xml:lang="en" schemeURI="https://www.wikidata.org/wiki/" subjectScheme="wikidata" valueURI="Q864365">Biographical Research</subject>

```

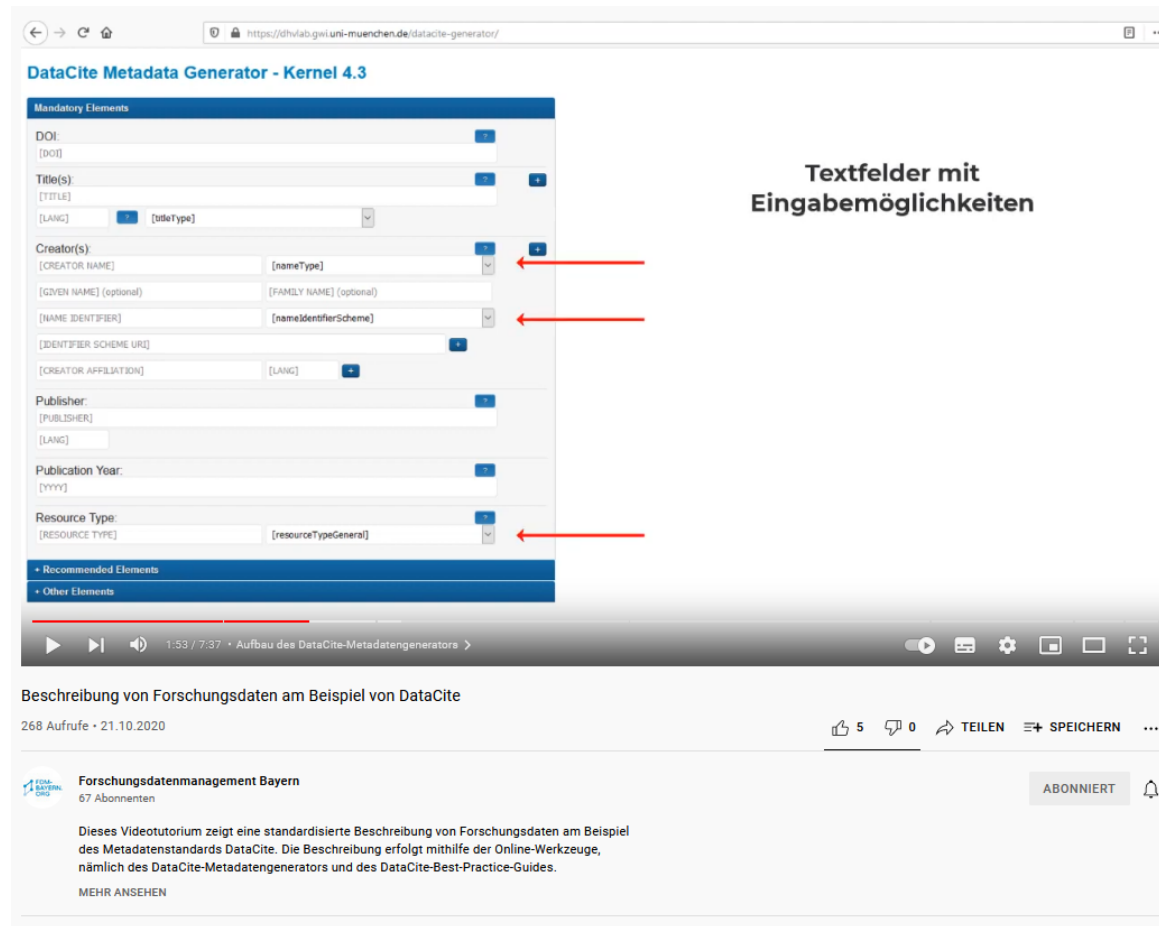
DataCite Metadata Generator

- Aufteilung in folgende Kategorien
 - Mandatory Elements
 - Recommended Elements
 - Other Elements

- Zusätzliche Funktionen des DataCite Metadata Generators:
 - Drop Down Menüs
 - „Live“-Erstellung einer XML-Metadaten-Datei (inkl. Download-Möglichkeit)
 - Erweiterte Hilfefunktionen

DataCite Metadata Generator

- Schulungsvideo zur Verwendung des Metadatengenerators



DataCite Metadata Generator - Kernel 4.3

Mandatory Elements

DOI: [DOI]

Title(s): [TITLE] [titleType]

Creator(s): [CREATOR NAME] [nameType] [GIVEN NAME] (optional) [FAMILY NAME] (optional) [NAME IDENTIFIER] [nameIdentifierScheme] [IDENTIFIER SCHEME URI] [CREATOR AFFILIATION] [LANG]

Publisher: [PUBLISHER] [LANG]

Publication Year: [YYYY]

Resource Type: [RESOURCE TYPE] [resourceTypeGeneral]

Recommended Elements

Other Elements

Textfelder mit Eingabemöglichkeiten

Beschreibung von Forschungsdaten am Beispiel von DataCite

268 Aufrufe • 21.10.2020

5 0 TEILEN SPEICHERN

Forschungsdatenmanagement Bayern
67 Abonnenten

ABONNIERT

Dieses Videotutorium zeigt eine standardisierte Beschreibung von Forschungsdaten am Beispiel des Metadatenstandards DataCite. Die Beschreibung erfolgt mithilfe der Online-Werkzeuge, nämlich des DataCite-Metadatengenerators und des DataCite-Best-Practice-Guides.

