

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

---

---

2019

## Dokumentation der Datenföderationsarchitektur

Tobias Gradl

DARIAH-DE Working Papers

Nr.39

---

Gradl, Tobias: Betrieb des DARIAH-DE Coordination Office  
Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2019  
(DARIAH-DE working papers 39)

Verfügbar:

PURL: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2019-11>

URN: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2019-11-7>

## 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:** Für die Komponenten der Datenföderationsarchitektur (DFA) wird mit der Veröffentlichung neuer Softwareversionen jeweils auch deren dedizierte Dokumentation generiert. Unter Anwendung der Prinzipien von Continuous Integration und Deployment wird dabei sichergestellt, dass Dokumentationsbausteine automatisiert erzeugt und verteilt werden. Hierdurch wird auch die Verwendung eines einheitlichen Erscheinungsbildes erleichtert, welches für die Weiterentwicklung, aber auch die Implementierung ergänzender Komponenten nachgenutzt werden kann. Dieser Beitrag beschreibt die verwendete Technologie zur automatisierten Erstellung und Verteilung der Dokumentation und bietet Entwicklern dadurch entsprechende Anknüpfungspunkte. Zudem bietet der Beitrag einen thematischen Einstieg in die DFA und deren Komponenten und stellt insbesondere Verweise zu Source Code, eigentlicher Dokumentation und den primären Instanzen zusammen.

**Keywords:** Forschungsdatenmanagement, Datenföderationsarchitektur, Dokumentation  
research data management, data federation architecture, documentation

# Dokumentation der Datenföderationsarchitektur

Tobias Gradl

Otto-Friedrich-Universität Bamberg, Fakultät für Wirtschaftsinformatik und Angewandte Informatik, Lehrstuhl für  
Medieninformatik



Tobias Gradl: „Dokumentation der Datenföderationsarchitektur“. *DARIAH-DE Working Papers* Nr. 39.  
Göttingen: DARIAH-DE, 2019. URN: [urn:nbn:de:gbv:7-dariah-2019-11-7](https://nbn-resolving.org/urn:nbn:de:gbv:7-dariah-2019-11-7).

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.



Dieser Beitrag ist ursprünglich im Februar 2019 als Report R 4.2.5 Dokumentation der Forschungsdaten-  
Föderationsarchitektur im Rahmen von DARIAH-DE (BMBF, Förderkennzeichen 01UG1610A bis J) entstanden.

## Zusammenfassung

Für die Komponenten der Datenföderationsarchitektur (DFA) wird mit der Veröffentlichung neuer Softwareversionen jeweils auch deren dedizierte Dokumentation generiert. Unter Anwendung der Prinzipien von Continuous Integration und Deployment wird dabei sichergestellt, dass Dokumentationsbausteine automatisiert erzeugt und verteilt werden. Hierdurch wird auch die Verwendung eines einheitlichen Erscheinungsbildes erleichtert, welches für die Weiterentwicklung, aber auch die Implementierung ergänzender Komponenten nachgenutzt werden kann.

Dieser Beitrag beschreibt die verwendete Technologie zur automatisierten Erstellung und Verteilung der Dokumentation und bietet Entwicklern dadurch entsprechende Anknüpfungspunkte. Zudem bietet der Beitrag einen thematischen Einstieg in die DFA und deren Komponenten und stellt insbesondere Verweise zu Source Code, eigentlicher Dokumentation und den primären Instanzen zusammen.

## Schlagwörter

Forschungsdatenmanagement, Datenföderationsarchitektur, Dokumentation

## Keywords

research data management, data federation architecture, documentation

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
<b>2</b>	<b>Dokumentation</b>	<b>6</b>
2.1	Repository . . . . .	8
2.2	Collection Registry . . . . .	9
2.3	Data Modeling Environment . . . . .	10
2.4	Generische Suche . . . . .	11
<b>3</b>	<b>Ausblick</b>	<b>12</b>
<b>4</b>	<b>Quellenverzeichnis</b>	<b>12</b>

## 1 Einleitung

Die DARIAH-DE Datenföderationsarchitektur (DFA) beschreibt ein Gesamtkonzept des Forschungsdatenmanagements und umfasst mehrere Softwarekomponenten, die den reproduzierbaren und nachhaltigen Umgang mit Forschungsdaten erleichtern:

