



GEORG-AUGUST-UNIVERSITÄT  
GÖTTINGEN



## GOEDOC - Dokumenten- und Publikationsserver der Georg-August-Universität Göttingen

---

---

2021

### Nachnutzung Git-basierter Sammlungen im Rahmen der Infrastrukturdienste von CLARIAH-DE

Tobias Gradl, Robin Jegan

DARIAH-DE Working Papers

Nr. 42

Gradl, T.; Jegan, R: Nachnutzung Git-basierter Sammlungen im Rahmen der Infrastrukturdienste von CLARIAH-DE

Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2021  
(DARIAH-DE working papers 42)

Verfügbar:

PURL: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2021-2>

URN: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2021-2-5>



## Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

*Erschienen in der Reihe*  
DARIAH-DE working papers

ISSN: 2198-4670

*Herausgeber der Reihe*  
DARIAH-DE, Niedersächsische Staats- und Universitätsbibliothek

Mirjam Blümm, Thomas Kollatz, Stefan Schmunk und Christof Schöch

---

---

**Abstract:** Repositorien auf der Basis von Git werden in zunehmendem Maße auch als Speicher von Forschungsdaten verwendet - auch weil sie Funktionalität wie Versionierung und Kollaboration bereits auf Protokollebene berücksichtigen und Anwenderinnen und Anwender durch ansprechende und funktionale Weboberflächen unterstützen. Über standardisierte Schnittstellen ist zudem ein problemloser Zugriff auf die Inhalte der Repositorien sichergestellt.

Die vorliegende Anleitung zeigt, wie Git-Repositorien mit Hilfe der Infrastrukturdienste von CLARIAH-DE dokumentiert, modelliert und in eine übergreifende Suche eingebunden werden können. Das gewählte Beispiel zur Integration der Lehrmaterialien von DARIAH Campus in den CLARIAH-DE Tutorial Finder ist dabei übertragbar auf eine Vielzahl von Anwendungsfällen in den Digital Humanities. Die beschriebenen Dienste der Collection Registry, des Data Modeling Environment und der Generic Search können neben den infrastrukturell betriebenen Instanzen auch für die Umsetzung eigener Integrations- und Suchlösungen installiert und betrieben werden.

**Keywords:** Digitale Forschungsinfrastrukturen, Forschungsdatenmanagement, Datenföderationsarchitektur, Anleitung

digital research infrastructures, research data management, data federation architecture, guide

# Nachnutzung Git-basierter Sammlungen im Rahmen der Infrastrukturdienste von CLARIAH-DE

Tobias Gradl

Robin Jegan

Otto-Friedrich-Universität Bamberg



Tobias Gradl, Robin Jegan: „Nachnutzung Git-basierter Sammlungen im Rahmen der Infrastrukturdienste von CLARIAH-DE“. *DARIAH-DE Working Papers* Nr. 42. Göttingen: DARIAH-DE, 2021.

URN: [urn:nbn:de:gbv:7-dariah-2021-2-5](https://nbn-resolving.org/urn:nbn:de:gbv:7-dariah-2021-2-5).

Dieser Beitrag erscheint unter der  
Lizenz [Creative-Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/) (CC-BY).

Die *DARIAH-DE Working Papers* werden von Mirjam Blümm,  
Thomas Kollatz, Stefan Schmunk und Christof Schöch  
herausgegeben.



## Zusammenfassung

Repositorien auf der Basis von Git werden in zunehmendem Maße auch als Speicher von Forschungsdaten verwendet – auch weil sie Funktionalität wie Versionierung und Kollaboration bereits auf Protokollebene berücksichtigen und Anwenderinnen und Anwender durch ansprechende und funktionale Weboberflächen unterstützen. Über standardisierte Schnittstellen ist zudem ein problemloser Zugriff auf die Inhalte der Repositorien sichergestellt.

Die vorliegende Anleitung zeigt, wie Git-Repositorien mit Hilfe der Infrastrukturdienste von CLARIAH-DE dokumentiert, modelliert und in eine übergreifende Suche eingebunden werden können. Das gewählte Beispiel zur Integration der Lehrmaterialien von DARIAH Campus in den CLARIAH-DE Tutorial Finder ist dabei übertragbar auf eine Vielzahl von Anwendungsfällen in den Digital Humanities. Die beschriebenen Dienste der Collection Registry, des Data Modeling Environment und der Generic Search können neben den infrastrukturell betriebenen Instanzen auch für die Umsetzung eigener Integrations- und Suchlösungen installiert und betrieben werden.

## Schlagwörter

Digitale Forschungsinfrastrukturen, Forschungsdatenmanagement, Datenföderationsarchitektur, Anleitung

## Keywords

digital research infrastructures, research data management, data federation architecture, guide

## Inhaltsverzeichnis

<b>Einleitung</b>	<b>4</b>
<b>Überblick</b>	<b>6</b>
<b>Schritt 1: Analyse von Sammlung, Inhalten und Kontext</b>	<b>7</b>
<b>Schritt 2: Beschreibung der Sammlung</b>	<b>8</b>
<b>Schritt 3: Modellierung der Inhalte</b>	<b>11</b>
Elementmodell . . . . .	11
Aufbereitung von Titeln zu URL-Bestandteilen . . . . .	12
Strukturierung von Inhaltsbereichen . . . . .	15
Volltext extrahieren . . . . .	18
Verfeinerung und Abschluss des Modells . . . . .	19
URL zur Webseite der Ressource . . . . .	19
Youtube-Daten anreichern . . . . .	20
Modellierungsergebnis . . . . .	22
<b>Schritt 4: Herstellung von Mappings</b>	<b>23</b>
<b>Schritt 5: Integration von Daten</b>	<b>26</b>
Sammlungsverwaltung in der Suche . . . . .	27
Datenmodell- und Indexverwaltung in der Suche . . . . .	28
Auffindbarkeit von Ressourcen . . . . .	28
<b>Zusammenfassung</b>	<b>29</b>
<b>Anhang</b>	<b>30</b>
In XML exportierte Beispieldaten im Quellformat . . . . .	30
In XML exportierte Beispieldaten im Zielformat . . . . .	33

## Einleitung

Im Rahmen von CLARIAH-DE<sup>1</sup> entsteht mit dem CLARIAH-DE Tutorial Finder<sup>2</sup> eine Suche über geisteswissenschaftliche Lehrmaterialien, die aus verschiedenen, heterogenen Sammlungen bezogen werden. Wesentliche Ziele der Suchlösung bestehen darin, digital arbeitenden Geisteswissenschaftlerinnen und Geisteswissenschaftlern einen integrativen Zugang zu relevanten Schulungsmaterialien bereitzustellen, und im Zuge dessen auch die Sichtbarkeit und Nachnutzbarkeit angebundener Sammlungen zu erhöhen. Technisch wird der Tutorial Finder durch eine spezifische Installation der DARIAH-DE Generic Search<sup>3</sup> realisiert.

Eine erste Bestandsaufnahme in CLARIAH-DE führte zu einer Zusammenstellung von 14 zu integrierenden Lehrmaterialsammlungen<sup>4</sup>. Dabei ist zu beachten, dass diese erste Auswahl nicht allein anhand der thematischen Relevanz und Bedeutung der Sammlungen für die zu entwickelnde Suchlösung erfolgte. Um die Anwendbarkeit der Generic Search als Lehrmaterialsuche einschätzen zu können wurde insbesondere auf die Heterogenität mit Bezug auf Medientypen, Formate und technische Zugangsformen geachtet. Neben dedizierten *Portalen* und *Internetseiten*, wie Teaching and Learning Materials Collection (TeLeMaCo)<sup>5</sup> umfassen die Einträge daher auch *Videosammlungen*, wie diejenige von forTEXT<sup>6</sup>. Mit Beispielen wie DARIAH-Campus<sup>7</sup> und The Programming Historian<sup>8</sup> finden sich in der derzeitigen Zusammenstellung auch Sammlungen, die für die Verwaltung ihrer Inhalte und Metadaten auf die Nutzung von *Git-Repositories* setzen. Bei Anbietern wie GitLab und GitHub oder auch in organisationseigenen Installationen<sup>9</sup> entsprechend konfiguriert, bieten diese eine unmittelbar zugängliche Schnittstelle für die Veröffentlichung und Nachnutzung von Inhalten.

Die vorliegende Anleitung dokumentiert die Anbindung eines solchen Git-Repositories an den CLARIAH-DE Tutorial Finder am Beispiel von DARIAH-Campus. Die in dieser Anleitung beschriebenen Integrationsschritte dienen dabei als Referenz für die Anbindung weiterer Git-Repositories an beliebige Installationen der DARIAH-DE Generic Search, wobei Sammlungsbeschreibung und Datenmodell inhaltlich jeweils spezifisch zu entwickeln sind.

Die Lehrmaterialsammlung von DARIAH-Campus wurde als Anwendungsbeispiel für diese Anleitung im Wesentlichen aus zwei Gründen ausgewählt:

1. Ressourcen liegen als Inhalte mit begleitenden Metadaten in einem öffentlich zugänglichen Git-Repository vor. Die Konfiguration eines Zugangs via Git weist im Vergleich zu Online-Dateien und OAI-PMH<sup>10</sup> die derzeit detailliertesten Konfigurationsoptionen in der Collection Registry (CR) auf.

---

<sup>1</sup><https://www.clariah.de>

<sup>2</sup><https://teaching.clariah.de>

<sup>3</sup><https://de.dariah.eu/web/guest/generische-suche>

<sup>4</sup><https://teaching.clariah.de/colreg-ui/collections/>

<sup>5</sup><https://telemaco.clarin-d.uni-saarland.de/hub/browse/>

<sup>6</sup>[https://www.youtube.com/channel/UC4aOD2eus\\_b\\_q66LD-Q0km9Q](https://www.youtube.com/channel/UC4aOD2eus_b_q66LD-Q0km9Q)

<sup>7</sup><https://github.com/DARIAH-ERIC/dariah-campus>

<sup>8</sup><https://github.com/programminghistorian/jekyll>

<sup>9</sup>bspw. die Installation der Otto-Friedrich-Universität Bamberg: <https://gitlab.rz.uni-bamberg.de>

<sup>10</sup><http://www.openarchives.org/OAI/openarchivesprotocol.html>

2. Ressourcen liegen als *unstrukturierte Textdateien* vor. Vorhandene Annotationen im *Markdown-Format* können im Rahmen der Modellierung verwendet werden, um strukturierte Daten zu gewinnen. Die Modellierungsschritte basieren auf der Anwendung von Grammatiken und Transformationsregeln und verdeutlichen die Mächtigkeit des *Data Modeling Environments* (DME) mit Bezug auf die Modellierung.

Die dargestellten Schritte basieren auf komplexen Rahmenbedingungen und sollen die jeweiligen Möglichkeiten von CR und DME andeuten. Für eine Anbindung strukturierter Daten sind vergleichbare Modellierungsaufwände deutlich geringer und im Bereich der Datenmodellierung oft auf den Import eines XML-Schemas<sup>11</sup> und die Herstellung einfacher Mappings beschränkt.

