



# **Dokumentation der Forschungsdaten- Föderationsarchitektur (R 4.2.5)**

**Version 04.02.2019**

**Cluster 4**

**Verantwortlicher Partner MInf-BA**

## **DARIAH-DE Überführung der digitalen Forschungsinfrastrukturen für die e-Humanities in die Operational Phase (Betriebsphase)**

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1610A bis J, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



**Bundesministerium  
für Bildung  
und Forschung**

**Projekt:** DARIAH-DE: Überführung der digitalen Forschungsinfrastrukturen für die e-Humanities in die Operational Phase (Betriebsphase)

**BMBF Förderkennzeichen:** 01UG1610A bis J

**Laufzeit:** März 2016 bis Februar 2019

**Dokumentstatus:** Final

**Verfügbarkeit:** öffentlich

**Autoren:**

Tobias Gradl, MInf-BA

**Revisionsverlauf:**

Datum	Autor	Kommentare
04.02.2019	Tobias Gradl	
06.02.2019	Tobias Gradl	Einarbeitung von Rückmeldungen der Konsortialleitung



Dieses Werk ist unter einer Creative Commons Lizenz vom Typ Namensnennung 3.0 Deutschland zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie <http://creativecommons.org/licenses/by/3.0/de/> oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.

## Inhaltsverzeichnis:

<b>1. Einleitung</b> .....	<b>4</b>
<b>2. 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. Ausblick</b> .....	<b>12</b>
<b>4. Abbildungsverzeichnis</b> .....	<b>13</b>

# 1. Einleitung

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

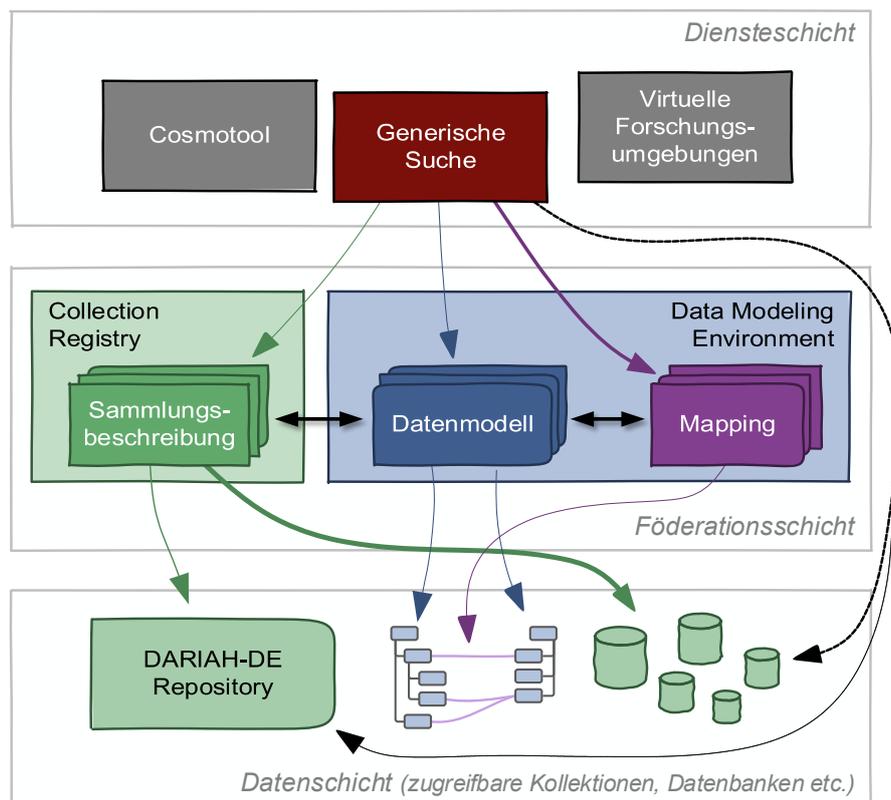
- Das **Repository** aggregiert verschiedene Dienste und Anwendungen und macht diese komfortabel nutzbar. Das Repository ermöglicht, Forschungsdaten nachhaltig und sicher zu speichern, sie mit Metadaten zu versehen und durch die Generische Suche auffindbar zu machen.
- Die **Collection Registry (CR)** ist ein Werkzeug zur Beschreibung von Sammlungen. Sammlungsbeschreibungen folgen dem umfangreichen, für kultur- und geisteswissenschaftliche Sammlungen entwickelten DARIAH Collection Description Data Model (DCDDM)<sup>1</sup>. Soweit vorhanden umfassen 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 umfangreiches 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 ggf. zu integrieren.