MEHR ANSEHEN

https://www.youtube.com/playlist?list=PL9ZaxDtkP6tz9dIppvNyLNj_xsUgdkMdyZ

DataCite Metadata Generator

Creator(s):

[CREATOR NAME]	[nameType]
[GIVEN NAME] (optional)	[FAMILY NAME] (optional)
[NAME IDENTIFIER]	[nameIdentifierScheme]
[IDENTIFIER SCHEME URI]	[nameIdentifierScheme]
[CREATOR AFFILIATION]	GND
	ORCID



Creator

- This field is mandatory.
- Consult section on [how to specify a person](#).
- Always prefer natural persons over institutions. If no natural person can be identified, you must set the attribute *nameType* to "Organizational".

Example

```
<creators>
  <creator>
    <creatorName nameType="Personal">Krefeld, Thomas</creatorName>
    <givenName>Thomas</givenName>
    <familyName>Krefeld</familyName>
    <nameIdentifier
      schemeURI="http://d-nb.info/gnd/"
      nameIdentifierScheme="GND">
      123778689
    </nameIdentifier>
    <nameIdentifier
      schemeURI="http://orcid.org/"
      nameIdentifierScheme="ORCID">
      0000-0001-9657-6052
    </nameIdentifier>
    <affiliation
      xml:lang="de">
      Institut für Romanische Philologie, Ludwig-Maximilians-Universität München
    </affiliation>
  </creator>
</creators>
```

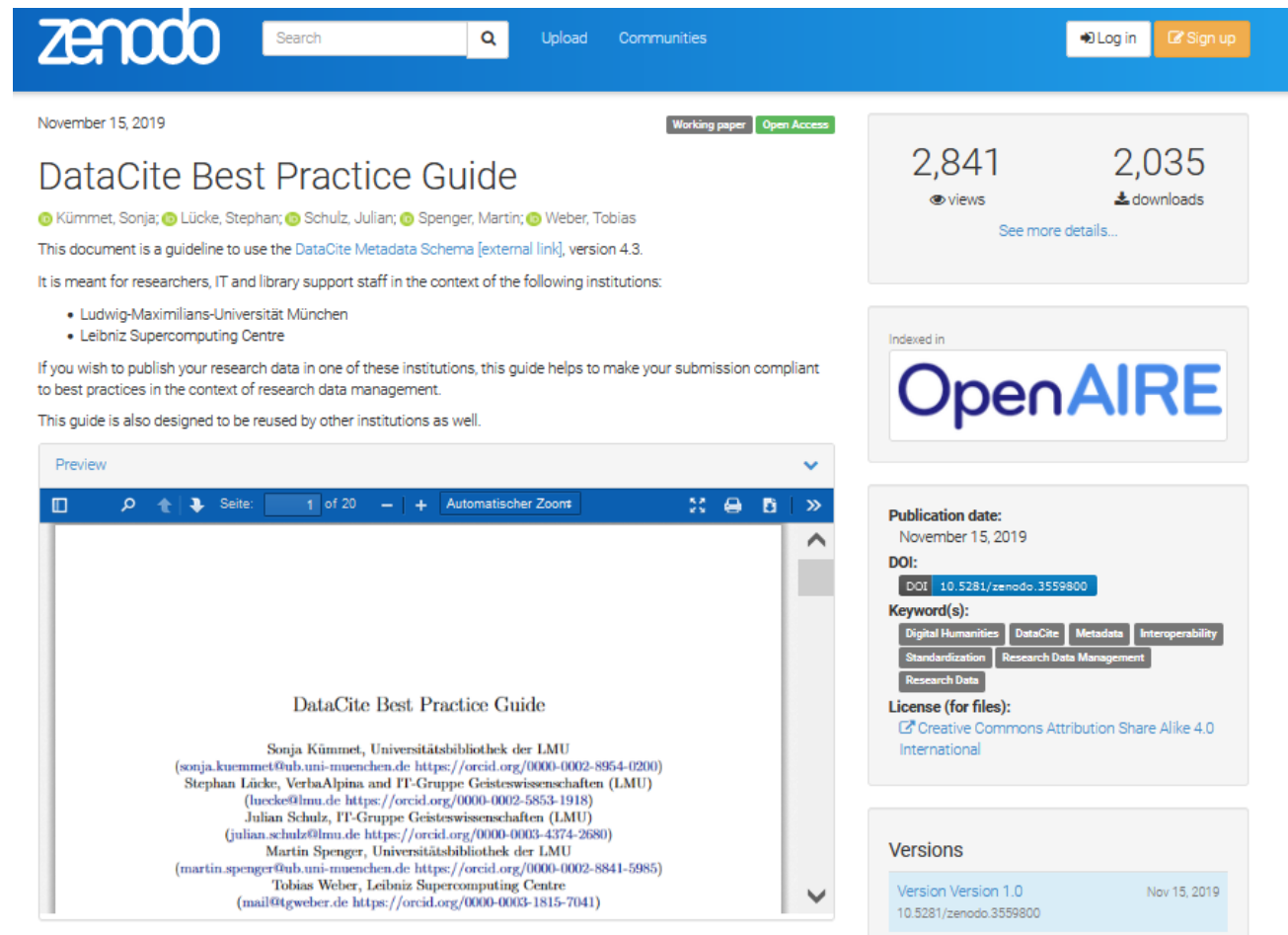
Best-Practice-Guide für DataCite - Ausgangslage

- Projekte der ITG weitgehend noch nicht mit Metadaten erschlossen
- Projektverantwortliche zeichneten ihre Projekte mit Hilfe des DataCite-Generators in DataCite aus
- Überprüfung der DataCite-XML-Files ergab: Trotz Verwendung des gleichen Metadatenschemas uneinheitliche, wenig interoperable Metadaten

Best-Practice-Guide für DataCite

- Zielgruppen:
 - Datenproduzenten (= Forschende)
 - Datenkuratoren
 - Infrastrukturanbieter

- Aufbau:
 - „General Best Practice“
 - „Best Practice for specific fields (19 Bereiche) mit Beispielen“
 - Exemplarische DataCite-Datensätze



The screenshot shows the Zenodo interface for the 'DataCite Best Practice Guide'. The header includes the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. The document is dated November 15, 2019, and is marked as a 'Working paper' and 'Open Access'. It has 2,841 views and 2,035 downloads. The document is indexed in OpenAIRE. The publication date is November 15, 2019, and the DOI is 10.5281/zenodo.3559800. The keywords are Digital Humanities, DataCite, Metadata, Interoperability, Standardization, Research Data Management, and Research Data. The license is Creative Commons Attribution Share Alike 4.0 International. The versions section shows Version 1.0 from November 15, 2019.