- Das **Repository** aggregiert verschiedene Dienste und Anwendungen und macht diese komfortabel nutzbar. Es ermöglicht die nachhaltige und sichere Speicherung von Forschungsdaten. Zudem können diese mit Metadaten versehen und durch die Generische Suche auffindbar gemacht werden.
- Die **Collection Registry (CR)** ist ein Werkzeug zur Beschreibung von Forschungsdatensammlungen. Beschreibungen folgen dem umfangreichen, für kultur- und geisteswissenschaftliche Sammlungen entwickelten DARIAH Collection Description Data Model (DCDDM)<sup>1</sup>. Soweit vorhanden beinhalten Beschreibungen auch Zugriffsmethoden auf die Daten selbst, wodurch auf diese durch die Generische Suche zugegriffen werden kann.
- Das **Data Modeling Environment (DME)** ist Nachfolger der Schema und Crosswalk Registry und löst diese als DFA Komponente ab. Die DME ist ein umfassendes Werkzeug zur Datenmodellierung und -integration basierend auf dem Konzept domänenspezifischer Sprachen. Mit Hilfe der DME werden die in Sammlungen verwendeten Datenmodelle und deren Relationen zu weiteren Datenmodellen beschrieben, um Sammlungen in größtmöglicher semantischer Tiefe nachnutzen zu können.
- Die **Generische Suche (GS)** ermöglicht als Suchmaschine die Suche in digital verfügbaren, über die DARIAH-DE Collection Registry (CR) nachgewiesenen Sammlungen. Die Generische Suche kann gleichermaßen als Breiten- und Tiefensuche eingesetzt werden und nutzt insbesondere die Datenmodelle und Mappings des DME, um Daten zu verarbeiten und zu integrieren.

Dieses Dokument fasst die wesentlichen Dokumentationsbestandteile zur den DFA Komponenten zusammen. Für weiterführende Informationen zu den einzelnen Diensten wird in den entsprechenden Abschnitten auf das jeweilige Dokumentationssystem verwiesen.

Als technische Basis für die Dokumentation der DFA Komponenten wird Sphinx<sup>2</sup> verwendet, ein Werkzeug bestehend aus Python Skripten zur Erzeugung verschiedener Formate (z. B. HTML, LaTeX, PDF etc.) aus reStructuredText<sup>3</sup> Markup. Durch die Verwendung derselben technologischen Basis und eines gemeinsamen look-and-feels<sup>4</sup> der Sphinx-generierten Dokumentationen wird die Erweiterbarkeit der Dokumentationen erleichtert. Durch die Design-Konsistenz wird zudem sichergestellt, dass für Nutzer bei der Recherche in den Dokumentationen die Zusammengehörigkeit der DFA bzw. DARIAH-DE Komponenten leicht ersichtlich ist.

---

<sup>1</sup><https://github.com/DARIAH-DE/DCDDM>

<sup>2</sup><http://www.sphinx-doc.org/>

<sup>3</sup><http://docutils.sourceforge.net/rst.html>

<sup>4</sup>[https://github.com/rtfd/sphinx\\_rtd\\_theme](https://github.com/rtfd/sphinx_rtd_theme)