Dieser Bericht 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.

---

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

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.



**Abbildung 1: Komponenten und Beziehungen in der DARIAH-DE**

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

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

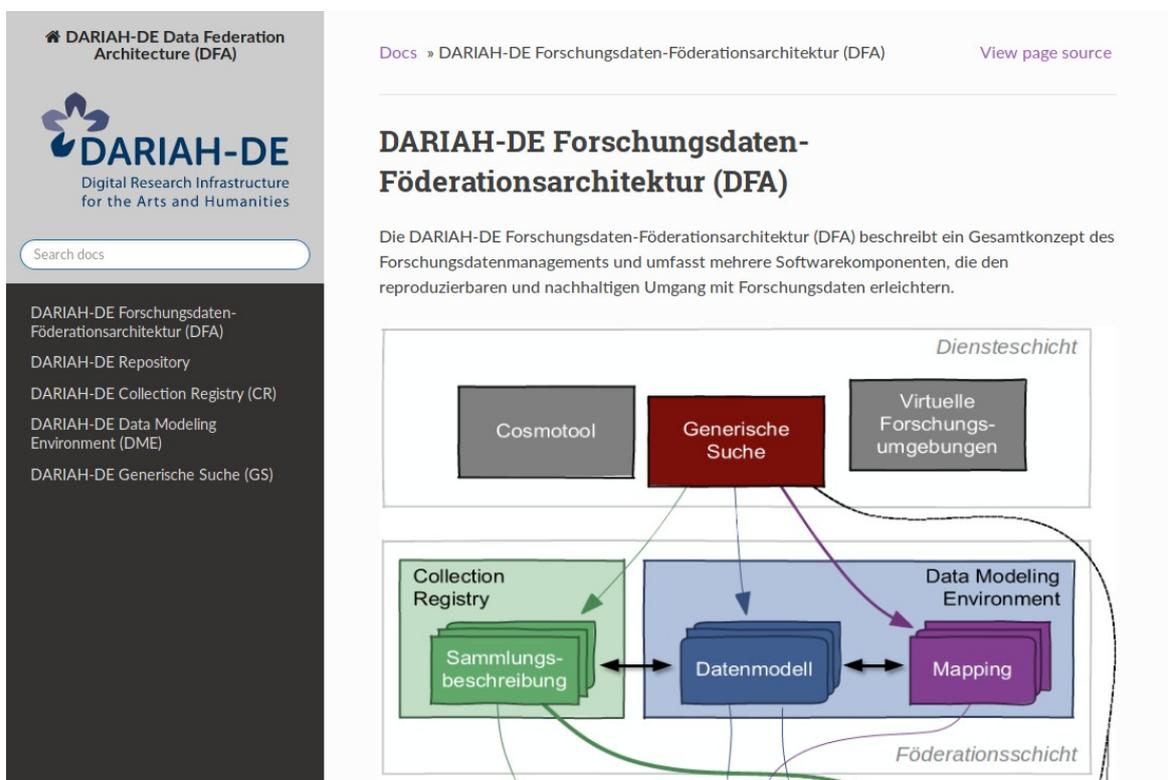
<sup>3</sup> <http://docutils.sourceforge.net/rst.html>

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

- 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 einen vereinfachten Zugang für Informationssuchende anzubieten wurde unter <https://dfa.de.dariah.eu/doc/> eine Übersichtsdokumentation (Abbildung 2) zur DFA eingerichtet.



**Abbildung 2: Übersichtseite der Dokumentation**

Eine inhaltliche Dokumentation der Komponenten findet an dieser Stelle nicht statt. Die Übersicht dient lediglich als logische Klammer um die Detaildokumentation der 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.