DataCite Best Practice Guide

Kümmet, Sonja; Lücke, Stephan; Schulz, Julian; Spenger, Martin; Weber, Tobias

This document is a guideline to use the [DataCite Metadata Schema \[external link\]](#), version 4.3.

It is meant for researchers, IT and library support staff in the context of the following institutions:

- Ludwig-Maximilians-Universität München
- Leibniz Supercomputing Centre

If you wish to publish your research data in one of these institutions, this guide helps to make your submission compliant to best practices in the context of research data management.

This guide is also designed to be reused by other institutions as well.

Preview

Seite: 1 of 20 Automatischer ZOOM

DataCite Best Practice Guide

Sonja Kümmet, Universitätsbibliothek der LMU
(sonja.kuemmet@ub.uni-muenchen.de <https://orcid.org/0000-0002-8954-0200>)

Stephan Lücke, VerbaAlpina and IT-Gruppe Geisteswissenschaften (LMU)
(luecke@lmu.de <https://orcid.org/0000-0002-5853-1918>)

Julian Schulz, IT-Gruppe Geisteswissenschaften (LMU)
(julian.schulz@lmu.de <https://orcid.org/0000-0003-4374-2680>)

Martin Spenger, Universitätsbibliothek der LMU
(martin.spenger@ub.uni-muenchen.de <https://orcid.org/0000-0002-8841-5985>)

Tobias Weber, Leibniz Supercomputing Centre
(mail@tweber.de <https://orcid.org/0000-0003-1815-7041>)

Publication date: November 15, 2019

DOI: DOI 10.5281/zenodo.3559800

Keyword(s): Digital Humanities, DataCite, Metadata, Interoperability, Standardization, Research Data Management, Research Data

License (for files): [Creative Commons Attribution Share Alike 4.0 International](#)

Versions

Version Version 1.0 10.5281/zenodo.3559800 Nov 15, 2019

<https://doi.org/10.5281/zenodo.3559800>

Best-Practice-Guide für DataCite

■ Innovative Veröffentlichungsform:

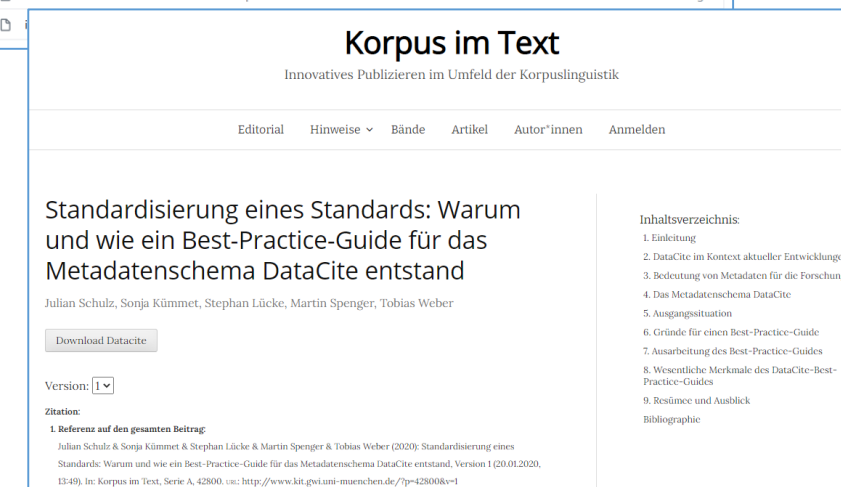
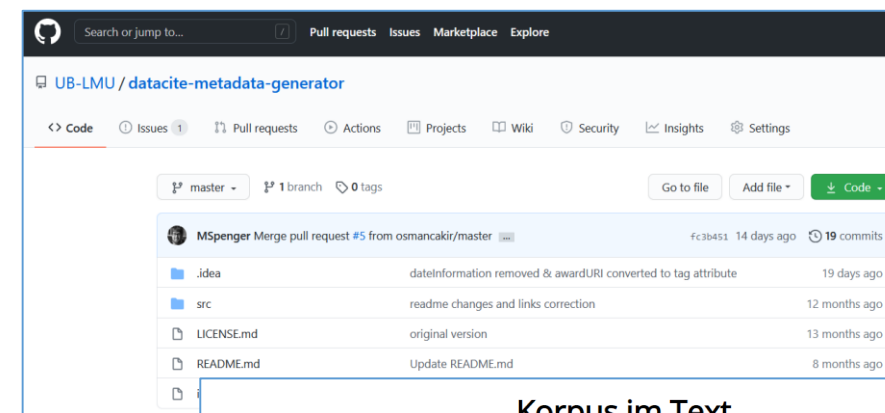
- [Zenodo](#): Best Practice Guide
- [GitHub](#): Guide und Beispiel-XML-Datensätze
- [Korpus im Text](#): Begleitpublikation (deutsch/englisch)

■ Nachnutzung des Best Practice Guides und Metadaten-Generators:

- Hohe Zugriffszahlen und Anfragen zur Nachnutzung