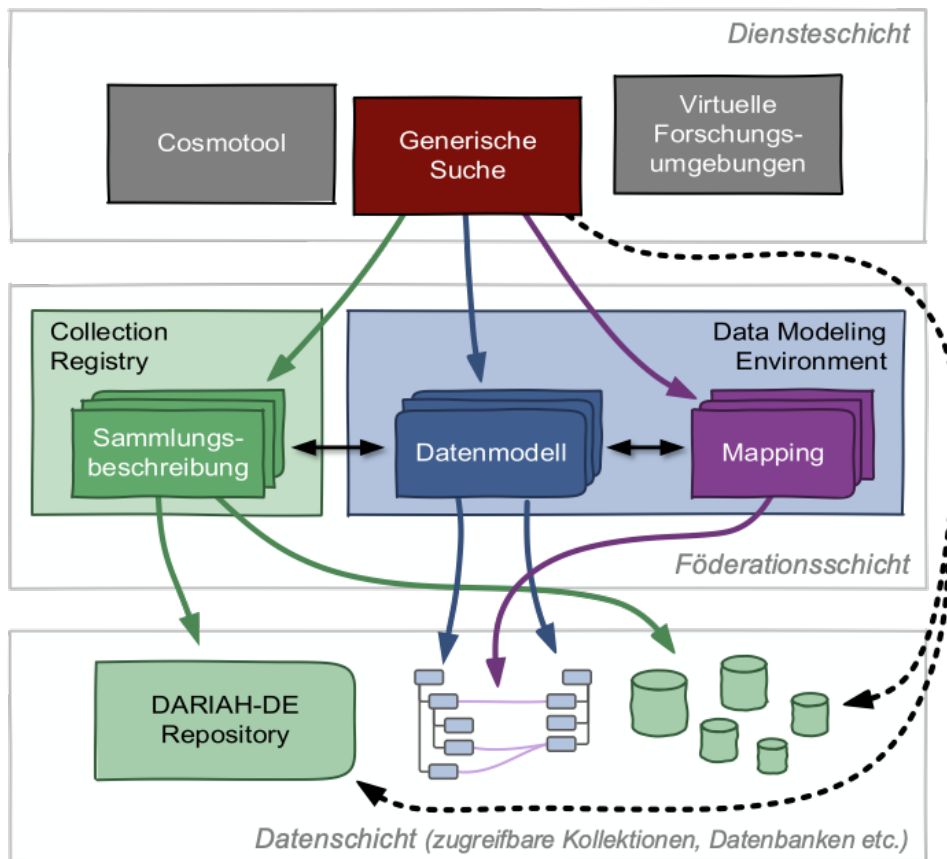


Abbildung 1: Komponenten und Beziehungen in der Datenföderationsarchitektur

Konzeptionell ist die DFA in Schichten organisiert. Sie ordnet ihre Komponenten der Datenschicht, Föderationsschicht und Diensteschicht zu und kann in jeder dieser Schichten jederzeit erweitert werden.

- Auf Ebene der *Datenschicht* werden Forschungsdaten gespeichert und verwaltet.
- Die Komponenten der *Föderationsschicht* haben eine nachweisende und deskriptive Funktion und bieten insbesondere Schnittstellen für die integrative Betrachtung von Forschungsdaten.
- Die *Diensteschicht* umfasst schließlich diejenigen Softwaresysteme, die auf Basis untergeordneter Schichten einen weiterführenden Nutzen für ihre Anwenderinnen und Anwender generieren.

## 2 Dokumentation

Für den Überblick über die architektonischen Zusammenhänge der DFA und um einen vereinfachten Zugang für Informationssuchende anzubieten wurde unter <https://dfa.de.dariah.eu/doc/> eine Übersichtsdokumentation (Abbildung 2) zur DFA eingerichtet.