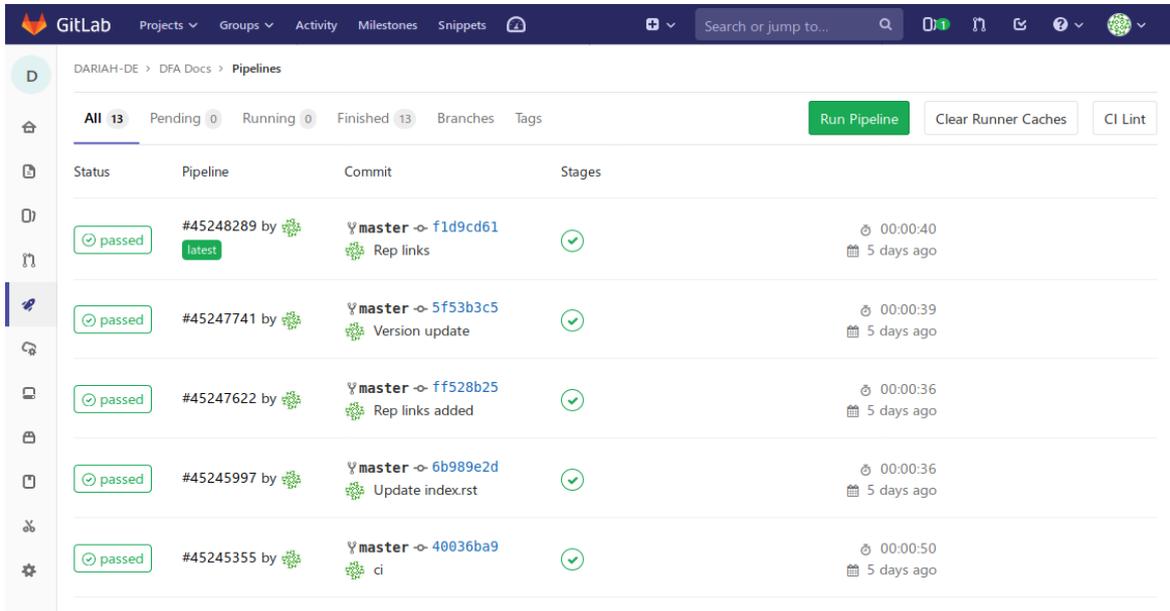


Abbildung 3: Ausgeführte GitLab CI/CD Pipelines für die DFA Dokumentation

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

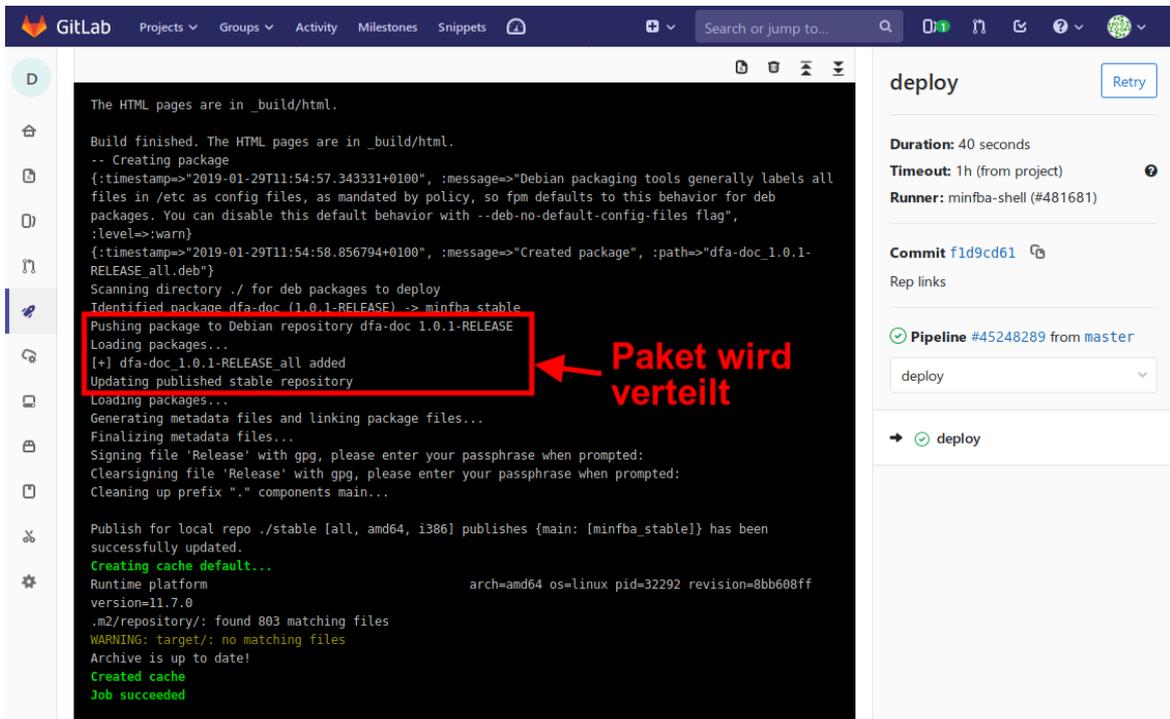


Abbildung 4: Ausführung des Deploy-Schrittes in der GitLab CI/CD Pipeline

5 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.