- Metadatengenerator u.a. an folgenden Einrichtungen im Beratungsangebot:



UNIVERSITÄT ZU LÜBECK



Merkmale des Best-Practice-Guides

- **Präzisierung** des DataCite-Standards: Vorgegebener Rahmen wird enger gefasst und konkretisiert
- **Spezifikation** von Konventionen bei mehrdeutigen Eingabemöglichkeiten
- **Normierung** der Eingabemöglichkeiten: Normvokabular anstelle tendenziell heterogener Stichwörter
- **Konformität** zum DataCite-Schema wird gewahrt

Mehrwert hinsichtlich Interoperabilität

- **Homogene Daten erleichtern Aggregation und Nachnutzung**
 - Konvertierung in maschineninterpretierbare Formate (z. B. RDF)
 - Erleichterter Austausch zwischen Forschungsdateninfrastrukturen
 - Erleichtertes Mapping auf andere Metadatenschemata

- **Standardisierung der Metadaten führt zu besserer Sichtbarkeit von Forschungsdaten**
 - Beitrag zu besserer Auffindbarkeit und Nachnutzbarkeit im Sinne der FAIR-Prinzipien
 - Gesteigerter Impact für Datenproduzenten

Übergeordnete Suchinfrastrukturen

- (Meta-)Daten werden in übergeordnete Suchinfrastrukturen eingebunden, z. B.:
 - Google Dataset Search
(<https://datasetsearch.research.google.com/>)
 - BASE (<https://www.base-search.net/>)
 - DataCite Commons
(<https://commons.datacite.org/>)

Google

Dataset Search

Nach Datensätzen suchen

Probieren Sie Coronavirus (bzw. COVID-19) oder education outcomes site: data.gov.

Weitere Informationen zu Dataset Search

Publication Year

☐ 2021
☐ 2020

3
 1,712
 2,198
 2,569
 1,913
 1,642
 3,267
 4,580

Standard-Suche Erweiterte Suche Browse Suchhistorie

Anmelden Deutsch



 BASE

 Bielefeld Academic Search Engine

Suchen Sie nach 154.341.311 Dokumenten aus 3.435 Quellen

Über BASE
 Was ist BASE?
 Quellen
 Statistik
 News
 Kontakt
 Impressum

Services / Hilfe
 Quelle vorschlagen
 Quelle validieren (DOI-URL)
 ORCID-ID eingeben
 JAR (PDF-Scanservice)
 GAZ (Schritt-für-Schritt)
 Weitere Tools

Weitere Infos
 Info
 FAQ
 Twitter
 BASE Blog
 GAZ Blog
 Login / Downloads



 ORCID

 Connect your research



 Pilotage

 Organisation

© by Universitätsbibliothek Bielefeld 2004-2019 | Powered by SciCloud & VuFind

DataCite Commons

Type to search...