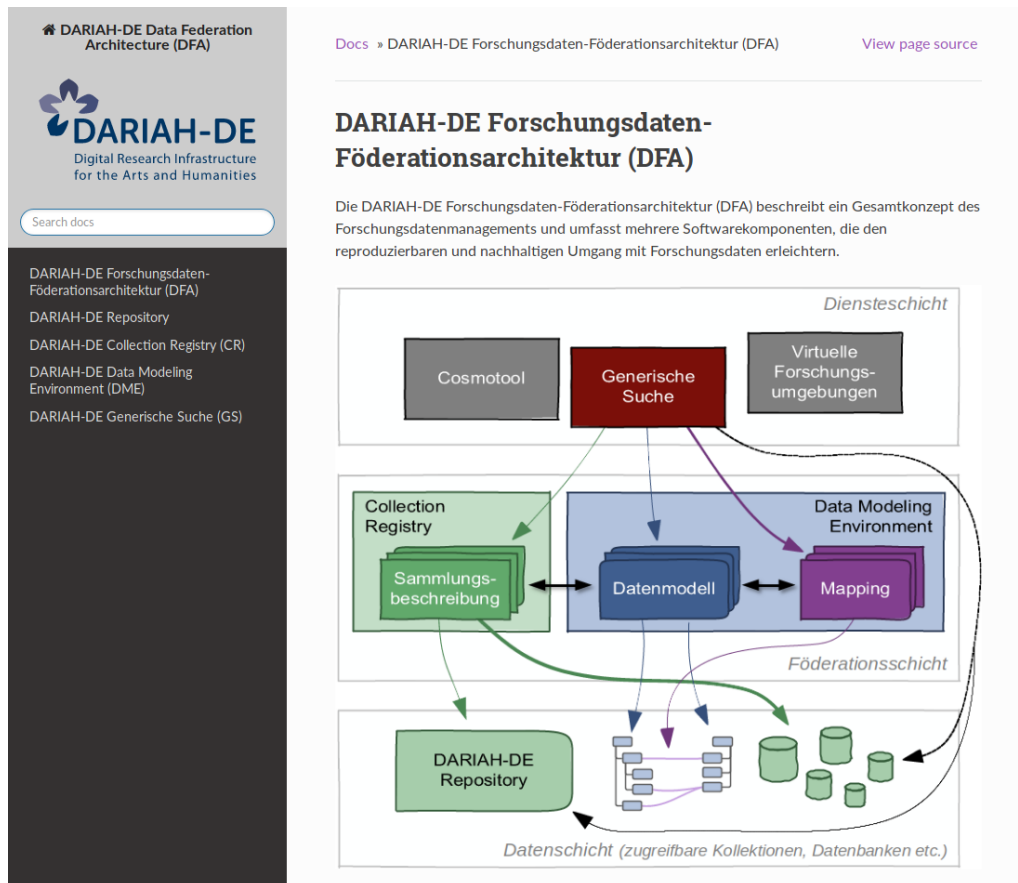


Abbildung 2: Übersichtssseite der Dokumentation

Eine inhaltliche Dokumentation der Komponenten findet an dieser Stelle nicht statt. Die Übersicht dient lediglich als logische Klammer um die Detaildokumentation der einzelnen Dienste. Für Anwender ist dies jedoch nicht ersichtlich, da die konsequente Verwendung derselben Dokumentationssoftware und entsprechender Konfiguration ein integriertes Bild vermittelt.

Die DFA Übersichtsdokumentation ist in einem eigenen GitLab Repository abgelegt: <https://gitlab.com/DARIAH-DE/dfa-docs>.



Eine GitLab CI/CD<sup>5</sup> Pipeline ist so konfiguriert, dass Änderungen in der Dokumentation in einem unmittelbaren Buildvorgang (vgl. Abbildung 3) münden, welcher überprüft, ob die Dokumentation erzeugt werden kann.

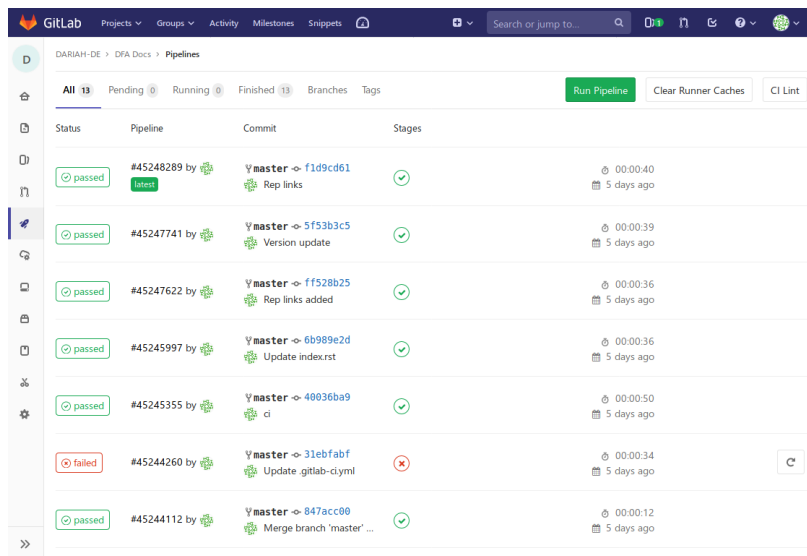


Abbildung 3: Ausgeführte GitLab CI/CD Pipelines für die DFA Dokumentation

Ein Installationspaket wird ggf. direkt auf einem APT Paketserver abgelegt (Abbildung 4), so dass dieses sehr einfach auf die entsprechenden Zielmaschinen verteilt werden kann.