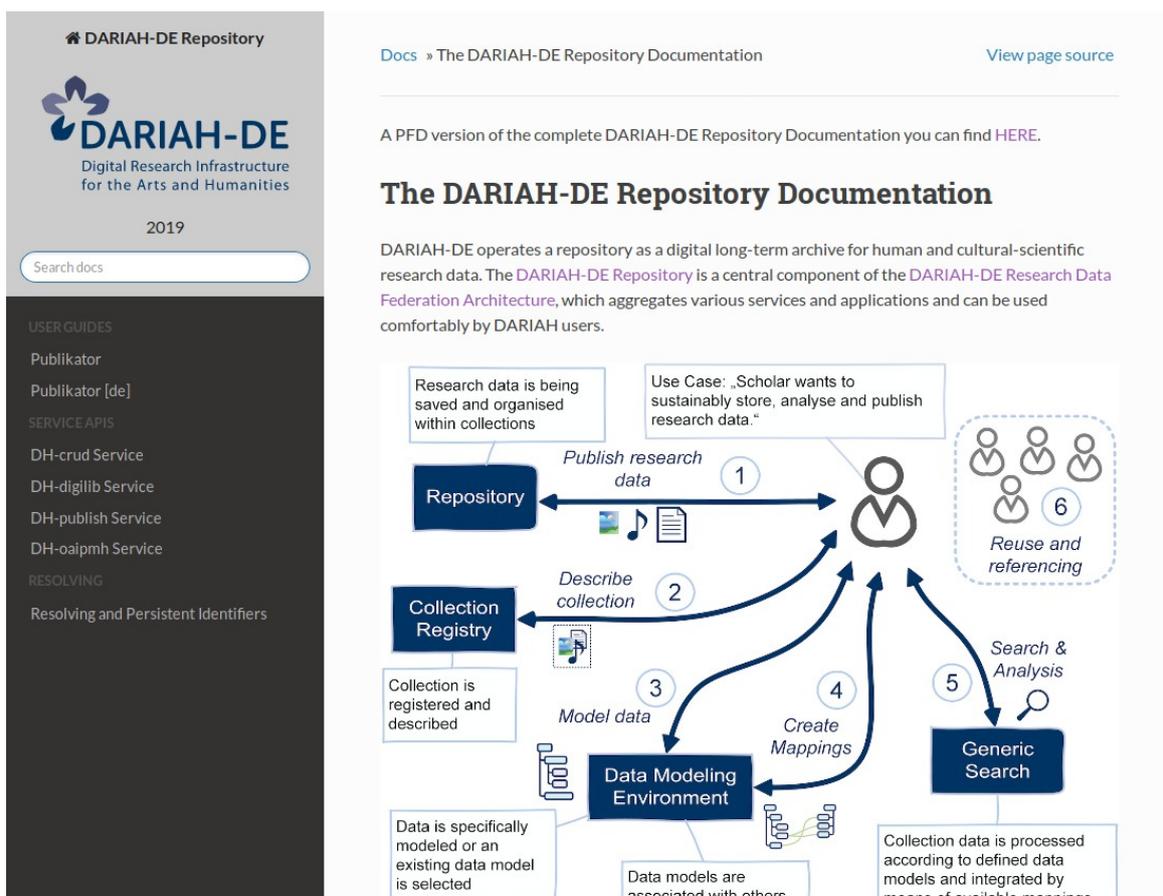


Abbildung 5: Dokumentation des DARIAH-DE Repository

<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 logo and a search bar. The main content area is light-colored and contains the following text:

**DARIAH-DE Collection Registry**

Docs » DARIAH-DE Collection Registry [View page source](#)