This Page Works People Organizations

<https://ror.org/05591te55>

Ludwig-Maximilians-Universität München
 LMU

Founded 1472

Links
 Homepage
 Wikipedia
 Twitter

Other Identifiers
 GRID grid.5252.0
 Crossref Funder ID 10.13039/501100005722
 Crossref Funder ID 10.13039/501100006192
 ISNI 000000041936973X
 Wikidata Q55044

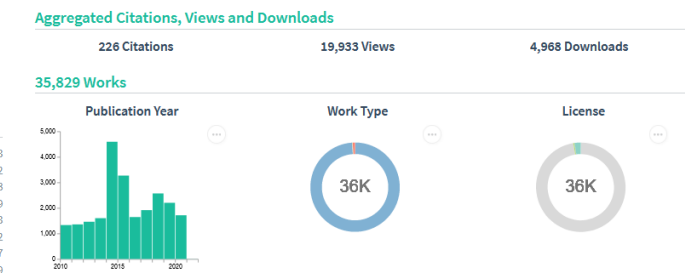
Geolocation
 48° 09' 03" N, 11° 34' 49" W

Germany Education DataCite Member

<https://ror.org/05591te55>

Share

Email
 Twitter
 Facebook

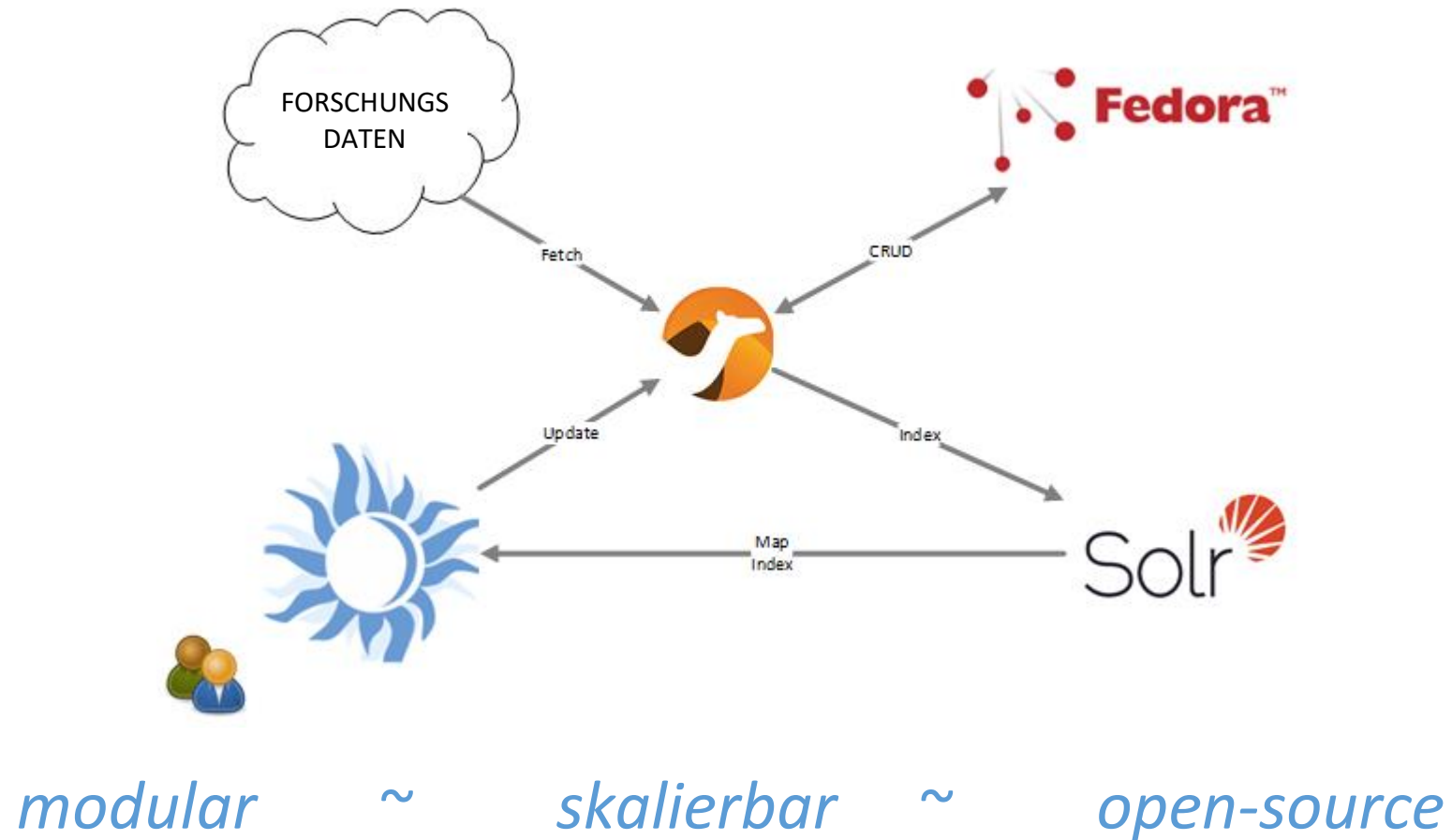


Discovery Service für Forschungsdaten

- Suche über verschiedene Projekte hinweg
- Großer Suchraum (Stemming)
- Intuitive Bedienbarkeit (z. B. nur einfache Suche)
- Ranking nach Relevanz
- Suchverfeinerung mit Facetten

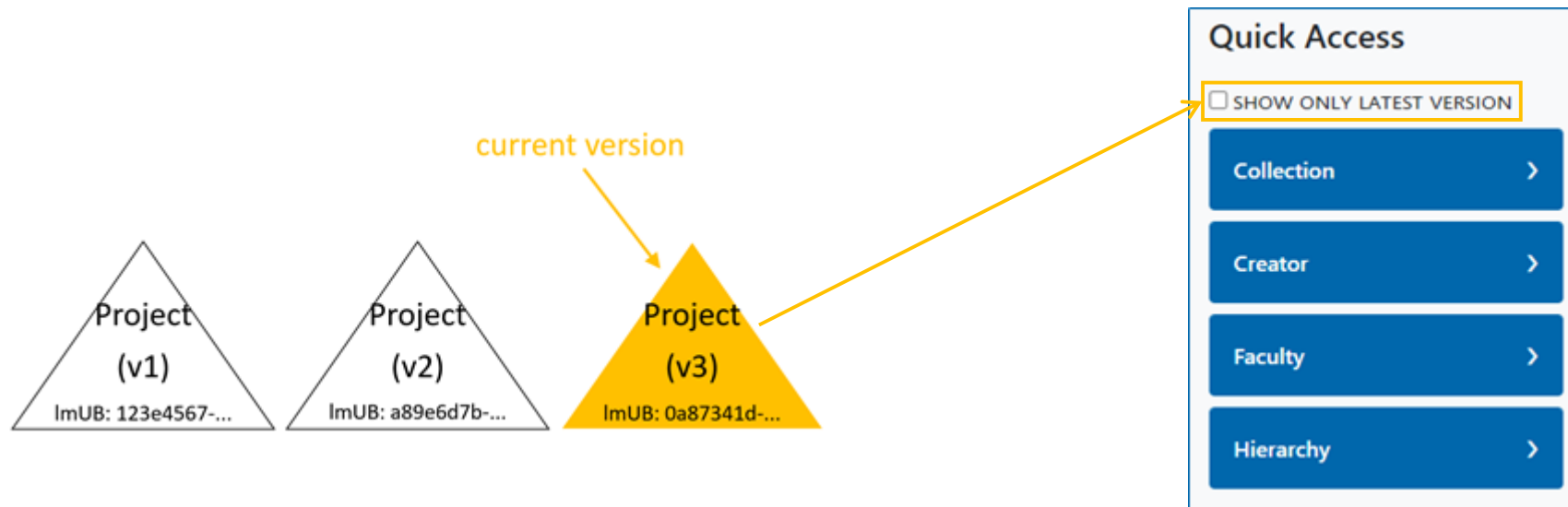
-
- Starke Berücksichtigung von Normdaten

Infrastruktur



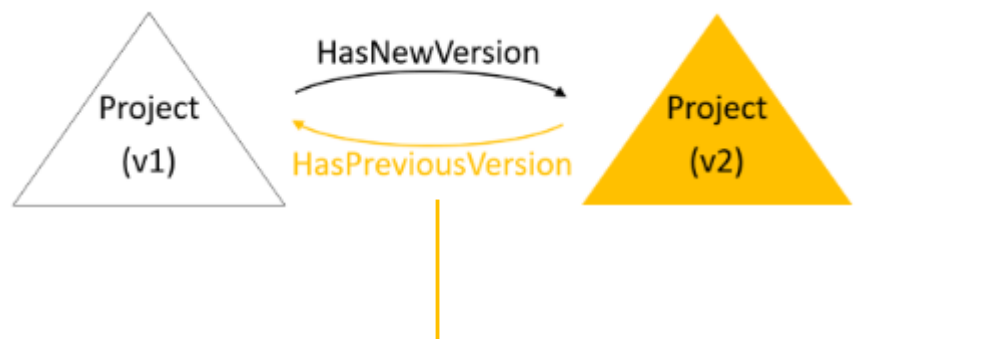
Versionierung