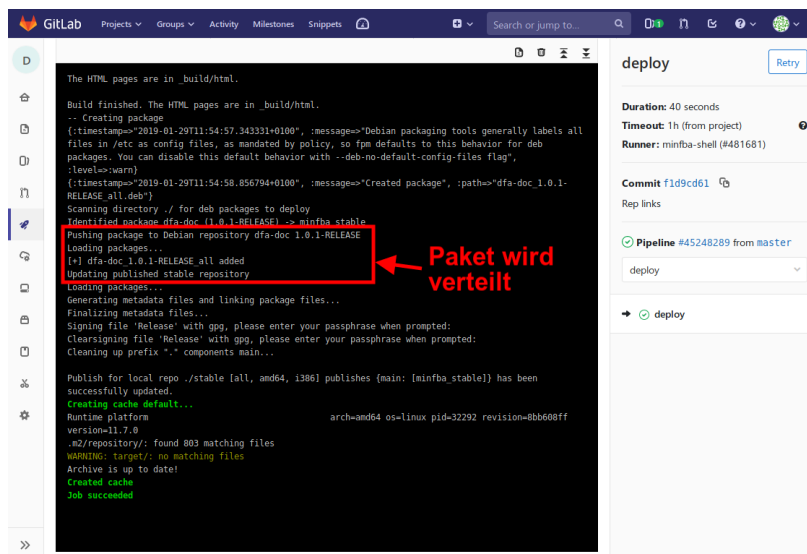


Abbildung 4: Ausführung des Deploy-Schrittes in der GitLab CI/CD Pipeline

<sup>5</sup>Continuous Integration / Continuous Deployment: <https://docs.gitlab.com/ee/ci/>

## 2.1 Repository

Produktivsystem: <https://repository.de.dariah.eu/>

Testinstanz: <https://dhrepworkshop.de.dariah.eu/>

**Dokumentation:** <https://repository.de.dariah.eu/doc>

Source Code: <https://repository.de.dariah.eu/doc/services/sourcecode.html>

Die Dokumentation des DARIAH-DE Repository umfasst derzeit den Nutzerguide, die Dokumentation der Service APIs und die Beschreibung der PID Auflösung auf Basis von DataCite DOI für Zitationen und EPIC Handle PIDs für Verwaltungszwecke.

Das Deployment des Repository erfolgt auf Basis von Jenkins als CI/CD Server und der Verteilung von Paketen auf eine Aptly Instanz: <https://ci.de.dariah.eu/aptly>. Gemeinsam mit den Softwarekomponenten wird auch die Dokumentation nach Änderungen erzeugt und verteilt.<sup>6</sup>

Neben der Online-Dokumentation steht für das DARIAH-DE Repository auch eine kompilierte Gesamtdokumentation in Form eines PDF Dokumentes zu Verfügung.