---

<sup>11</sup><http://www.edition-w3.de/TR/2001/REC-xmlschema-0-20010502/>

## Überblick

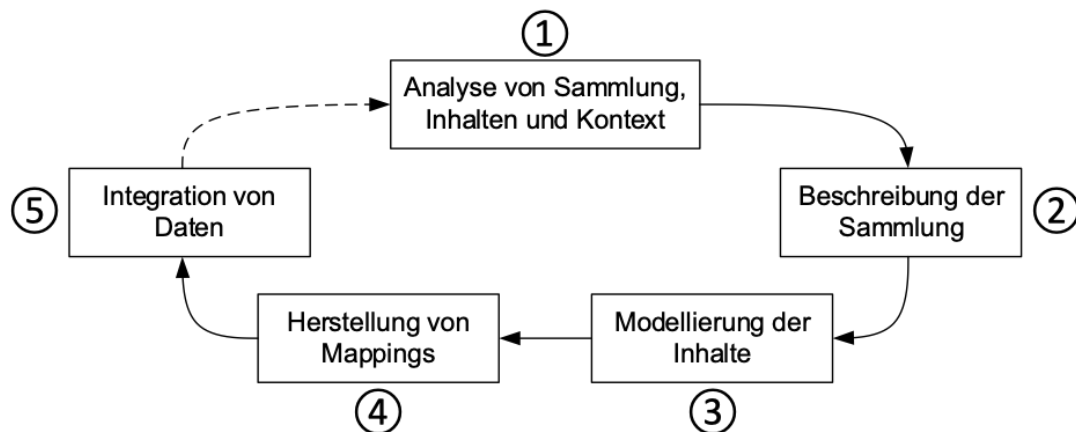


Abbildung 1: Sammlungs- und Datenintegration im Überblick

Die Integration einer Sammlung und ihrer Daten folgen einem strukturierten Prozess (vgl. Abbildung 1). Zunächst sind *Informationen über die Sammlung und die beinhalteten Ressourcen* einzuholen und zu bewerten ①. Kenntnisse zur *gewünschten Nachnutzung* der Daten sind ebenso von Bedeutung, da der jeweilige Verwendungskontext oft bereits die Modellierung der Daten, vor allem aber deren anschließendes Mapping beeinflusst. Mit Hilfe der CR werden *Sammlungen erfasst, beschrieben und insbesondere die Zugriffsmöglichkeit* (hier Git) *spezifiziert* ②.

Im DME werden die *Sammlungsinhalte modelliert* ③, wobei neben der schematischen Abbildung der Daten insbesondere Regeln zu deren Anreicherung und Bereinigung modelliert werden können. Ebenso wird durch *Mappings in der DME Interoperabilität der Daten* ④ hergestellt, welche für eine *Integration* beispielsweise im Rahmen von Suchdiensten Anwendung findet. Die Anbindung von Datenbeständen führt zu einer *übergreifenden Analysierbarkeit, Durchsuchbarkeit und Interpretierbarkeit von Inhalten* ⑤ – im Kontext dieser Anleitung im Rahmen des CLARIAH-DE Tutorial Finders. Beschreibungs- und Modellierungsergebnisse können schließlich bei Bedarf iterativ verfeinert werden.

Mit der vorliegenden Anleitung verfolgt CLARIAH-DE die Absicht, diese Art und Weise der Einbindung von externen Ressourcen in eine eigene Suche als Best Practice vorzustellen und zur Umsetzung in anderen Kontexten zu ermutigen. Die Bereitstellung von Ressourcen über eine standardisierte Schnittstelle wie Git wird dabei selbst bereits als Best Practice empfohlen. Sämtliche der verwendeten Komponenten können, wie im Beispiel des Tutorial Finders, direkt verwendet werden. Sie stehen jedoch auch für eine individuelle Installation bspw. zur Umsetzung eigener Suchlösungen zur Verfügung.<sup>12</sup>

Die Integration der Ressourcen in den CLARIAH-DE Tutorial Finder wird im Folgenden anhand der Sammlung und Inhalte von *DARIAH-Campus* beschrieben. Sämtliche Teilschritte können auch in den jeweiligen Diensten nachvollzogen werden.

<sup>12</sup>Vgl. <https://dfa.de.dariah.eu/doc/dfa/>



## Schritt 1: Analyse von Sammlung, Inhalten und Kontext

Für die Integration von Daten in die Infrastruktur von CLARIAH-DE mit Hilfe von CR und DME sind die Begriffe der Ressource und der Sammlung relevant und werden im vorliegenden Dokument wie folgt verwendet. Eine *Ressource* bildet im vorliegenden Kontext ein individuell beschreibbares Artefakt, welches als Material für Lehrende, Studierende und Forschende in den Digital Humanities relevant sein kann, bspw. ein Tutorial, ein Artikel oder eine Prozessdokumentation. Ressourcen werden durch den Tutorial Finder auffindbar gemacht, wobei an dieser Stelle nicht wesentlich zwischen der *primären* Ressource, also den Lehrmaterialien selbst und *sekundären* Ansichten auf diese unterschieden wird: Abbildung 2 zeigt die Darstellung einer Ressource – hier ein Artikel – in DARIAH-Campus<sup>13</sup>. Diese browser- und damit nutzerorientierte Sicht basiert auf einer in GitHub (vgl. Abbildung 3) hinterlegten Datei im Markdown<sup>14</sup> Format. In dieser findet sich neben vorangestellten Metadaten insbesondere auch der vollständige Artikel. Das primäre Material findet sich somit unmittelbar und vollständig im Git-Repository wieder. Daneben bildet eine Vielzahl von Ressourcen in DARIAH-Campus auch externe Ressourcen ab, bspw. YouTube-Videos<sup>15</sup> oder Schulungsmaterialien<sup>16</sup>. In diesen Fällen befindet sich die primäre Ressource nicht innerhalb des Git-Repository, ist jedoch ebenso über den Tutorial Finder auffindbar.

The screenshot shows the DARIAH-Campus website interface. At the top, there is a navigation bar with the DARIAH-CAMPUS logo and links for Resources, Topics, Sources, Course Registry, and About. A search icon and a 'Contact' button are on the right. The main content area features the title 'Controlled Vocabularies and SKOS' in a large, bold font. Below the title, the author's name 'Ksenia Zaytseva' and the date 'Mar 23, 2020' are displayed. To the right of the author information, the source is listed as 'DARIAH' and the topics as 'SKOS, Controlled Vocabularies, Semantic Web, Information Architecture, Metadata'. On the left side of the article, there is a 'CITE AS' section with the full citation and a 'Copy citation' button, followed by a 'REUSE CONDITIONS' section. The main body of the article starts with an 'Introduction' and a section 'What are Controlled Vocabularies?'. The text explains that controlled vocabularies are knowledge organization systems used for organizing and classifying data. A quote from Patricia Harpring (2010) is included, defining a controlled vocabulary. A 'TABLE OF CONTENTS' sidebar on the right lists the article's sections: Introduction, What are Controlled Vocabularies?, Role of CVs in Semantic web, CVs in a Digital Humanities context, SKOS Standard, About SKOS, SKOS Basics, Concept, Labels, Semantic relationships, Documenting concepts, Concept Scheme, Grouping concepts within one vocabulary, Advanced SKOS, Mapping concepts to external vocabularies and Linked Open Data (LOD) resources.

Abbildung 2: Browserdarstellung einer Ressource in DARIAH-Campus

<sup>13</sup><https://campus.dariah.eu/resource/controlled-vocabularies-and-skos>

<sup>14</sup><https://github.github.com/gfm/>

<sup>15</sup><https://campus.dariah.eu/resource/cultural-big-data-building-a-european-internet-of-cultural-things> mit Verweis auf <https://www.youtube.com/watch?v=6Vjhm6SJhpo>

<sup>16</sup><https://campus.dariah.eu/resource/game-play-design-in-the-arts-and-humanities> mit Verweis auf <https://teach.dariah.eu/course/view.php?id=57>

DARIAH-Campus ist damit ein Beispiel für eine *Sammlung* im vorliegenden Kontext, indem diese sowohl eine inhaltliche, als auch eine technische (das Git-Repository) Klammer um einzelne Ressourcen abbildet.<sup>17</sup>

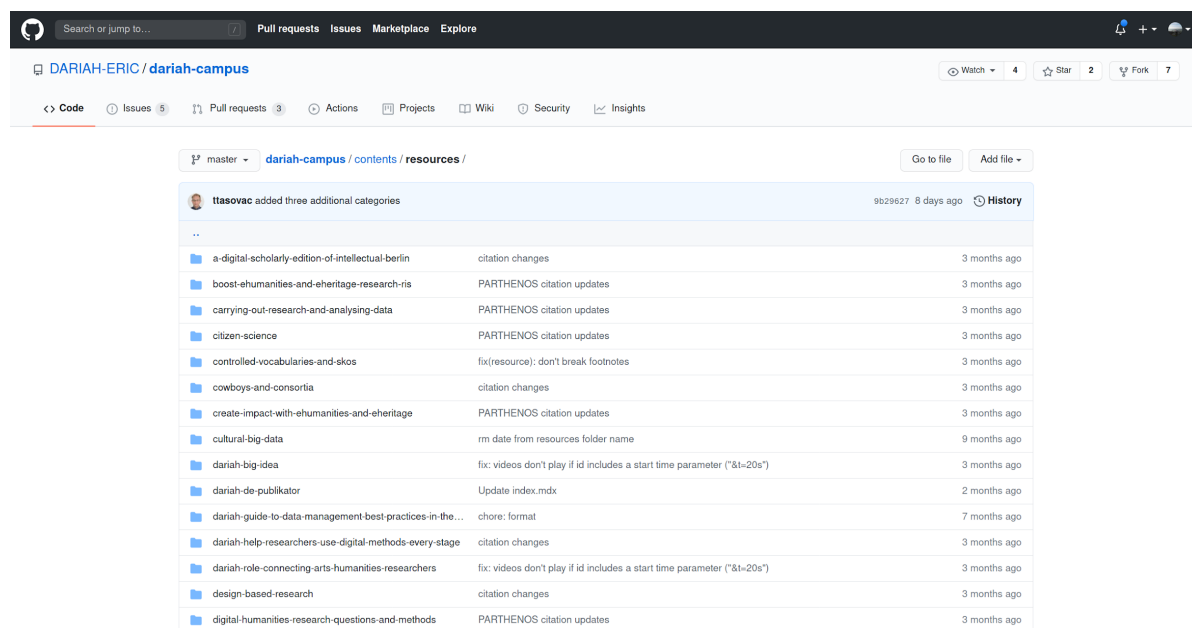


Abbildung 3: Repository von DARIAH-Campus auf GitHub

## Schritt 2: Beschreibung der Sammlung

Das Ziel der Integration der DARIAH-Campus Materialien besteht darin, diese in den CLARIAH-DE Tutorial Finder einzubinden und damit im Rahmen einer übergreifenden Suche zugänglich zu machen.

Um die Zugänglichkeit und Nutzbarkeit einer Sammlung und der beinhalteten Daten herzustellen, ist ein entsprechender *Eintrag in der CR anzulegen*<sup>18</sup>. Eine solche Eintragung kann dabei grundsätzlich durch authentifizierte<sup>19</sup> Nutzerinnen und Nutzer jederzeit selbst vorgenommen werden. In der Praxis hat sich gezeigt, dass eine korrekte und vollständige Sammlungsbeschreibung zumeist durch Sammlungsverantwortliche selbst oder in enger Abstimmung mit diesen entwickelt wird. Da der Tutorial Finder als technisch unabhängige Anwendung realisiert wird, unterscheidet sich die Instanz der CR von der generischen DARIAH-DE Instanz der CR<sup>20</sup>.

<sup>17</sup><https://github.com/DARIAH-ERIC/dariah-campus/tree/master/contents/resources>

<sup>18</sup><https://teaching.clariah.de/colreg-ui/collections/5eba5f516117b1100fe31678>

<sup>19</sup>Mitglieder über die DFN-AAI angeschlossener Einrichtungen, vgl. <https://www.aai.dfn.de/>

<sup>20</sup><https://colreg.de.dariah.eu/>

Die für die Beschreibung einer Sammlung zur Verfügung stehenden Metadaten-Felder sind im Rahmen des DARIAH Collection Description Data Model (DCDDM)<sup>21</sup> definiert. Wenngleich nur wenige Eingabefelder verpflichtend sind, so dient eine möglichst umfassende und vollständige Beschreibung der Auffindbarkeit, Nutzbarkeit und Dokumentation der Sammlungen.