Möglichkeit der Sucheingrenzung auf die jeweils aktuellste Version eines Forschungsdatums



Versionierung - Pt. 2

Möglichkeit der Navigation von einer neueren Version zu einer älteren Version, und umgekehrt



VerbaAlpina instances for the morphological type "Teie(r (gem.)"

14 Previous Version

MAIN INFORMATION

Creator/Author: Krefeld, Thomas 
[Institut für Romanische Philologie, Ludwig-Maximilians-Universität Münche

Lücke, Stephan 
[IT-Gruppe Geisteswissenschaften (ITG), Ludwig-Maximilians-Universität Mün

Contributors: [show](#) 

Year of Publication: 2021

Version: 19/2

Funding: Deutsche Forschungsgemeinschaft (DFG): 253900505 

Faculty: Faculty for Languages and Literatures

RELATIONS

Is Part Of: [Q START SEARCH](#)

Has Part: [Q START SEARCH](#)

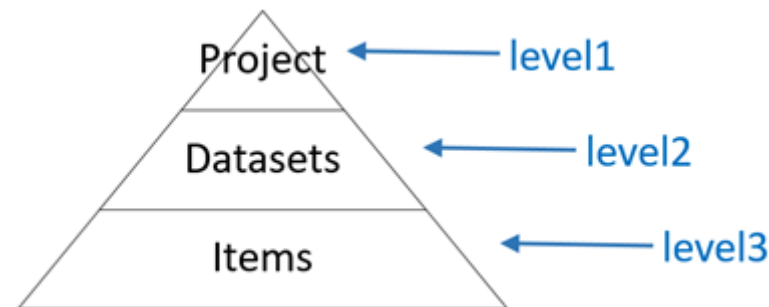
Is Identical To: [Find Record @ Verba Alpina](#) 

Is Variant Form Of: [Find Record @ Verba Alpina](#) 

Granularität

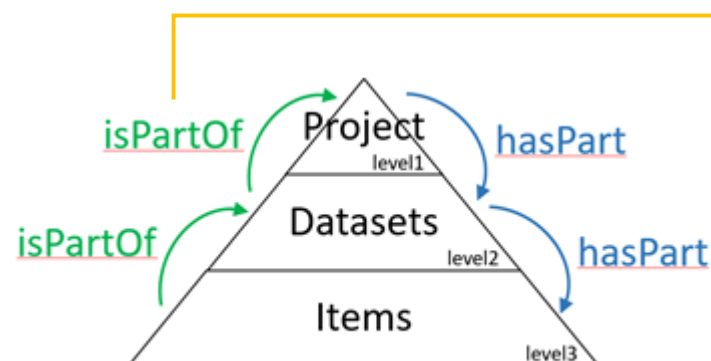
Was bedeutet das?