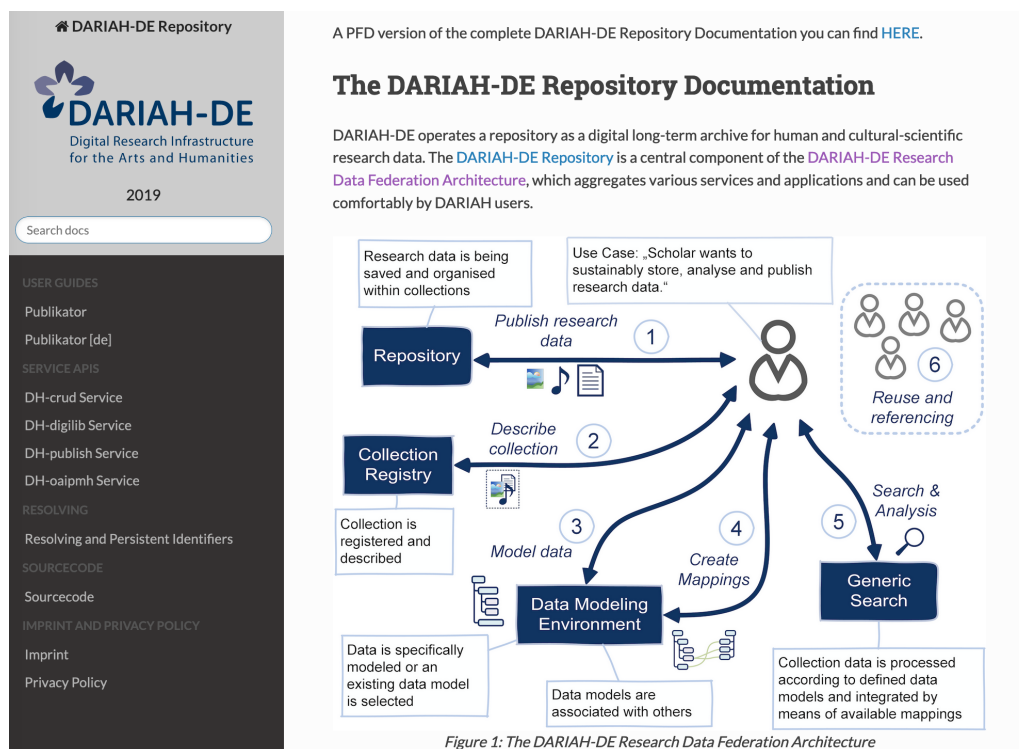


Abbildung 5: Dokumentation des DARIAH-DE Repository

<sup>6</sup><https://wiki.de.dariah.eu/display/DARIAH3/DARIAH-DE+Release+Management#DARIAH-DEReleaseManagement-ReleasenmitderAptlyGUI> (interne Wikiseite, Anmeldung erforderlich)

## 2.2 Collection Registry

---

Produktivsystem: <https://colreg.de.dariah.eu/>

Testinstanz: <https://dfatest.de.dariah.eu/colreg-ui>

**Dokumentation:** <https://dfa.de.dariah.eu/doc/colreg>

Source Code: <https://gitlab.com/DARIAH-DE/colreg>

---

Die Dokumentation der DARIAH-DE Collection Registry wurde in vergangenen und der aktuellen Projektphase an unterschiedlichen Stellen – vornehmlich Seiten des DARIAH-DE Wikis – sorgfältig gesammelt und aufbereitet.

Für die Zusammenstellung der Dokumentation zur Collection Registry wurden die unterschiedlichen Fragmente konsolidiert. Das Ergebnis besteht wie bei den übrigen DFA Komponenten aus einer Sphinx Dokumentation (Abbildung 6).

Die Dokumentation der Collection Registry wird mit Hilfe von GitLab CI/CD gemeinsam mit dem Anwendungscode erzeugt und per Apt-Repository verteilt.

The screenshot shows the documentation page for the DARIAH-DE Collection Registry. On the left is a dark sidebar with the DARIAH-DE logo and a search bar. The main content area has a white background with a blue header. The title is 'DARIAH-DE Collection Registry' and the section is 'Einführung'. The text describes the registry as a free tool for describing collections and mentions the DARIAH Collection Description Data Model (DCDDM). The sidebar contains a search bar and a list of navigation links: 'DARIAH-DE Collection Registry', 'Einführung', 'Die wissenschaftliche Sammlung', 'Der Nutzen der Collection Registry', '...für existierende Sammlungen, mit oder ohne technische Schnittstelle zu den (digitalen) Objekten', '...für Sammlungen, die noch nicht öffentlich verfügbar sind und die veröffentlicht werden sollen', 'Wo finde ich die Collection Registry?', 'Wie benutze ich die Collection Registry', 'Wie melde ich mich an?', 'Nach der Anmeldung', 'Das Anlegen einer neuen Sammlungsbeschreibung und das Bearbeiten von Entwürfen', 'Das Anlegen eines neuen Akteurs', 'Das Speichern einer neuen Sammlungsbeschreibung', 'Das Veröffentlichen einer Sammlungsbeschreibung', and 'Best Practice Empfehlungen'. The main content area has a blue header with the title 'DARIAH-DE Collection Registry' and a 'View page source' link. Below the title is the subtitle 'Die DARIAH-DE Collection Registry' and the section 'Einführung'. The text describes the registry as a free tool for describing collections and mentions the DARIAH Collection Description Data Model (DCDDM). The text also mentions the DARIAH-DE Data Federation Architecture and the DARIAH-DE Portal.