Abbildung 4: Sammlungseditor der CR

Die Beschreibung von Sammlungen erfolgt im Sammlungseditor. Authentifizierte Nutzerinnen und Nutzer legen diese zunächst als so genannten Sammlungsentwurf an, der z. B. in Zusammenarbeit zwischen Sammlungsverantwortlichen, Domänenexpertinnen und Domänenexperten iterativ und kollaborativ entwickelt und verfeinert werden kann. Durch die explizite Veröffentlichung der Sammlungsbeschreibung wird diese öffentlich sichtbar und Daten können durch angeschlossene Dienste, wie den CLARIAH-DE Tutorial Finder genutzt werden

Sammlungsbeschreibungen dienen in erster Linie zunächst der Dokumentation von Sammlungen. Soll über diesen Zweck hinaus eine Nutzbarkeit der Inhalte einer Sammlung ermöglicht werden, ist zusätzlich zu den ohnehin obligatorischen Pflichtfeldern die Angabe einer unterstützten Schnittstelle erforderlich. Erst durch die Bereitstellung von Informationen über Möglichkeiten des Zugriffs auf beinhaltete Daten können die an die CR *angeschlossenen Dienste*, wie der CLARIAH-DE Tutorial Finder auf Daten zugreifen, diese verarbeiten und bereitstellen.

Die Beschreibung eines *Zugriffs auf Basis von Git* wird in Abbildung 5 am Beispiel von DARIAH-Campus dargestellt. Für die Spezifikation werden verschiedene Felder angeboten, von denen *URI*, *der Dateityp*, sowie *anzuwendende Zugriffsmuster erforderlich sind*.

<sup>21</sup><https://github.com/DARIAH-DE/DCDDM>

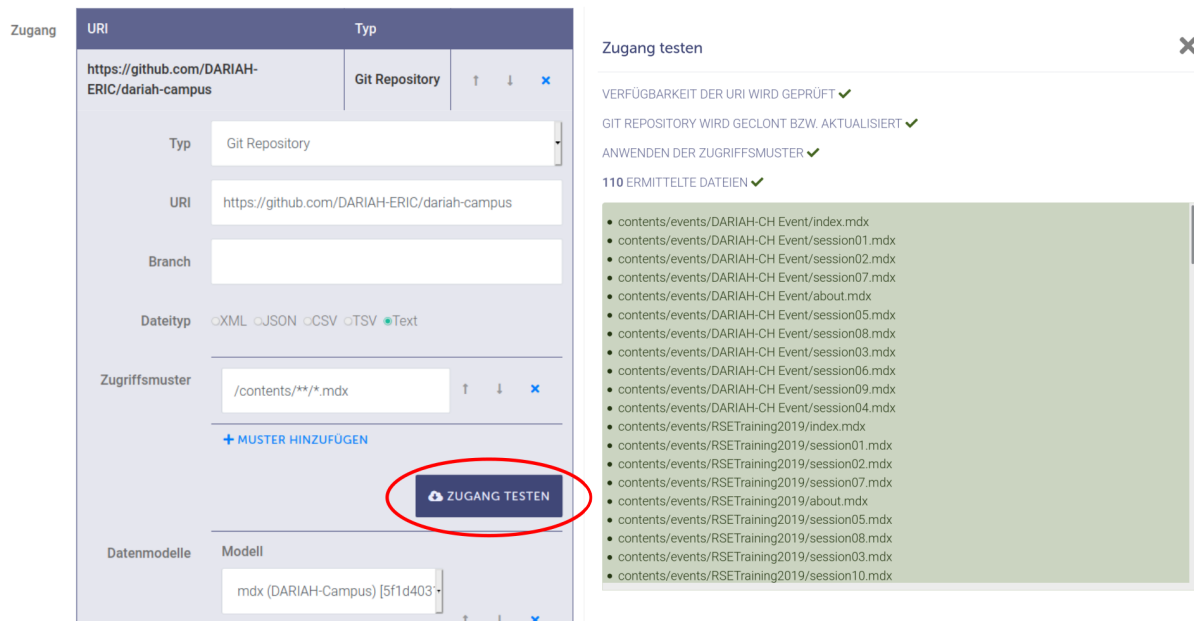


Abbildung 5: Browserdarstellung einer Ressource in DARIAH-Campus

- *URI*: die über das Internet öffentlich zugängliche Adresse des Git-Repositorium. Es werden derzeit ausschließlich Git-Repositorium unterstützt, die einen öffentlichen, lesenden Zugriff ohne Authentifizierung erlauben.
- *Branch*: erlaubt die Spezifikation des zu verwendenden Branches oder Tags; ohne Angabe wird der Standard-Branch des jeweiligen Repositoriums verwendet.
- *Dateityp*: die Angabe des Dateityps ist wichtig für nachgelagerte Dienste, die auf Basis dieser Angabe und der Spezifikation des Datenmodells Daten verarbeiten.
- *Zugriffsmuster* als Glob-Pfadnamen<sup>22</sup>; zumeist sind nicht alle in einem Repositorium enthaltenen Dateien auch als Sammlungsdaten zu verstehen; die notwendige Angabe von Zugriffsmustern erlaubt eine entsprechende Filterung von Dateien.
- *Datenmodell*: Verweis auf das Datenmodell in der DME; im Beispiel wurde das Markdown-Format von DARIAH-Campus individuell modelliert (vgl. Abschnitt Datenmodellierung), worauf an dieser Stelle verwiesen wird.

Mit Hilfe von Zugriffsmustern werden die in dem beschriebenen Git-Repositorium enthaltenen Dateien gefiltert und definiert.

Können die in einem Git-Repositorium enthaltenen und relevanten Dateien nicht mit Hilfe einer einzelnen Kombination aus Zugriffsmuster, Dateityp und Datenmodell beschrieben werden, so können *beliebig viele Zugänge angelegt* werden, mit denen unterschiedliche *Gruppen von Dateien* beschrieben werden. Die *Inhalte der Sammlung ergeben sich dann aus der Kombination aller definierten Zugänge*.

<sup>22</sup>siehe z. B. [https://en.wikipedia.org/wiki/Glob/\\_\(programming\)](https://en.wikipedia.org/wiki/Glob/_(programming))

Beispiel:

Zugang 1: Dateityp [XML], Zugriffsmuster [mods\_\*.xml], Datenmodell [MODS]

Zugang 2: Dateityp [XML], Zugriffsmuster [dc\_\*.xml], Datenmodell [DublinCore]

### Schritt 3: Modellierung der Inhalte

Mit Hilfe eines im DME beschriebenen Datenmodells werden angeschlossene Dienste in die Lage versetzt, *Sammlungsinhalte zu verarbeiten* und nach den *im Modell spezifizierten Regeln aufzubereiten*. Der Schritt der Datenmodellierung ist für die weitere Nutzung von Daten immer dann erforderlich, wenn nicht bereits ein passendes Modell im DME vorliegt. In jedem Fall sollte die Funktionalität eines bereits vorhandenen oder spezifisch erstellten Datenmodells durch die Anwendung von Beispieldaten geprüft werden.

#### Elementmodell

Die Farben der Knoten im Elementmodell (vgl. Abbildung 6) zeigen den jeweiligen Typ der Elemente an:

- **Blaue Elemente** bilden die *Parser-Sicht* ab und repräsentieren die Struktur der vorliegenden Daten, beispielsweise XML-Elementen oder Spalten einer CSV-Datei.
- **Goldene Knoten** unmittelbar unter blauen oder violetten Knoten beinhalten *Grammatiken zur Definition von Elementwerten*. Durch deren Anwendung werden implizite strukturelle Gegebenheiten expliziert und verfügbar gemacht. Die Spezifikation von Transformationsregeln wird dadurch vereinfacht oder oft erst ermöglicht. Für unzerlegbare (atomare) Elementwerte leitet eine so genannte *Passthrough-Grammatik* Werte direkt an Transformationsregeln weiter.
- **Goldene Knoten** unter Grammatiken beinhalten *Transformationsregeln* beispielsweise zur Zuweisung von Daten an untergeordnete Labels.
- **Violette Knoten** stellen als Labels *produzierte Elemente* dar, die nicht als solche in ursprünglichen Daten auftauchen, sondern durch die Anwendung von Grammatiken und Transformationsregeln erzeugt und befüllt werden.

Der Bildschirmausschnitt in Abbildung 6 stellt eine Ansicht auf den Datenmodell-Editor des DME dar. Zum Elementmodell von DARIAH-Campus<sup>23</sup> passende Beispieldaten wurden zur Veranschaulichung im Eingabebereich links eingefügt. Durch Anwendung des Datenmodells auf diese Daten produzierte Ergebnisse werden im rechten Bereich angezeigt.<sup>24</sup>

<sup>23</sup><https://teaching.clariah.de/dme/model/editor/5f1d40316117b104a5fd3944/>

<sup>24</sup>Die Standardanordnung der Bereiche unterscheidet sich von der des Bildschirmausschnittes. Die Bereiche können durch Drag-und-Drop der Tabs frei angeordnet werden.

Dadurch dass die Ressourcen in DARIAH-Campus zunächst als unstrukturierte Text-Dateien vorliegen, entsteht das Elementmodell aus nur einem Knoten **MDXfile**, der beliebigen Text aufnimmt. Durch die Anwendung von Grammatiken (z. B. **ParseText**) und Transformationsregeln (z. B. **ExtractMetadata**) werden strukturelle Regeln der Daten expliziert und semi-strukturierte Daten gewonnen.

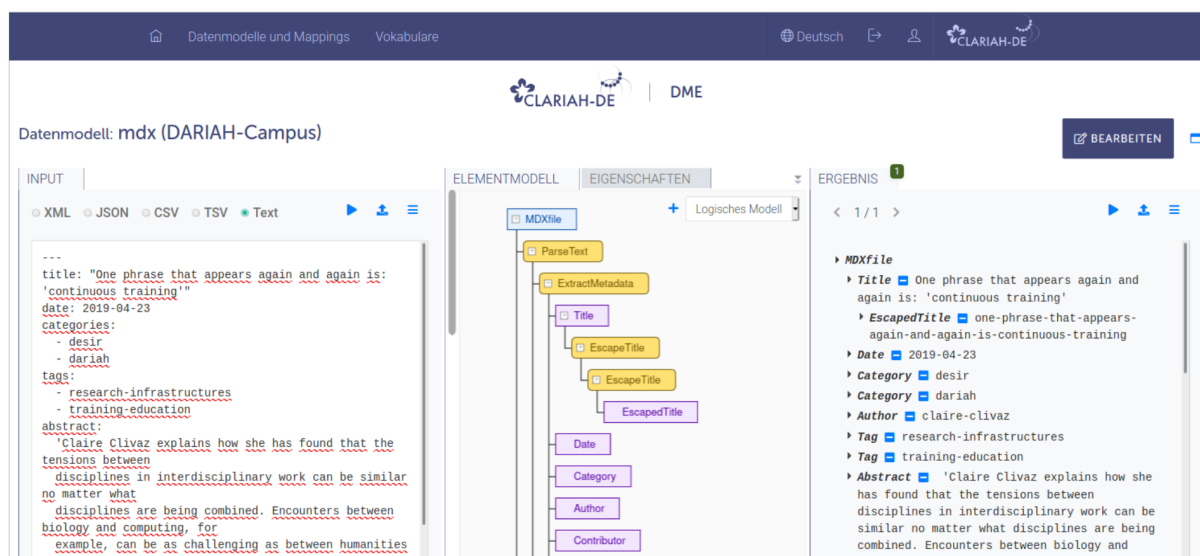


Abbildung 6: Datenmodell von DARIAH-Campus im DME

Die zur Modellierung der DARIAH-Campus Daten entwickelten Grammatiken **ParseText**, **ExtractFulltext** und **EscapeTitle** finden sich als Sammlung im DARIAH-DE Repository<sup>25</sup>.