- Beschrieben werden nicht nur das Projekt als Ganzes, sondern auch darin enthaltene Datensätze
- *Ziel:*
Nicht nur das Projekt als Ganzes, sondern auch darin enthaltene Datensätze sollen referenzierbar und (separat) auffindbar sein.



Granularität - Pt. 2

Möglichkeit der Navigation von einem übergeordneten zu einem untergeordneten Forschungsdatum, und umgekehrt



VerbaAlpina instances for the morphological type "Teie(r (gem.)"

[Previous Version](#)

MAIN INFORMATION

Creator/Author: Krefeld, Thomas 
[Institut für Romanische Philologie, Ludwig-Maximilians-Universität Münche

Lücke, Stephan 
[IT-Gruppe Geisteswissenschaften (ITG), Ludwig-Maximilians-Universität MÜN

Contributors: [show](#) ▾

Year of Publication: 2021

Version: 19/2

Funding: Deutsche Forschungsgemeinschaft (DFG): 253900505 

Faculty: Faculty for Languages and Literatures

RELATIONS

Is Part Of: [START SEARCH](#)

Has Part: [START SEARCH](#)

Is Identical To: [Find Record @ Verba Alpina](#) 

Is Variant Form Of: [Find Record @ Verba Alpina](#) 

Verwendung von Normdaten

- Eindeutige Identifizierung von Personen, Orten und Stichwörtern
- Verlinkung über entsprechende Icons in der Einzeltrefferanzeige

GND

ORCID

GLOTTOLOG

WIKIDATA

GEONAMES

rdUB als internes Datenformat von Discover

- „DataCite-Anwendungsprofil“ - Kompromiss zwischen der Verwendung eines Standards und anwendungsspezifischer Erfordernisse
- Erweiterung von DataCite um Felder wie z. B. *hierarchy*, *currentVersion*, *collection*, *metadataUrl*, *checksum*, ...
- Struktur:

```

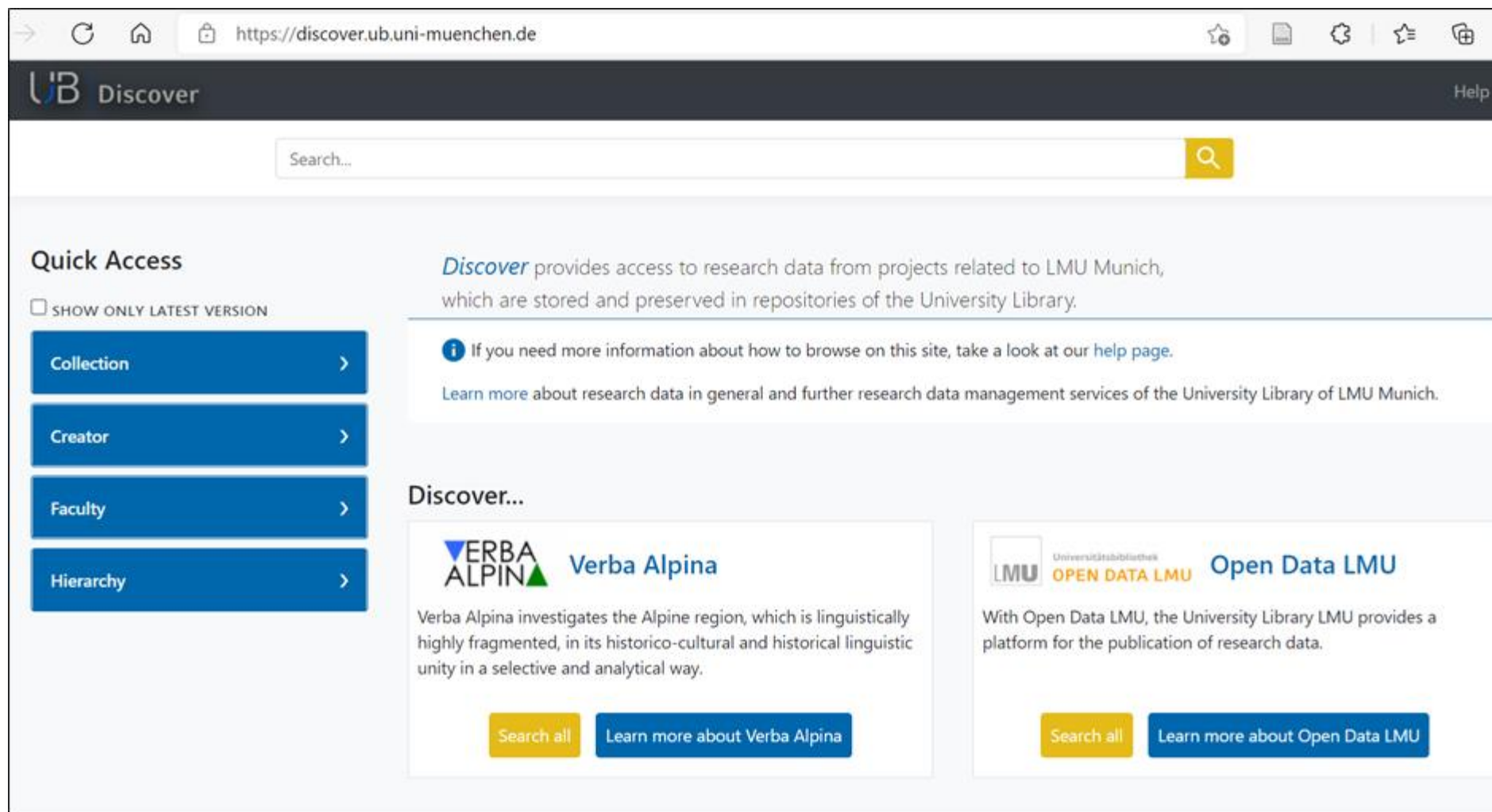
<lmUB:rData>
  <datacite:resource>
    <datacite:identifier>
    ...