Abbildung 6: Dokumentation der DARIAH-DE Collection Registry

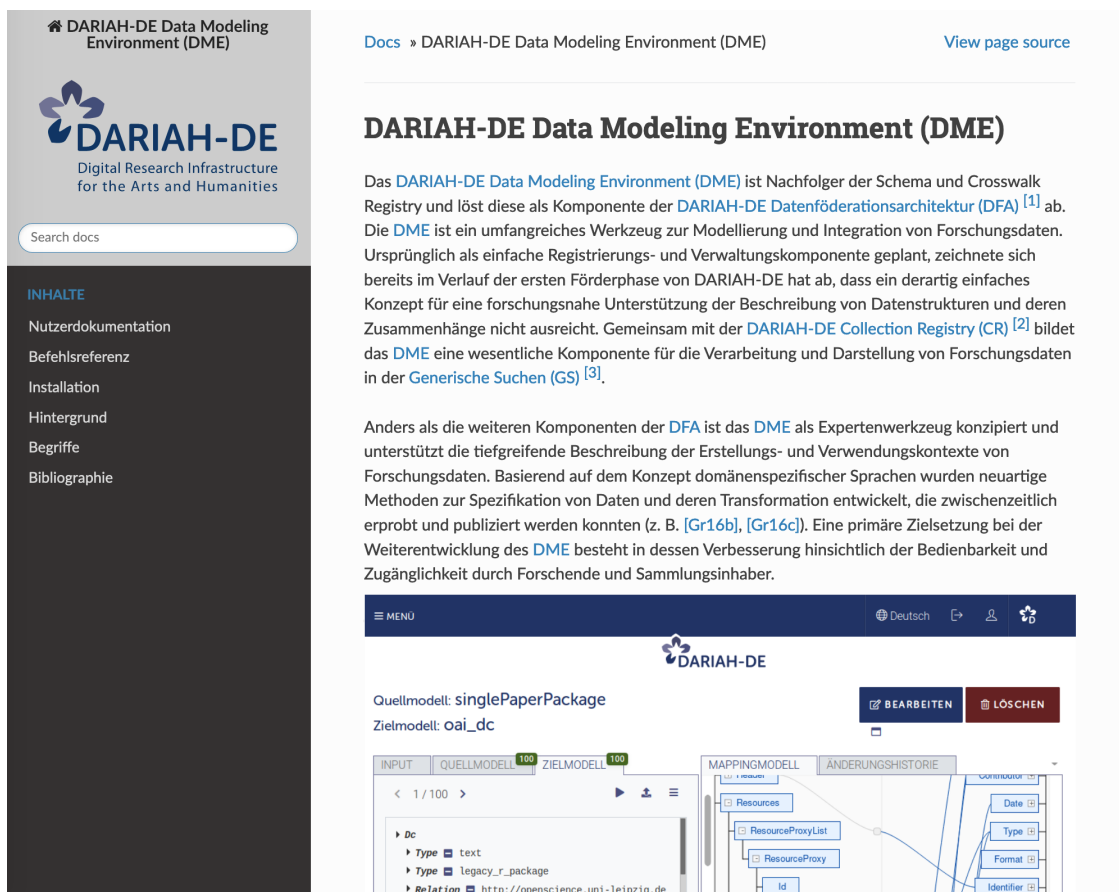
## 2.3 Data Modeling Environment

---

Produktivsystem:	<a href="https://dme.de.dariah.eu/">https://dme.de.dariah.eu/</a>
Testinstanz:	<a href="https://dfatest.de.dariah.eu/dme">https://dfatest.de.dariah.eu/dme</a>
<b>Dokumentation:</b>	<a href="https://dfa.de.dariah.eu/doc/dme">https://dfa.de.dariah.eu/doc/dme</a>
Source Code:	<a href="https://gitlab.rz.uni-bamberg.de/dariah/dme">https://gitlab.rz.uni-bamberg.de/dariah/dme</a>

---

Die Dokumentation des DME als Expertensystem stellt besondere Herausforderungen an die zu erstellende Dokumentation, da diese von einfachen Hilfestellungen für Erstanwender hin zu einer Befehlsreferenz von Transformationskommandos für fortgeschrittene Nutzer verschiedene Aspekte abdecken sollte. Die Dokumentation der DME befindet sich aus diesem Grund weiterhin im Aufbau. Modellierungsbeispiele und Best-Practice-Guidelines entstehen derzeit im Wesentlichen durch Rückkopplungen aus Machbarkeitsstudien, Workshops und nutzenden Projekten.