### Aufbereitung von Titeln zu URL-Bestandteilen

Die Funktionalität zur Bereinigung von Titeln wird an dieser Stelle vorangestellt, da die hierfür erforderliche sprachliche Definition von **EscapeTitle**<sup>26</sup> die einfachste der drei Grammatiken darstellt. Die im Modellierungsablauf zuvor erforderliche Extraktion inhaltlicher Blöcke durch die **ParseText** Grammatik ist deutlich komplexer und wird im Rahmen des nächsten Abschnittes adressiert.

Die weitere Verarbeitung des Titels hat den Hintergrund, dass auf deren Basis *die URL zu der repräsentativen Webseite der Ressource* auf DARIAH-Campus zu bilden ist. Diese URL ist erforderlich, um einen Verweis von entsprechenden Suchergebnissen des Tutorial Finders auf die Ressource in ihrem Sammlungskontext anbieten zu können.

<sup>25</sup><https://dx.doi.org/10.20375/0000-000C-E12B-6>

<sup>26</sup><https://doi.org/10.20375/0000-000C-E12C-5>

Der Titel eines DARIAH-Campus Beispieleintrags:

*„One phrase that appears again and again is: ,continuous training“*

ist zu transformieren in einen URL-Suffix:

*one-phrase-that-appears-again-and-again-is-continuous-training*

um abschließend die folgende URL generieren zu können:

<https://campus.dariah.eu/resource/one-phrase-that-appears-again-and-again-is-continuous-training>

```
/*
* Mit Hilfe dieser Grammatik werden die Titel der DARIAH-Campus
* Ressourcen verarbeitet und in eine Form überführt, die für die
* Erzeugung von Links auf die DARIAH-Campus Website erforderlich ist.
*
* Beispiel:
* Titel: „One phrase that appears again and again is: ,continuous training“
* URL: https://campus.dariah.eu/resource/one-phrase-that-appears-again-and-again-is-
*       continuous-training
*/
// -----
// Syntaktische Regeln
// -----
// Die zu extrahierende Titelangabe ist ggf. eingebettet in Zeichen
title : NON_LETTERS? Linktitle NON_LETTERS?;
// Der extrahierte Titel muss mit Buchstaben oder Zahlen beginnen
// und enden. Dazwischen sind Symbole möglich
linktitle: LETTERS ((LETTERS | NON_LETTERS)* LETTERS)?;
// -----
// Lexikalische Regeln
// -----
// Buchstaben und Zahlen
LETTERS : [a-zA-Z0-9]+;
// Symbole werden gemeinsam mit einem folgenden Leerzeichen
// durch ein ,-, Symbol ersetzt
NON_LETTERS : ~[a-zA-Z0-9]+ { setText(„-,“); };
```

#### *Grammatik 1: EscapeTitle zur Transformation von Titeln in die URL-konforme Variante*

Der Einstieg in die Verarbeitung des Titels in Grammatik 1 besteht in der Parserregel `title`, die neben dem inhaltlichen `linktitle` optionale, ggf. mehrfache vor- und nachgestellte Leerzeichen und Symbole (`NON_LETTERS`) identifiziert. Der für die Produktion der URL relevante Inhalt von `linktitle` wird durch

die lexikalischen Bestandteile **LETTERS** und **NON\_LETTERS** gebildet, wobei als erstes und letztes Zeichen nur **LETTERS** erlaubt sind.

Die lexikalischen Regeln werden so geformt, wie diese für die Formulierung der syntaktischen Regeln erforderlich sind:

- **LETTERS**: alle Zeichen a-z, A-Z und Zahlen 0-9, die als gültige Zeichen in die URL übernommen werden können
- **NON\_LETTERS**: Alle übrigen Zeichen werden durch ein „-“ ersetzt; hierbei werden unmittelbar aufeinanderfolgende **NON\_LETTERS** als Gruppe ersetzt

Mittels Rechtsklick →Anzeigen (nach Authentifizierung Rechtsklick →Bearbeiten) auf die jeweiligen Knoten im Elementmodell werden *Editoren für die Grammatik bzw. Transformationsfunktion* angezeigt. Abbildung 7 zeigt das Ergebnis einer Anwendung der im linken Bereich eingepflegten Grammatik **EscapeTitle** auf einen Beispieltitel (rechts oben in der Abbildung). Aufgrund der sprachlichen Interpretation von Daten wird so ein Syntaxbaum als Zwischenansicht produziert (in der Abbildung rechts unten).

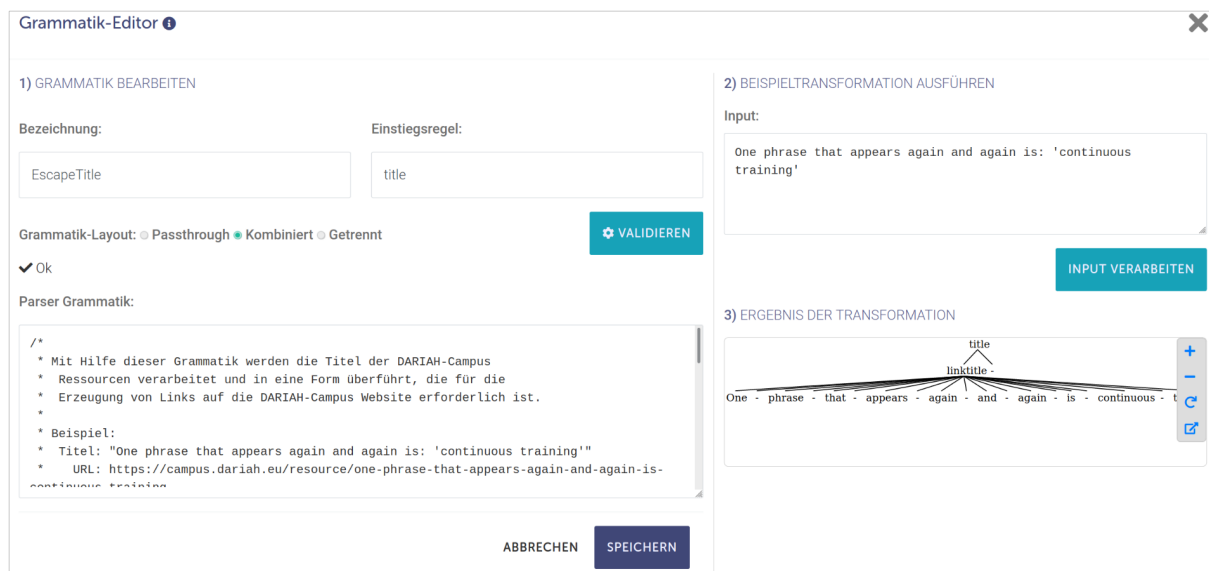


Abbildung 7: Dialog für die interaktiven Definition der Grammatik **EscapeTitle**

Die nachgelagerte Transformationsfunktion in Abbildung 8 selektiert mit **linktitle** unmittelbar den nicht-terminalen Knoten des Syntaxbaumes und weist dessen Inhalt dem Element **EscapedTitle** zu. Durch die zusätzliche Anwendung des **toLower** Kommandos im Core-Namensraum des DME werden Titel im Rahmen der Wertzuweisung in Kleinbuchstaben konvertiert.



Abbildung 8: Dialog für die interaktiven Definition der Funktion *EscapeTitle*

Datenmodelle, Grammatiken und Transformationsregeln werden oft anhand ausgewählter Beispieldaten entwickelt. Die Modellierung führt jedoch zu einer Spezifikation von *Verarbeitungslogik auf schematischer Ebene*. Einmal entwickelt können Modelle mit ihren Grammatiken und Regeln *auf eine unbeschränkte Menge von Daten ausgeführt werden* - soweit diese eine *strukturelle Homogenität* zu den verwendeten Beispieldaten aufweisen.

### Strukturierung von Inhaltsbereichen

Die grundlegende Strukturierung von Dateiinhalten fällt sowohl in grammatikalischer Spezifikation als auch anschließender Transformationsfunktionalität umfangreicher und komplexer aus. Die Grammatik *ParseText*<sup>27</sup> wird nur in Auszügen in dieses Dokument übernommen.

Ressourcen von DARIAH-Campus liegen in Form von Textdokumenten vor und sind zunächst als unstrukturierter Text zu behandeln. Die Textdateien beinhalten Auszeichnungen nach Markdown-Syntax. Markdown<sup>28</sup> ist eine einfache Auszeichnungssprache, die das Verfassen und Lesen von Texten für das Web erleichtert und zudem einfach in XHTML konvertiert werden kann. Zu diesem Zweck wird das Markdown-Format auch in DARIAH-Campus eingesetzt: Markdown-Dokumente werden im Rahmen des GatsbyJS<sup>29</sup> Seitengenerators in HTML konvertiert und angezeigt<sup>30</sup>.

<sup>27</sup><https://doi.org/10.20375/0000-000C-E12E-3>

<sup>28</sup><https://daringfireball.net/projects/markdown/syntax>

<sup>29</sup><https://www.gatsbyjs.org>

<sup>30</sup><https://github.com/DARIAH-ERIC/dariah-campus/blob/master/src/utlis/markdown-to-html.js>

Wenngleich Markdown primär präsentationsrelevante Auszeichnungen beinhaltet, so können mit Hilfe der Grammatik *ParseText* (Grammatik 2) sprachliche Bestandteile der Dokumente so herausgearbeitet werden, dass eine weiterführende Nachnutzung der Ressourcen erreicht werden kann.

```
// -----  
// Syntaktische Regeln  
// -----  
  
// Das Gesamtdokument besteht aus einem Metadaten-Block und einem  
// (optionalen) Markdown-Inhaltsbereich  
mdx: md_block body?;  
  
// Der Metadatenblock wird eingeleitet und abgeschlossen durch die Zeichenfolge ---  
// dazwischen finden sich Metadaten-Elemente  
md_block: MD_OPEN MD_NEWLINE* md_element+ MD_CLOSE;  
  
// Metadaten-Elemente sind entweder:  
// - einzeilige (auch leere) Schlüssel/Wert-Paare,  
// - mehrzeilige Schlüssel/Wertlisten (jedes Listenelement eingeleitet durch  
//   ' -' [MD_WS MD_WS MD_DASH]) oder  
// - mehrzeilige Schlüssel/Wert-Paare, bei denen aufgrund der Länge des Wertes  
//   Zeilenumbrüche eingefügt sind  
md_element : md_key md_keyval_sep md_value? MD_NEWLINE+  
            | md_key md_keyval_sep MD_NEWLINE (MD_WS? MD_WS? MD_DASH MD_WS md_list_item  
            MD_NEWLINE)+  
            | md_key md_keyval_sep md_value_multiline+ MD_NEWLINE+;  
...  
// -----  
// Lexikalische Regeln  
// -----  
  
// Metadaten-, Inhalts- und Box-Bereich werden als unterschiedliche Teilsprachen  
// interpretiert. Durch die Anwendung von Modi werden diese lexikalisch  
// unterschiedlich behandelt  
  
MD_OPEN      : '---'  -> pushMode(METADATA);  
BODY_OPEN    : .      -> pushMode(BODY);  
  
//-----  
//---- Metadaten  
//-----  
mode METADATA;  
MD_CLOSE     : '---'  -> pushMode(BODY);
```

```

MD_CATCHALL      : ~[\-:\n #|"]+;
MD_DASH          : '-';
...