  </datacite:resource>
</lmUB:rData>

```

} DataCite-Datensatz

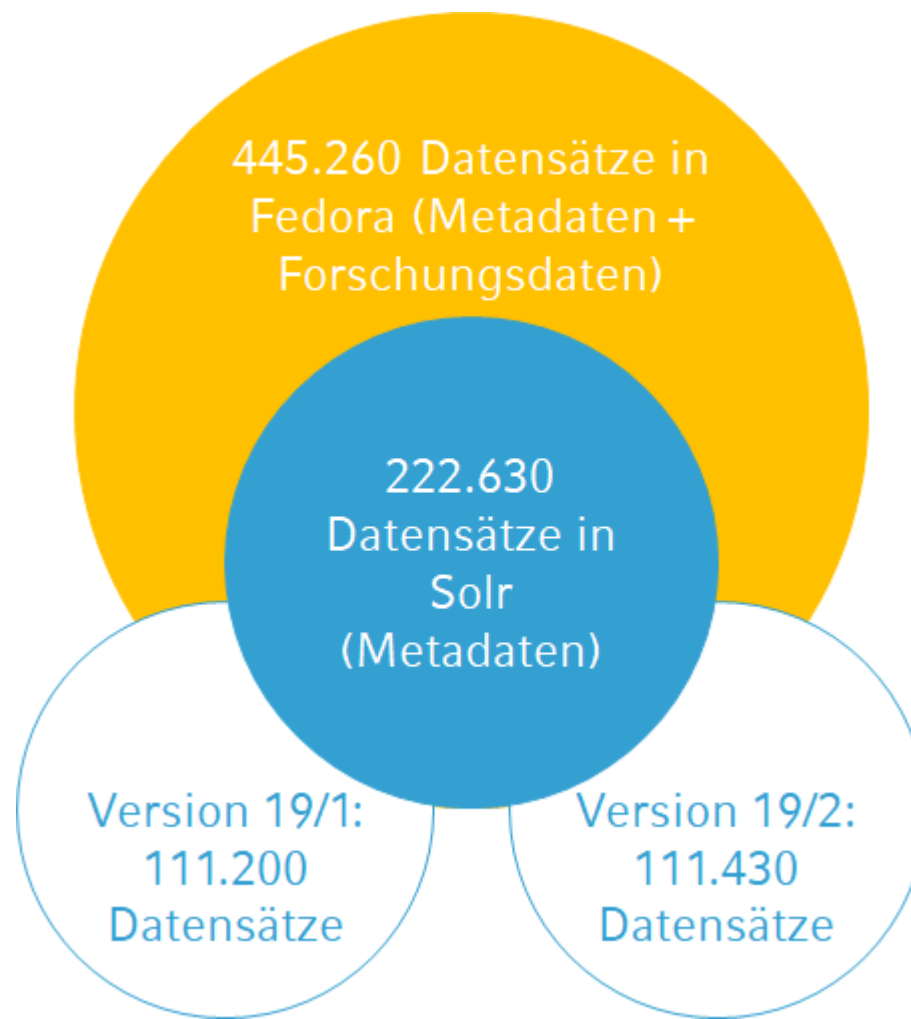
- Dokumentation: <https://github.com/UB-LMU/rdUB>

UB Discover - Live-Vorführung



The screenshot shows the UB Discover website interface. At the top, the browser address bar displays <https://discover.ub.uni-muenchen.de>. The website header features the 'UB Discover' logo and a 'Help' link. Below the header is a search bar with the placeholder text 'Search...'. On the left side, under the 'Quick Access' section, there is a checkbox labeled 'SHOW ONLY LATEST VERSION' and four blue buttons with white text: 'Collection', 'Creator', 'Faculty', and 'Hierarchy', each with a right-pointing arrow. The main content area on the right contains a paragraph explaining that 'Discover' provides access to research data from projects related to LMU Munich, which are stored and preserved in repositories of the University Library. Below this, there is an information icon and a link to the 'help page'. Further down, there is a section titled 'Discover...' with two featured projects. The first project is 'Verba Alpina', which investigates the Alpine region, described as linguistically highly fragmented, in its historico-cultural and historical linguistic unity in a selective and analytical way. It includes a yellow 'Search all' button and a blue 'Learn more about Verba Alpina' button. The second project is 'Open Data LMU', which provides a platform for the publication of research data. It also includes a yellow 'Search all' button and a blue 'Learn more about Open Data LMU' button.

Das FDM-Projekt in Zahlen



Vielen Dank für Ihre Aufmerksamkeit!

Web: www.fdm-bayern.org

Twitter: @fdm_bayern

Kontakt: forschungsdaten@ub.uni-muenchen.de