Das **DARIAH-DE Data Modeling Environment (DME)** ist Nachfolger der Schema und Crosswalk Registry und löst diese als Komponente der **DARIAH-DE Datenföderationsarchitektur (DFA)** <sup>[1]</sup> ab. Die **DME** ist ein umfangreiches Werkzeug zur Modellierung und Integration von Forschungsdaten. Ursprünglich als einfache Registrierungs- und Verwaltungskomponente geplant, zeichnete sich bereits im Verlauf der ersten Förderphase von DARIAH-DE hat ab, dass ein derartig einfaches Konzept für eine forschungsnaher Unterstützung der Beschreibung von Datenstrukturen und deren Zusammenhänge nicht ausreicht. Gemeinsam mit der **DARIAH-DE Collection Registry (CR)** <sup>[2]</sup> bildet das **DME** eine wesentliche Komponente für die Verarbeitung und Darstellung von Forschungsdaten in der **Generische Suchen (GS)** <sup>[3]</sup>.

Anders als die weiteren Komponenten der **DFA** ist das **DME** als Expertenwerkzeug konzipiert und unterstützt die tiefgreifende Beschreibung der Erstellungs- und Verwendungskontexte von Forschungsdaten. Basierend auf dem Konzept domänenspezifischer Sprachen wurden neuartige Methoden zur Spezifikation von Daten und deren Transformation entwickelt, die zwischenzeitlich erprobt und publiziert werden konnten (z. B. [\[Gr16b\]](#), [\[Gr16c\]](#)). Eine primäre Zielsetzung bei der Weiterentwicklung des **DME** besteht in dessen Verbesserung hinsichtlich der Bedienbarkeit und Zugänglichkeit durch Forschende und Sammlungsinhaber.

Abbildung 7: Dokumentation des DARIAH-DE Data Modeling Environment

## 2.4 Generische Suche

Produktivsystem: <https://search.de.dariah.eu/>

Testinstanz: <https://dfatest.de.dariah.eu/search>

**Dokumentation:** <https://dfa.de.dariah.eu/doc/search>

Source Code: <https://gitlab.rz.uni-bamberg.de/dariah/search>

Auch für die Generische Suche wird eine Dokumentation auf Basis von Sphinx erzeugt und verteilt. Die Dokumentation umfasst die Nutzerdokumentation sowie Hilfestellung zur Administration und Installation der Suche.

The screenshot shows the documentation page for the DARIAH-DE Generic Search. The top left corner features the DARIAH-DE logo and a search bar. The main navigation menu includes 'INHALTE', 'Nutzerdokumentation', 'Verwaltung', 'Installation', and 'Begriffe'. The page title is 'DARIAH-DE Generische Suche'. The content explains that the Generic Search (GS) is part of the DARIAH-DE Data Federations Architecture (DFA), enabling search in digital collections through the DARIAH-DE Collection Registry (CR). It notes that the search is free and without registration, and can be used for both broad and deep searches. The text further describes the search's adaptability, including facet-based searches and dynamic visualization techniques supported by the DARIAH-DE Data Modeling Environment (DME). At the bottom, there is a search interface with a search bar, a language dropdown set to 'Deutsch', and a word cloud of search terms.

Abbildung 8: Dokumentation der DARIAH-DE Generischen Suche

### 3 Ausblick

Die Dokumentation der DFA mit ihren Komponenten wurde in konsistentem Erscheinungsbild auf Basis von Sphinx installiert und steht Anwendern als Hilfestellung zur Verfügung.

Mit der Integration der Dokumentationsbausteine in den Sourcecode und die gemeinsame Verteilung über Continuous Integration und Deployment wird sichergestellt, dass für Entwickler möglichst geringe Hürden für die weiterführende Dokumentation ihrer Komponenten überwunden werden müssen. So kann bei weiterhin wachsendem Funktionsumfang der Komponenten eine ständige Aktualisierung und Erweiterung der Dokumentation erfolgen.

### 4 Quellenverzeichnis

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

<http://www.sphinx-doc.org/>

<http://docutils.sourceforge.net/rst.html>

[https://github.com/rtfd/sphinx\\_rtd\\_theme](https://github.com/rtfd/sphinx_rtd_theme)

Continuous Integration / Continuous Deployment: <https://docs.gitlab.com/ee/ci/>

<https://wiki.de.dariah.eu/display/DARIAH3/DARIAH-DE+Release+Management#DARIAH-DEReleaseManagement-ReleasenmitderAptlyGUI> (interne Wikiseite, Anmeldung erforderlich)