```

## Grammatik 2: ParseText zur Differenzierung von Metadaten und Inhalten

Die Zielsetzung der **ParseText** Grammatik besteht in der Differenzierung sprachlich unterschiedlicher Bereiche. So enthält eine Datei (**mdx**) einen Metadatenblock (**md\_block**) und einen optionalen Inhaltsbereich (**body?**). Der Metadatenblock wird eingeleitet durch **MD\_OPEN** (die Zeichenfolge **---**) und optionale Zeilenumbrüche (**MD\_NEWLINE**). Es wird abgeschlossen durch **MD\_CLOSE** (erneut **---**). **MD\_OPEN** und **MD\_CLOSE** identifizieren zwar gleiche Tokens **---**, befinden sich jedoch in unterschiedlichen Lexer-Modi<sup>31</sup>. Mit Hilfe solcher Modi können lexikalische Regeln je Bereich unterschiedlich und ohne Wechselwirkungen definiert werden. So bildet die Struktur der Metadatenelemente (**md\_element**) mit unterschiedlichen Formen von Schlüssel-Wert-Paaren eine Teilsprache, die unabhängig vom Inhaltsbereich (**body**) durch die Anwendung von Modi umgesetzt werden kann.

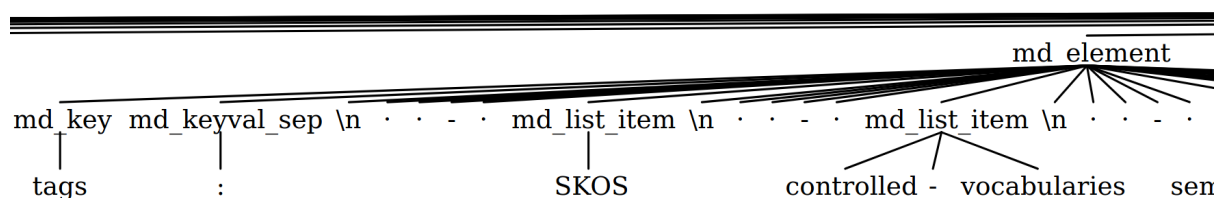


Abbildung 9: Ausschnitt eines Syntaxbaums nach Anwendung der ParseTree Grammatik

Durch Anwendung der Grammatik auf Markdown Dokumente von DARIAH-Campus werden Strukturbestandteile für eine Weiterverarbeitung der Daten nutzbar. Abbildung 9 zeigt einen Ausschnitt eines Syntaxbaumes an der Stelle einer **tags** Metadatenangabe.

Die nachgelagerte Transformationsfunktion kann nun eine Selektion anhand der syntaktischen Regeln vornehmen. Eine entsprechende Anweisung zur Zuweisung von Werten an zu erzeugende **Tag** Elemente ist in Funktion 1 dargestellt. Die Funktion iteriert dabei durch diejenigen **md\_element** Knoten, deren untergeordneter **md\_key** **"tags"** lautet und entnimmt entsprechende Werte. Durch die Zusammenfassung aller **md\_value** Knoten durch **core.concat** werden mehrzeilige Eingaben zusätzlich berücksichtigt. **md\_list\_item** selektiert Listenelemente.

```

@md_element[@md_key=="tags"]
    Tag = core.concat(@md_value);
    Tag = @md_list_item;
[endif]

```

Funktion 1: Zuweisung des Tag-Zielelementes basierend auf Knoten des Syntaxbaumes

<sup>31</sup><https://github.com/antlr/antlr4/blob/master/doc/lexer-rules.md/#lexical-modes>

Die Anwendung der **ParseTree** Grammatik und der Transformationsfunktionen erzeugen - transformiert in XML - die semistrukturierten Daten 1.<sup>32</sup>

```
<dme:Title><dme:EscapedTitle>one-phrase-that-appears-again-and-again-is--continuous-  
training</dme:EscapedTitle>One phrase that appears again and again is: 'continuous  
training'</dme:Title>  
<dme:Date>2019-04-23</dme:Date>  
<dme:Category>desir</dme:Category>  
<dme:Category>dariah</dme:Category>  
<dme:Author>claire-clivaz</dme:Author>  
<dme:Tag>research-infrastructures</dme:Tag>  
<dme:Tag>training-education</dme:Tag>  
<dme:Abstract> 'Claire Clivaz (...)</dme:Abstract>  
<dme:Type>video</dme:Type>
```

*Daten 1: Export transformierter Beispieldaten in XML*

### **Volltext extrahieren**

Durch die Anwendung der **ParseTree** Grammatik im vorausgegangenen Abschnitt wurden Metadaten von Inhalten abgetrennt. Der Inhaltsbereich wurde einem **Markdown** Element zugewiesen, welches nun die textuellen Inhalte des Dokuments inklusive Markdown Auszeichnungen beinhaltet. Für eine Durchsuchbarkeit der textuellen Daten sind diese nun von eingebetteten Annotationen zu bereinigen, so dass ausschließlich Volltext extrahiert und letztlich durch die Suche indexiert werden kann.

Die hierzu entwickelte **ExtractFulltext**<sup>33</sup> Grammatik unterscheidet beispielsweise auf lexikalischer Ebene zwischen text- und annotationsbedingten Zeichen und verwirft letztere. Die zahlreichen *lexikalischen* Regeln in Grammatik 3 dienen der Identifikation von strukturellen Bereichen, sowie der Unterscheidung zwischen relevanten und irrelevanten Tokens. Die nur wenigen *syntaktischen* Regeln fassen inhaltlich relevante Tokentypen schließlich zusammen und bilden aus diesen **text** und **heading** Knoten.

```
// -----  
// Syntaktische Regeln  
// -----  
// Syntaktisch besteht das Dokument aus Überschriften und Textblöcken  
document : (heading | text)+;  
// Die Textregel fasst die erlaubten lexikalischen Elemente für Inhalte zusammen  
text : (TEXT | ASTERISK | NL | OPENQUOTE | WS | CODEBLOCK_CONTENT | CODEBLOCK_CLOSE |  
      CODEINLINE_CONTENT | CLOSETAG)+;  
// Inhalte von Überschriften werden durch einen Lexer MODE abgebildet und werden dadurch
```

<sup>32</sup>Datensätze werden zu Gunsten der Lesbarkeit gekürzt. Die ungekürzten Daten finden sich jeweils im Anhang.

<sup>33</sup><https://doi.org/10.20375/0000-000C-E12D-4>

```
// durch andere lexikalische Elemente repräsentiert, als die Textregel
heading : HEADING_CONTENT HEADING_CLOSE;

// -----
// Lexikalische Regeln
// -----
// Überschriften werden durch #-Zeichen in einer neuen Zeile angezeigt
// Die Regel erkennt solche Überschriften, verweist auf den HEADING Mode und
// verwirft die \n# Token, da diese inhaltlich irrelevant sind
HEADING_OPEN  : '\n#' HEADING_TYPE? { setText("\n\n"); } -> skip, pushMode(HEADING);
// Inline-Code und Code-Blöcke erlauben viele Zeichen, die in normalen Inhaltsbereichen
// nicht zulässig sind. Es wird in entsprechende Modi verwiesen. Die Symbole zur
// Kennzeichnung von Codeblöcken werden verworfen.
CODEBLOCK_OPEN : ' ' -> skip, pushMode(CODEBLOCK);
CODEINLINE_OPEN : '`' -> skip, pushMode(CODEINLINE);
// Bei Blockzitat-Angaben wird das vorangestellte '\n>' verworfen
OPENQUOTE      : '\n> ' { setText(" "); };
// Reguläre Zeilenumbrüche werden durch ein Leerzeichen ersetzt
NL              : '\n' { setText(" "); };
...
```

Grammatik 3: ExtractFulltext zur Extraktion von Volltext aus Markdown

### Verfeinerung und Abschluss des Modells

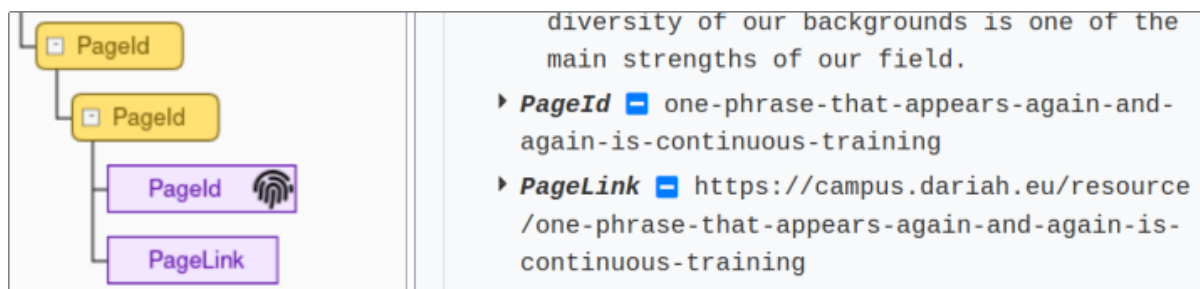
Mit der grundlegenden Strukturierung des Dokumentes, der Extraktion von Metadaten und der Bereinigung textueller Inhalte ist der wesentliche Teil der Datenmodellierung abgeschlossen. Es können jedoch weitere Metadaten produziert werden, die für eine spätere Indexierung der Daten und deren Präsentation im Rahmen des Tutorial Finders nützlich sind und das Elementmodell ergänzen.

Derartige Modellierungsschritte sind in sehr vielen Fällen nicht unmittelbar aus den Daten ableitbar, sondern ergeben sich oft erst durch die konzeptuelle Betrachtung des Zielkontexts oder die tatsächliche Überführung der Daten in diesen (Mapping im nächsten Abschnitt).

### URL zur Webseite der Ressource

Eine nach der bisherigen Modellierung noch offene Aufgabe besteht in der Konstruktion der URL zur Repräsentation der Ressource in DARIAH-Campus, auf die im Suchergebnis verwiesen werden soll. In Fällen, in denen die Metadaten der Ressource einen Wert `slug` definieren, findet letzterer hierfür ohne weitere Transformation Anwendung. Eine Analyse von Ressourcen und deren Darstellung in DARIAH-Campus zeigt, dass die benötigte URL zumeist jedoch aus dem Titel erzeugt wird. Die Basis hierfür besteht dann in dem durch `EscapeTitle` erzeugten `EscapedTitle` Element.

Die URL wird unmittelbar unter dem Wurzelknoten **MDXfile** modelliert. Die Grammatik **PageId** ist eine so genannte Passthrough-Grammatik, leitet also einen übergebenen Input unverändert an die Transformationsfunktion **PageId** weiter.



Grammatik 4: *PageId* und *PageLink* mit Beispielergebnissen

Die Elemente **PageId** und **PageLink** werden in Abhängigkeit von der Existenz des **Slug** Metadatum erzeugt. Funktion 2 beinhaltet aus diesem Zweck eine if/else-Anweisung. Als Eingabewert steht bei Verwendung einer Passthrough-Grammatik die vollständige, übergeordnete Elementhierarchie zur Verfügung.

```
[core.nullorepty(@PageId.ParseText.Slug)==true]
  PageId = @PageId.ParseText.Title.EscapedTitle;
  PageLink = core.concat("https://campus.dariah.eu/resource/",
                        @PageId.ParseText.Title.EscapedTitle);
[else]
  PageId = @PageId.ParseText.Slug;
  PageLink = core.concat("https://campus.dariah.eu/resource/", @PageId.ParseText.Slug);
[endif]
```

Funktion 2: Zuweisung von *PageId* und *PageLink* in Abhängigkeit des *Slug* Elements

## Youtube-Daten anreichern

Als weiteres Detail finden sich in den entsprechenden Ressourcen von DARIAH-Campus Hinweise auf Bezeichner von Youtube Videos (Daten 2).

```
<VideoCard
  title="One phrase that appears again and again is: 'continuous training'"
  id="mI1yfjAkwrM"
  image="images/clivaz-training.jpg" />
```

Daten 2: Verweis auf ein YouTube Video in einer DARIAH Campus Ressource

Entsprechende Angaben werden im Rahmen der **ParseText** Grammatik behandelt und führen zu der Erzeugung eines **Video** Elements mit untergeordneter **Id** (Abbildung 10). Durch die Übermittlung einer REST-basierten Anfrage an die YouTube Data API<sup>34</sup>

```
VideoListResponse = file.embed(
  core.concat("https://www.googleapis.com/youtube/v3/videos?",
    "part=contentDetails%2Csnippet%2CrecordingDetails%2C%20",
    "statistics%2C%20status%2C%20topicDetails&id=",
    @FetchDetails,
    "&key=AIZAyDGCZrJbOUQmykVadtFvVrWbTuaf3BzuU"
  )
);
```

### Funktion 3: Aufruf der YouTube Data API

werden die zu YouTube Videos verfügbaren Metadaten bezogen und als **VideoListResponse** Element mit untergeordneter Hierarchie in jeweilige Datensätze aufgenommen.