### DARIAH-DE Collection Registry

Die DARIAH-DE Collection Registry

#### Einführung

Die **DARIAH-DE Collection Registry** ist ein von DARIAH-DE bereitgestelltes, kostenfreies Tool, um Sammlungen zu beschreiben. Das bedeutet, dass an dieser Stelle Referenzen auf Sammlungsdaten beschrieben und gespeichert (und – soweit vorhanden – eine Zugriffsmethode auf die Daten angegeben) werden und nicht die Daten von Sammlungen selbst. Sie beschreiben also in der Collection Registry Ihre Sammlung – samt technischen Schnittstellen, und Sie können dafür auf ein weit umfangreicheres Beschreibungsschema (**DARIAH Collection Description Data Model – DCDDM**) zugreifen als dies bei der Veröffentlichung mit z. B. **Dublin Core Simple** möglich ist.

Die Collection Registry ist verankert in der DARIAH-DE **Data Federation Architecture** und spielt dort eine wichtige Rolle im Forschungsdatenzzyklus und im Forschungsdatenmanagement. Weiterführende Empfehlungen zum Forschungsdatenmanagement finden Sie auf dem **DARIAH-DE Portal**.

#### Die wissenschaftliche Sammlung

Als wissenschaftliche Sammlung werden in DARIAH-DE keine lediglich nach formalen Kriterien aufgebauten Digitalisat- oder Volltextsammlungen verstanden (siehe **Report zu Sammlungskonzepten aus DARIAH-DE II**), sondern Sammlungen, die nach inhaltlichen, forschungsfragenbezogenen oder fachlichen Kriterien gebildet wurden. Diese sind dabei möglicherweise aus formalen Sammlungen gebildet worden, können aber auch aus mehreren anderen solcher formalen Sammlungen zusammengefasst worden sein. In diesem Sinne sind formale Sammlungen Rohdaten und Grundlage für die Erstellung von wissenschaftlichen Sammlungen. Diese sind darüber hinaus Aggregationen von Forschungs- und Metadaten, die von WissenschaftlerInnen zur Beantwortung von Forschungsfragen innerhalb bestimmter Forschungskontexte zusammengestellt und mit weiteren Informationen wie Normdaten oder Annotationen maschinenlesbar angereichert werden.

Eine Sammlung wissenschaftlich relevanter Forschungsdaten (siehe **Definition nach Altendorf/Schmuck 2014**)

Abbildung 6: Dokumentation der DARIAH-DE Collection Registry

## 2.3. Data Modeling Environment

**Produktivsystem:** <https://dme.de.dariah.eu>

**Testinstanz:** <https://dfatest.de.dariah.eu/dme>

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

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

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.

DARIAH-DE Data Modeling Environment (DME)

DARIAH-DE  
Digital Research Infrastructure  
for the Arts and Humanities

Search docs

INHALTE

- Nutzerdokumentation
- Befehlsreferenz
- Installation
- Hintergrund
- Begriffe
- Bibliographie

Docs » DARIAH-DE Data Modeling Environment (DME) [View page source](#)

### DARIAH-DE Data Modeling Environment (DME)

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 forschungsnahe 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.

#### Wichtige Links

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

#### Inhalte

- Nutzerdokumentation
- Befehlsreferenz
- Installation
- Hintergrund
- Begriffe

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 page layout includes a dark sidebar on the left with a search bar and navigation links. The main content area has a header with the DARIAH-DE logo and the title 'DARIAH-DE Generische Suche'. Below the header is a paragraph describing the search system. A section titled 'Inhalte' lists the navigation items. At the bottom, there is a search interface with a search bar, a 'Suchen' button, and a word cloud visualization of search results.

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. Abbildungsverzeichnis

Abbildung 1: Komponenten und Beziehungen in der DARIAH-DE.....	5
Abbildung 2: Übersichtsseite der Dokumentation.....	6
Abbildung 3: Ausgeführte GitLab CI/CD Pipelines für die DFA Dokumentation.....	7
Abbildung 4: Ausführung des Deploy-Schrittes in der GitLab CI/CD Pipeline.....	7
Abbildung 5: Dokumentation des DARIAH-DE Repository.....	8
Abbildung 6: Dokumentation der DARIAH-DE Collection Registry.....	9
Abbildung 7: Dokumentation des DARIAH-DE Data Modeling Environment.....	10
Abbildung 8: Dokumentation der DARIAH-DE Generischen Suche.....	11