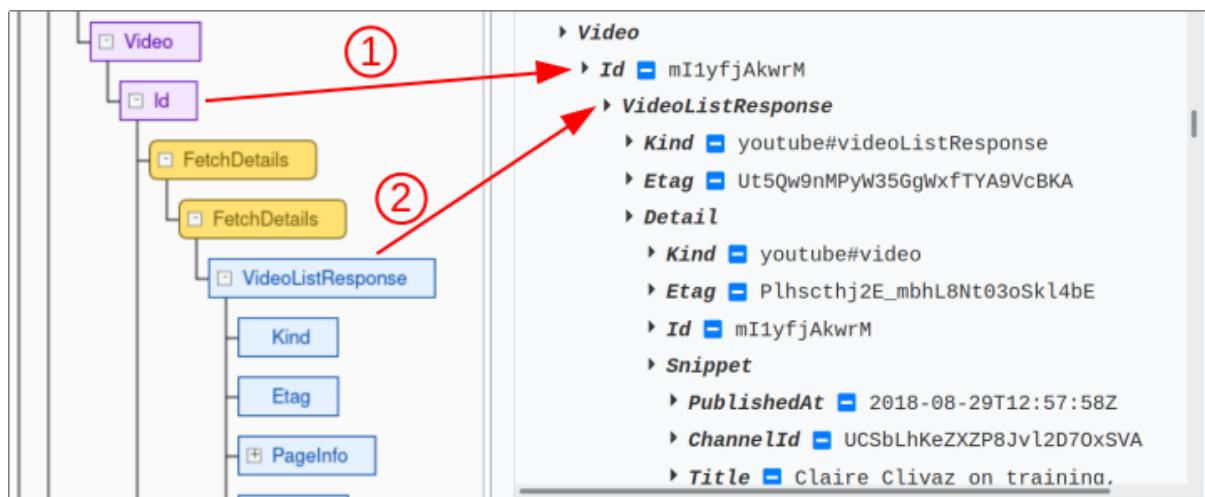


Abbildung 10: Anreicherung um YouTube Metadaten

<sup>34</sup><https://developers.google.com/youtube/v3>

## Modellierungsergebnis

Als Ergebnis der angeführten Modellierungsschritte steht auf Grundlage der Markdown-annotierten Inhalte und der Anwendung von Funktionalität zur Definition und Transformation von Daten nun eine Elementhierarchie bereit (Daten 3).

```
<dme:MDXfile xmlns:dme="https://teaching.clariah.de/dme/model/5f1d40316117b104a5fd3944/">
  <dme:Title>
<dme:EscapedTitle>one-pharse-that-appears-again-and-again-is--continuous-
training</dme:EscapedTitle>      One phrase that appears again and again is: 'continuous
training'</dme:Title>
  <dme:Date>2019-04-23</dme:Date>
  <dme:Category>desir</dme:Category>
  <dme:Category>dariah</dme:Category>
  <dme:Author>claire-clivaz</dme:Author>
  <dme:Tag>research-infrastructures</dme:Tag>
  <dme:Tag>training-education</dme:Tag>
  <dme:Abstract> 'Claire Clivaz explains how she has found that the tensions between
disciplines in interdisciplinary work can be similar no matter ...</dme:Abstract>
  <dme:Type>video</dme:Type>
  <dme:Video>
    <dme:Id>
      <VideoListResponse>
        <Kind>youtube#videoListResponse</Kind>
        <Etag>Ut5Qw9nMPyW35GgWxfTYA9VcBKA</Etag>
        <Detail>
          <Kind>youtube#video</Kind>
          <Etag>PlhscThj2E_mbhL8Nt03oSk14bE</Etag>
          <Id>mI1yfjAkwrM</Id>
          <Snippet>
            <PublishedAt>2018-08-29T12:57:58Z</PublishedAt>
            <ChannelId>UCSbLhKeZXP8Jv12D70xSVA</ChannelId>
            <Title>Claire Clivaz on training, education and research infrastructures</Title>
            <Description>Toma Tasovac speaks with Claire Clivaz from the Swiss Institute of
Bioinformatics about the role of training and ...</Description>
            <Thumbnails>
              <Default>
                <Url>https://i.ytimg.com/vi/mI1yfjAkwrM/default.jpg</Url>
                <Width>120</Width>
                <Height>90</Height>
              </Default>
              ...
            </Thumbnails>
          </Snippet>
        </Detail>
      </VideoListResponse>
    </Id>
  </dme:Video>
</dme:MDXfile>
```



```

<dme:Links>
  <dme:Resource>https://www.youtube.com/watch?v=mI1yfjAkwrM</dme:Resource>
<dme:Details>https://www.googleapis.com/youtube/v3/videos?part=contentDetails%2CrecordingDe
tails%2C%20statistics%2C%20status%2C%20topicDetails&id=mI1yfjAkwrM&key=AIzaSyDGCZrJ
bOUQmykVadtFVwVrWbTuaf3BzuU</dme:Details>
  <dme:Embedded>
    <dme:Enabled>true</dme:Enabled>
    <dme:URL>https://www.youtube.com/embed/mI1yfjAkwrM</dme:URL>
  </dme:Embedded>
</dme:Links>mI1yfjAkwrM</dme:Id>
</dme:Video>
<dme:Markdown>
  <dme:ExtractedFulltext> Claire Clivaz explains how she has found that the tensions
  between disciplines in interdisciplinary work can be similar no ...
  </dme:ExtractedFulltext>
</dme:Markdown>
<dme:PageId>one-phrase-that-appears-again-and-again-is-continuous-training</dme:PageId>
<dme:PageLink>https://campus.dariah.eu/resource/one-phrase-that-appears-again-and-again-
is-continuous-training</dme:PageLink>
</MDXfile>

```

*Daten 3: Datensatz nach Strukturierung und Anreicherung mit Hilfe des DME (gekürzt)*

## Schritt 4: Herstellung von Mappings

Die durch die Modellierung entstandene Elementhierarchie führt zu einer angereicherten und strukturierten Repräsentation von Daten, welche nun zur Herstellung von Mappings eingesetzt werden kann. Für den im Beispiel modellierten Anwendungsfall des CLARIAH-DE Tutorial Finders besteht das derzeit gewählte Integrationsmodell (Facettierung der Suche, Filtern von Suchergebnissen etc.) im DataCite Metadata Schema<sup>35</sup>, welches um das **Fulltext**-Element erweitert wurde.

Der Bildschirmausschnitt in Abbildung 11 zeigt eine Übersicht des Mappings des DARIAH-Campus Datenmodells auf das erweiterte DataCite Datenmodell<sup>36</sup>. Das Farbschema der Elementtypen findet sich auch in der Mappingansicht, wobei die Unterscheidung zwischen **Parser-Elementen** und **produzierten** Elementen für das Mapping unerheblich ist. In den **goldfarbenen Knoten** sind Grammatiken und Transformationsregeln aller verbundenen Eingangselemente visuell gekapselt. Mittels Rechtsklick → Anzeigen (nach Authentifizierung Rechtsklick → Bearbeiten) auf diese **Knoten** können Grammatiken und Regeln eingesehen und angepasst werden.

Ein wesentlicher Grundgedanke für die Modellierung im Rahmen des DME besteht darin, Datenmodelle in der Verarbeitung, Bereinigung und Anreicherung von Daten möglichst umfassend zu gestalten,

<sup>35</sup><https://schema.datacite.org>

<sup>36</sup><https://teaching.clariah.de/dme/mapping/editor/5f2148346117b104a5fd3c02/>

die Mappings in dieser Hinsicht aber möglichst minimal. Dies hat insbesondere den Grund, dass Erweiterungen am Datenmodell durch sämtliche Mappings unmittelbar nachgenutzt werden können. Am Mapping implementierte Funktionalität wird jeweils nur bei dessen Ausführung verwendet.

Ein wichtiges, wenngleich nicht immer trennscharfes Entscheidungsmerkmal, ob eine Grammatik oder Transformation am Modell oder Mapping einzurichten ist, besteht in der Kontextualität der zu erzeugenden Elemente. Betreffen diese ausschließlich den konkreten Verwendungskontext, so sollten die Elemente am Mapping modelliert werden. Sind sie dagegen unabhängig von einer konkreten Verwendung und universell für die modellierten Daten gültig, sollte eine Modellierung am Datenmodell erfolgen.

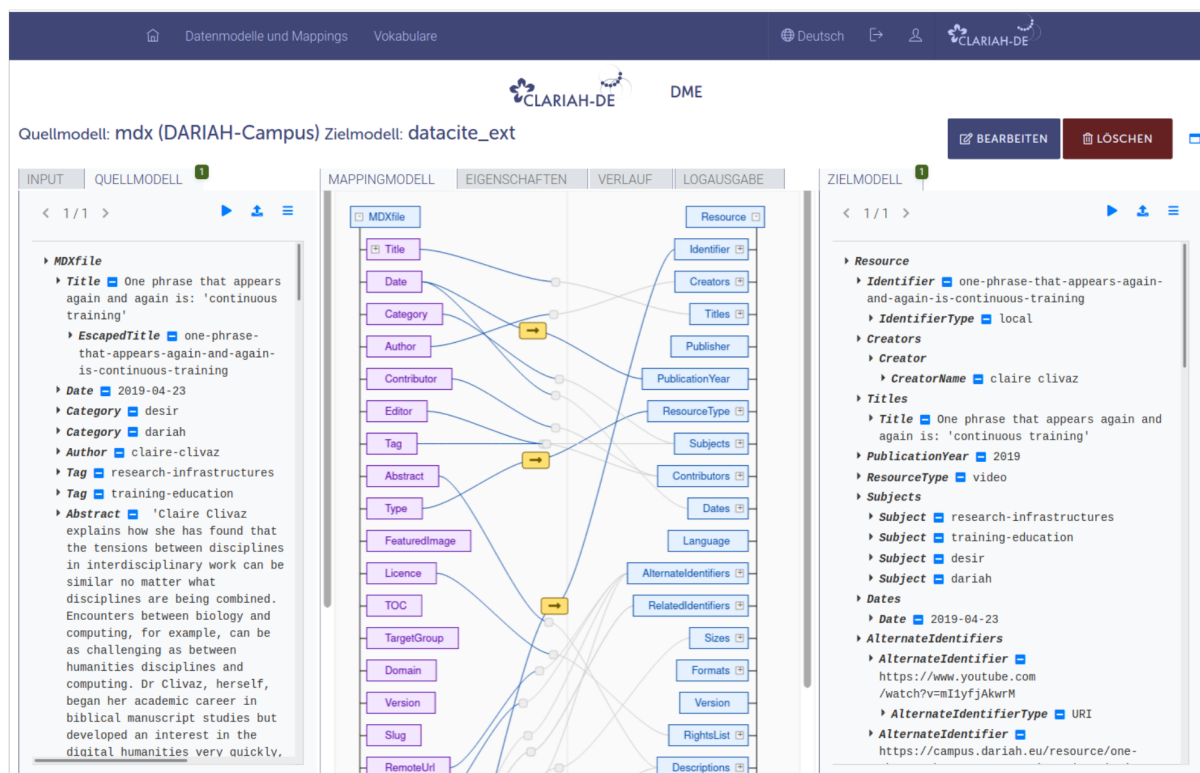


Abbildung 11: Mapping des Datenmodells von DARIAH-Campus zu DataCite

Durch die Anwendung des Mappings auf Beispieldaten werden diese in das Zielformat überführt. Abbildung 11 zeigt im rechten Bereich einen Ausschnitt der transformierten Daten im Zielformat, ein Export als XML führt zu nachfolgender Ausgabe.

```

<ns1:resource>
  <ns1:identifier ns1:identifierType="local">one-phrase-that-appears-again-and-again-is-
continuous-training</ns1:identifier>
  <ns1:creators>
    <ns1:creator>
      <ns1:creatorName>claire clivaz</ns1:creatorName>
    </ns1:creator>
  </ns1:creators>
  <ns1:titles>
    <ns1:title>One phrase that appears again and again is: 'continuous training'</ns1:title>
  </ns1:titles>
  <ns1:publicationYear>2019</ns1:publicationYear>
  <ns1:resourceType>video</ns1:resourceType>
  <ns1:subjects>
    <ns1:subject>research-infrastructures</ns1:subject>
    <ns1:subject>training-education</ns1:subject>
    <ns1:subject>desir</ns1:subject>
    <ns1:subject>dariah</ns1:subject>
  </ns1:subjects>
  <ns1:dates>
    <ns1:date>2019-04-23</ns1:date>
  </ns1:dates>
  <ns1:alternateIdentifiers>
    <ns1:alternateIdentifier
      ns1:alternateIdentifierType="URI">https://www.youtube.com/watch?v=mI1yfjAkwrM
    </ns1:alternateIdentifier>
    <ns1:alternateIdentifier
      ns1:alternateIdentifierType="URI">https://campus.dariah.eu/resource/one-phrase-that-
appears-again-and-again-is-continuous-training</ns1:alternateIdentifier>
  </ns1:alternateIdentifiers>
  ...
  <ns1:descriptions>
    <ns1:description ns1:descriptionType="Abstract"> 'Claire Clivaz explains how she has
found that the tensions ...</ns1:description>
    <ns1:description ns1:descriptionType="Other">Toma Tasovac speaks with Claire Clivaz from
the Swiss Institute of ...</ns1:description>
  </ns1:descriptions>
  <fulltext> Claire Clivaz explains how she has found that the tensions between disciplines
in interdisciplinary ...</fulltext>
</ns1:resource>

```

*Daten 4: In erweitertes DataCite überführter Datensatz von DARIAH Campus (gekürzt)*

## Schritt 5: Integration von Daten

Mit der Eintragung der Sammlung in die CR, der Modellierung der Daten und der Herstellung des Mappings auf DataCite stehen dem CLARIAH-DE Tutorial Finder sämtliche Informationen zur Verfügung, die für den Abruf, die Indexierung und die *Integration* der DARIAH-Campus Daten erforderlich sind. Mit der Eintragung der Sammlung, der Modellierung und des Mappings der Daten sind die *Aufgaben des Sammlungs- und Datenexperten abgeschlossen*. Administrative Aufgaben finden in produktiven Instanzen der Generischen Suche in zyklischen Abständen automatisch statt, so dass Ressourcen nach einem erfolgten Indexierungslauf aufgefunden werden können.

Die folgenden Bildschirmausschnitte sind aus Gründen der Vollständigkeit hier angeführt. Eine manuelle Interaktion mit den administrativen Dialogen der Generischen Suche durch Nutzerinnen und Nutzer ist typischerweise nicht erforderlich.

The screenshot displays the CLARIAH-DE search interface. At the top, there is a navigation bar with 'Suchen', 'Daten', and 'Hilfe' menus, along with a language selector set to 'Deutsch' and a user profile icon. The main header features the CLARIAH-DE logo and the text 'LEHRMATERIAL SUCHE'. Below this, the 'DARIAH-Campus' collection is highlighted. The 'Sammlung' section shows the collection's ID and a source URL. The 'Verfügbare Datenquellen' section is divided into three panels: 'ENDPUNKT' (Endpoint), 'DATENABRUF & INDEXIERUNG' (Data Retrieval & Indexing), and 'DATENSATZ' (Dataset). The 'ENDPUNKT' panel shows a green checkmark and details about the GitHub repository. The 'DATENABRUF & INDEXIERUNG' panel shows a green checkmark and a plug icon, indicating online status, along with buttons for 'LOGAUSGABE ANZEIGEN', 'DATEN LÖSCHEN', and 'NEU INDEXIEREN'. The 'DATENSATZ' panel shows a green checkmark and details about the dataset, including the number of indexed documents and secondary data. At the bottom, the 'VERFÜGBARE AKTIONEN' (Available Actions) section contains buttons for 'ONLINEDATEN CRAWLEN' and 'DATEN ZURÜCKSETZEN...'.

CLARIAH-DE LEHRMATERIAL SUCHE

DARIAH-Campus

Sammlung

ID: 5f21ab8f4f0ceab08ec7c2ca

Quelle: <https://c105-229.cloud.gwdg.de/colreg-ui/collections/5eba5f516117b1100fe31678>

Verfügbare Datenquellen

mdx (DARIAH-Campus)

ENDPUNKT

URL: <https://github.com/DARIAH-ERIC/dariah-campus>

Zugriff: Git Repository

Dateityp: Text

ID: 5f21ab8f4f0ceab08ec7c2c9

Status: Zugreifbar

DATENABRUF & INDEXIERUNG

Online (29.07.2020 19:17:07)

ID: 5f21aec64f0c2a438ece2334

Dateien: 14 (727.0 kB)

[LOGAUSGABE ANZEIGEN](#)

[DATEN LÖSCHEN](#)

[NEU INDEXIEREN](#)

DATENSATZ

Datenmodell: mdx (DARIAH-Campus)

Alias: nicht definiert

ID: 5f1d40316117b104a5fd3944

[i](#) Indexierte Dokumente: 110

[i](#) Sekundärdaten: 162 Dateien (2.2 MB)

Status: Inaktiv

VERFÜGBARE AKTIONEN

[ONLINEDATEN CRAWLEN](#) [DATEN ZURÜCKSETZEN...](#)

Abbildung 12: Ansicht auf die Sammlung DARIAH-Campus im Tutorial Finder

Nach der Übernahme von Sammlungsinformationen aus der CR steht ein korrespondierender Eintrag in der *Sammlungsadministration* der Suche zur Verfügung. Über diesen Eintrag können Administratoren der Suche den Zustand von Sammlungsdaten in der Suche einsehen und verschiedene Aufgaben, wie die Durchführung eines Online-Crawls, ausführen.

## Datenmodell- und Indexverwaltung in der Suche

**Data Modeling Environment (API)**

URL:  
<https://c105-229.cloud.gwdg.de/dme/>

Status: ✔ 138ms

Letzter Abgleich: ✔ 31. Juli 2020 11:57:21 MESZ

Nächster Abgleich: ⌚

**Datenmodelle**

Zeige: 10 Filter:

**SYNCHRONISIEREN**

	Name ↑ ↓	Index ↑ ↓	Dokumente ↑ ↓	
<span style="color: green;">✔</span>	datacite_ext	gs_mww2_5eb11e106117b10b7b6c7c11	0	<a href="#">INDEX ERSTELLEN</a> <a href="#">ENTFERNEN</a>
<span style="color: green;">✔</span>	mdx (DARIAH-Campus)	gs_mww2_5f1d40316117b104a5fd3944	110	<a href="#">INDEX ERSTELLEN</a> <a href="#">ENTFERNEN</a>
<span style="color: green;">✔</span>	mdx (Programming Historian)	gs_mww2_5f23d1f66117b1047c30558b	140	<a href="#">INDEX ERSTELLEN</a> <a href="#">ENTFERNEN</a>

Abbildung 13: Datenmodelle im CLARIAH-DE Tutorial Finder

Auch für Datenmodelle bietet die Suche eine einfache Verwaltungsansicht. In dieser können insbesondere Indices neu erstellt bzw. aus der Suche entfernt werden.

## Auffindbarkeit von Ressourcen

Nach Abschluss der typischerweise automatisiert durchgeführten Verwaltungsaufgaben in der Suchinstanz sind die Ressourcen der hinzugefügten Sammlung, hier DARIAH-Campus, indexiert, integriert und auffindbar.

**Suchoptionen**

☐ Erläuterungen anzeigen

20 Ergebnisse je Anfrage

**Durchsuchte Kollektionen**

- 1 DARIAH-Campus
- DHdKanal Video-Tutorials
- Teaching and Learning Materials Collection

[AUSWÄHLEN...](#) [ALLE](#)

**Einfache Suche**

Suche: "Agile Dynamic System Development Methods"

[Erweiterte Suche](#)

**Verfügbare Filter**

GENRE

- data-management (1)
- data-modeling (1)
- events (1)
- project-management (1)
- software-development (1)

[WEITERE...](#)

**1 von 1 Ressourcen**

DARIAH-CAMPUS

**'MaDiH: Research Software Engineering Training'**

[Detailseite](#)

[data-modeling](#) [software-development](#) [data-management](#) [project-management](#) [events](#)

**TREFFER IM DOKUMENT**

**System Development Methods (DSDM)** as well as various theoretical and practical approaches to digital cultural heritage.

Hosted by King's Digital Lab (KDL) at King's College London, the workshop introduced participants to best practices in project management, the **Agile Dynamic**

Abbildung 14: Darstellung eines Suchergebnisses von DARIAH-Campus

## Zusammenfassung

Die hier präsentierte Anleitung zeigt eine Möglichkeit auf, Git-basierte Ressourcen mithilfe von Infrastrukturdiensten aus CLARIAH-DE in eine Suche zu integrieren und damit übergreifend nachzunutzen. Als Anwendungsfall wurde hier DARIAH-Campus gewählt, anhand dessen die schrittweise Durchführung beginnend mit der Analyse der Sammlung bis hin zur letztendlichen Integration von Daten beschrieben und dokumentiert wird. Analog können weitere Sammlungen bzw. Repositorien angebunden werden, die ähnliche Rahmenbedingungen wie DARIAH-Campus mitbringen und die ebenso in einer Anwendung wie dem CLARIAH-DE Tutorial Finder indexiert werden sollen.

Zu den Rahmenbedingungen zählen neben den Metadaten, den eigentlichen Ressourcen (beispielsweise Volltexte) auch die Referenzen auf Browser-Repräsentationen oder anderen Darstellungen dieser Daten. Git-Repositorien als Sammlungsschnittstelle bieten sich hier hervorragend an, da bereits detailreiche Konfigurationsoptionen für diese Art der Schnittstelle bereitstehen. Die Daten können sowohl als strukturierte Daten wie auch unstrukturiert vorliegen. Letztere sind im Fall von DARIAH-Campus zu verarbeiten, welche dadurch jedoch auch die umfangreiche Modellierungs-, Transformations- und Mappingoptionen zeigt, die durch das DME ermöglicht werden, und somit aus unstrukturierten Markdown Textdateien strukturierte Daten zu gewinnen.

Die Integration von DARIAH Campus Ressourcen in die CLARIAH-DE Infrastruktur schafft die Möglichkeit, die Informationen, die in dieser Sammlung vorhanden sind und vormals oft nicht (oder nicht in übergreifender Form) durchsuchbar waren, zu indexieren und letztendlich im CLARIAH-DE Tutorial Finder zu durchsuchen<sup>37</sup>. Dabei ist nicht nur die einfache Volltextsuche verfügbar, sondern auch eine facettierte erweiterte Suche, welche Forschenden in den Geisteswissenschaften und darüber hinaus explorative Suchen ermöglicht.

---

<sup>37</sup><https://teaching.clariah.de/search/>

## Anhang

### *In XML exportierte Beispieldaten im Quellformat*

```
<?xml version='1.0' encoding='UTF-8'?>
<dme:MDXfile xmlns:dme="https://teaching.clariah.de/dme/model/5f1d40316117b104a5fd3944/">
  <dme:Title>
    <dme:EscapedTitle>one-phrase-that-appears-again-and-again-is--continuous-
    training</dme:EscapedTitle>    One phrase that appears again and again is: 'continuous
    training'</dme:Title>
  <dme:Date>2019-04-23</dme:Date>
  <dme:Category>desir</dme:Category>
  <dme:Category>dariah</dme:Category>
  <dme:Author>claire-clivaz</dme:Author>
  <dme:Tag>research-infrastructures</dme:Tag>
  <dme:Tag>training-education</dme:Tag>
  <dme:Abstract> 'Claire Clivaz explains how she has found that the tensions between
  disciplines in interdisciplinary work can be similar no matter what disciplines are being
  combined. Encounters between biology and computing, for example, can be as challenging as
  between humanities disciplines and computing. Dr Clivaz, herself, began her academic career
  in biblical manuscript studies but developed an interest in the digital humanities very
  quickly, at a time when the impact of computing was being felt in the humanities more
  widely. She explains the usefulness of the DH Course Registry in finding university-based,
  formal, DH training in Switzerland. However, she argues that informal opportunities to
  learn are crucial. One phrase that appears again and again in the digital humanities, she
  states, is: continuous training.'</dme:Abstract>
  <dme:Type>video</dme:Type>
  <dme:Video>
    <dme:Id>
      <VideoListResponse>
        <Kind>youtube#videoListResponse</Kind>
        <Etag>Ut5Qw9nMPyW35GgWxfTYA9VcBKA</Etag>
        <Detail>
          <Kind>youtube#video</Kind>
          <Etag>PlhscThj2E_mbhL8Nt03oSk14bE</Etag>
          <Id>mI1yfjAkwrM</Id>
          <Snippet>
            <PublishedAt>2018-08-29T12:57:58Z</PublishedAt>
            <ChannelId>UCSbLhKeZXP8Jv12D70xSVA</ChannelId>
            <Title>Claire Clivaz on training, education and research infrastructures</Title>
            <Description>Toma Tasovac speaks with Claire Clivaz from the Swiss Institute of
            Bioinformatics about the role of training and education in digital research
```



infrastructures. Recorded in the context of DESIR, a DARIAH-EU project, co-funded by the Horizon 2020 Framework Programme of the European Union.</Description>

<Thumbnails>

<Default>

<Url>https://i.ytimg.com/vi/mI1yfjAkwrM/default.jpg</Url>

<Width>120</Width>

<Height>90</Height>

</Default>

<Medium>

<Url>https://i.ytimg.com/vi/mI1yfjAkwrM/mqdefault.jpg</Url>

<Width>320</Width>

<Height>180</Height>

</Medium>

<High>

<Url>https://i.ytimg.com/vi/mI1yfjAkwrM/hqdefault.jpg</Url>

<Width>480</Width>

<Height>360</Height>

</High>

</Thumbnails>

<ChannelTitle>Центар за дигиталне хуманистичке науке</ChannelTitle>

<LiveBroadcastContent>none</LiveBroadcastContent>

</Snippet>

<ContentDetails>

<Duration>PT19M44S</Duration>

<Dimension>2d</Dimension>

<Definition>hd</Definition>

<Caption>>false</Caption>

<LicensedContent>>false</LicensedContent>

<Projection>rectangular</Projection>

</ContentDetails>

<Status>

<UploadStatus>processed</UploadStatus>

<PrivacyStatus>public</PrivacyStatus>

<License>youtube</License>

<Embeddable>>true</Embeddable>

<PublicStatsViewable>>true</PublicStatsViewable>

<MadeForKids>>false</MadeForKids>

</Status>

<Statistics>

<ViewCount>29</ViewCount>

<FavoriteCount>0</FavoriteCount>

</Statistics>

<TopicDetails>

```

    <RelevantTopicIds>
      <dme:Topic>/m/098wr</dme:Topic>/m/098wr</RelevantTopicIds>
    <RelevantTopicIds>
      <dme:Topic>/m/098wr</dme:Topic>/m/098wr</RelevantTopicIds>
    <TopicCategories>
      <dme:WikipediaReference>
        <dme:Page>Society</dme:Page>
        <dme:Server>en.wikipedia.org</dme:Server>
      </dme:WikipediaReference>https://en.wikipedia.org/wiki/Society</TopicCategories>
    </TopicDetails>
  </Detail>
  <PageInfo>
    <TotalResults>1</TotalResults>
    <ResultsPerPage>1</ResultsPerPage>
  </PageInfo>
</VideoListResponse>
<dme:Links>
  <dme:Resource>https://www.youtube.com/watch?v=mI1yfjAkwrM</dme:Resource>
  <dme:Details>https://www.googleapis.com/youtube/v3/videos?part=contentDetails%2CcreatingDetails%2C%20statistics%2C%20status%2C%20topicDetails&id=mI1yfjAkwrM&key=AIZA5yDGCZrJbOUQmykVadtFVwVrWbTuaf3BzuU</dme:Details>
  <dme:Embedded>
    <dme:Enabled>true</dme:Enabled>
    <dme:URL>https://www.youtube.com/embed/mI1yfjAkwrM</dme:URL>
  </dme:Embedded>
  </dme:Links>mI1yfjAkwrM</dme:Id>
</dme:Video>
<dme:Markdown>
  <dme:ExtractedFulltext> Claire Clivaz explains how she has found that the tensions between disciplines in interdisciplinary work can be similar no matter what disciplines are being combined. Encounters between biology and computing can be as challenging as between humanities disciplines and computing. Who trains new students? How do we transfer techniques across these different 'worlds'? She began her academic career in biblical manuscript studies but developed an interest in the digital humanities very quickly, at a time when the impact of computing was being felt in the humanities more widely. She explains the usefulness of the DH Course Registry in finding university-based, formal, DH training in Switzerland. However, she argues that informal opportunities to learn are crucial. One phrase that appears again and again in the digital humanities, she states, is: continuous training. This is part of a series entitled 'What is the Role of Training and Education in Research Infrastructures?'. DARIAH Director Toma Tasovac spoke to fellow DARIAHns and colleagues from prospective DARIAH countries about their own educational backgrounds and the role of training and education in digital research infrastructures. The interviews show that there is no

```

single educational trajectory which produces DH scholars. The diversity of our backgrounds is one of the main strengths of our field.

```
</dme:ExtractedFulltext>
</dme:Markdown>
<dme:PageId>one-phrase-that-appears-again-and-again-is--continuous-training</dme:PageId>
<dme:PageLink>https://campus.dariah.eu/resource/one-phrase-that-appears-again-and-again-
is--continuous-training</dme:PageLink>
</MDXfile>
```

*Daten 5: Datensatz nach Strukturierung und Anreicherung mit Hilfe des DME*

### *In XML exportierte Beispieldaten im Zielformat*

```
<?xml version='1.0' encoding='UTF-8'?>
<ns1:resource xmlns:dme="https://teaching.clariah.de/dme/model/5eb11e106117b10b7b6c7c11/"
xmlns:ns1="http://datacite.org/schema/kernel-4"
xmlns:ns2="http://www.w3.org/2001/XMLSchema">
<ns1:identifier ns1:identifierType="local">one-phrase-that-appears-again-and-again-is-
continuous-training</ns1:identifier>
<ns1:creators>
<ns1:creator>
<ns1:creatorName>claire clivaz</ns1:creatorName>
</ns1:creator>
</ns1:creators>
<ns1:titles>
<ns1:title>One phrase that appears again and again is: 'continuous training'</ns1:title>
</ns1:titles>
<ns1:publicationYear>2019</ns1:publicationYear>
<ns1:resourceType>video</ns1:resourceType>
<ns1:subjects>
<ns1:subject>research-infrastructures</ns1:subject>
<ns1:subject>training-education</ns1:subject>
<ns1:subject>desir</ns1:subject>
<ns1:subject>dariah</ns1:subject>
</ns1:subjects>
<ns1:dates>
<ns1:date>2019-04-23</ns1:date>
</ns1:dates>
<ns1:alternateIdentifiers>
<ns1:alternateIdentifier
```

```

ns1:alternateIdentifierType="URI">https://www.youtube.com/watch?v=mI1yfjAkwrM
</ns1:alternateIdentifier>
<ns1:alternateIdentifier
ns1:alternateIdentifierType="URI">https://campus.dariah.eu/resource/one-phrase-that-
appears-again-and-again-is-continuous-training</ns1:alternateIdentifier>
</ns1:alternateIdentifiers>
<ns1:relatedIdentifiers>
  <ns1:relatedIdentifier ns1:resourceTypeGeneral="IFrame"
ns1:relatedIdentifierType="URI">https://www.youtube.com/embed/mI1yfjAkwrM
  </ns1:relatedIdentifier>
  <ns1:relatedIdentifier ns1:resourceTypeGeneral="Image"
ns1:relatedIdentifierType="URI">https://i.ytimg.com/vi/mI1yfjAkwrM/hqdefault.jpg
  </ns1:relatedIdentifier>
  <ns1:relatedIdentifier ns1:resourceTypeGeneral="Thumbnail"
ns1:relatedIdentifierType="URI">https://i.ytimg.com/vi/mI1yfjAkwrM/hqdefault.jpg
  </ns1:relatedIdentifier>
</ns1:relatedIdentifiers>
<ns1:sizes>
  <ns1:size>PT19M44S</ns1:size>
</ns1:sizes>
<ns1:descriptions>
  <ns1:description ns1:descriptionType="Abstract"> 'Claire Clivaz explains how she has
found that the tensions between disciplines in interdisciplinary work can be similar no
matter what disciplines are being combined. Encounters between biology and computing,
for example, can be as challenging as between humanities disciplines and computing.
Dr Clivaz, herself, began her academic career in biblical manuscript studies but
developed an interest in the digital humanities very quickly, at a time when the
impact of computing was being felt in the humanities more widely. She explains the
usefulness of the DH Course Registry in finding university-based, formal, DH training
in Switzerland. However, she argues that informal opportunities to learn are crucial.
One phrase that appears again and again in the digital humanities, she states, is:
continuous training.'</ns1:description>
  <ns1:description ns1:descriptionType="Other">Toma Tasovac speaks with Claire Clivaz
from the Swiss Institute of Bioinformatics about the role of training and education
in digital research infrastructures. Recorded in the context of DESIR,
a DARIAH-EU project, co-funded by the Horizon 2020 Framework Programme of the
European Union.</ns1:description>
</ns1:descriptions>
<fulltext> Claire Clivaz explains how she has found that the tensions between
disciplines in interdisciplinary work can be similar no matter what disciplines are
being combined. Encounters between biology and computing can be as challenging as
between humanities disciplines and computing. Who trains new students? How do we
transfer techniques across these different 'worlds'? She began her academic career in

```

biblical manuscript studies but developed an interest in the digital humanities very quickly, at a time when the impact of computing was being felt in the humanities more widely. She explains the usefulness of the DH Course Registry in finding university-based, formal, DH training in Switzerland. However, she argues that informal opportunities to learn are crucial. One phrase that appears again and again in the digital humanities, she states, is: continuous training. This is part of a series entitled 'What is the Role of Training and Education in Research Infrastructures?'. DARIAH Director Toma Tasovac spoke to fellow DARIAHns and colleagues from prospective DARIAH countries about their own educational backgrounds and the role of training and education in digital research infrastructures. The interviews show that there is no single educational trajectory which produces DH scholars. The diversity of our backgrounds is one of the main strengths of our field. [</fulltext>](#)  
[</ns1:resource>](#)

*Daten 6: In erweitertes DataCite überführter Datensatz von DARIAH Campus*