



Weiterentwicklung der Collection Registry – Schnittstellen (M 4.2.2.)

**Version 1.0
Cluster 4
Verantwortlicher Partner MInfBa, SUB**

DARIAH-DE Aufbau von Forschungsinfrastrukturen für die e-Humanities

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1110A bis N, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



**Bundesministerium
für Bildung
und Forschung**

Projekt: DARIAH-DE: Aufbau von Forschungsinfrastrukturen für die e-Humanities

BMBF Förderkennzeichen: 01UG1110A bis N

Laufzeit: März 2011 bis Februar 2016

Dokumentstatus: Final

Verfügbarkeit: Öffentlich

Autoren:

Stefan E. Funk, SUB Göttingen
Tobias Gradl, MInf Uni Bamberg
Stefan Schmunk, SUB Göttingen

Revisionsverlauf:

Datum	Autor	Kommentare
11. Januar 2016	Schmunk	Gliederung und Draft Fassung
18. Januar 2016	Funk & Gradl	Überarbeitung
19. Januar 2016	Funk, Gradl, Schmunk	Finalisierung

Inhaltsverzeichnis

1. Einleitung.....	4
2. Aktueller Stand der Collection Registry.....	5
3. Literaturverzeichnis	10
4. Abbildungsverzeichnis.....	11

1. Einleitung

Im Rahmen der zweiten Förderphase von DARIAH-DE wurde die Collection Registry (CR) weiterentwickelt. In der ersten Phase von 2011 bis 2014 lag der Fokus auf der Konzeption und einer ersten Implementierung eines Prototypen. Dieser Prototyp wurde zu Beginn der zweiten Förderphase erneuten umfangreichen Tests durch fachwissenschaftliche NutzerInnen unterzogen und auf dieser Basis ein Relaunch umgesetzt, der insbesondere auf eine Steigerung der Usability und einer Verbesserung der graphischen Nutzeroberfläche und der damit verbundenen Nutzerführung und -interaktion abzielte. Besonderes Augenmerk wurde in diesem Kontext auf zwei Aspekte gelegt. Einerseits wurden in den Entwicklungsprozess der CR FachwissenschaftlerInnen unmittelbar einbezogen, um unmittelbar Funktionalitätstest durchführen zu können, so dass die benötigte Entwicklungszeit verkürzt werden konnte und andererseits wurde ein Fokus darauf gelegt, den verwendeten Softwarestack zu anderen in DARIAH-DE entwickelten und betriebenen Komponenten anzugleichen. Aus diesem Grund wurde die Collection Registry auf Spring, Bootstrap und Java umgestellt – Entwicklungsarbeiten, die vor allem deshalb umgesetzt wurden, um die Administration und Pflege mehrerer Komponenten zu vereinfachen und auch einen Dauerbetrieb nachhaltig umsetzen zu können. Weiterhin wurde die Collection Registry an das angepasste zugrunde liegende Datenmodell DCDDM¹ angepasst, das das vorherige DCLAP² ersetzt.

An dieser Stelle sei insbesondere Christoph Plutte gedankt, der die Collection Registry konzipierte und der die Entwicklungsarbeiten von 2011 bis Frühjahr 2015 durchführte, ebenfalls seinem Nachfolger Marco Jürgens – beide Berlin-Brandenburgische Akademie der Wissenschaften –, der in der zweiten Jahreshälfte die Umstellung auf Spring und Bootstrap umsetzte. Die von den beiden Kollegen begonnenen Arbeiten wurden von Tobias Gradl (Uni Bamberg) fertiggestellt, der die Weiterentwicklung ab Januar 2016 übernahm, nachdem die Projektmittel der BBAW im Rahmen der Cluster 4 Tätigkeiten Ende 2015 ausgelaufen waren.

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

² https://wiki.de.dariah.eu/download/attachments/14651583/M1.2.2_Collection_Registry_incl_DCLAP.pdf.

2. Aktueller Stand der Collection Registry

Die Collection Registry ist ein zentrales Online-Verzeichnis, in dem Beschreibungen von Daten- und Forschungsdatensammlungen registriert und beschrieben werden können. Die in der Collection Registry abgelegten Sammlungsbeschreibungen stehen sowohl in maschinen- als auch menschenlesbarer Form zur Verfügung. Ziel der Collection Registry ist es, einen Service anzubieten, in dem geistes- und kulturwissenschaftliche Sammlungsbeschreibungen hinterlegt und zugleich von WissenschaftlerInnen nachgenutzt werden können. Hierbei stehen vor allem folgende vier Szenarien im Mittelpunkt:

1. WissenschaftlerInnen und Forschungseinrichtungen können eigene Sammlungsbeschreibungen von digitalen Forschungsdaten in der Collection Registry hinterlegen bzw. angeben, so dass diese an einer zentralen Stelle von Dritten gefunden werden können. Die Collection Registry hat hierbei das Ziel, als zentrales Nachweisinstrument zu dienen, in dem Sammlungsbeschreibungen hinterlegt werden können, und auch die Sammlungen selbst können – sofern es sich um digitale Daten handelt und deren Schnittstellen angegeben werden –, in der Generischen Suche indexiert und damit auch durchsucht werden.
2. Neben diesem Standard-Use-Case können WissenschaftlerInnen und Einrichtungen die Collection Registry auch als Werkzeug nutzen, um Sammlungsbeschreibungen von Dritten, also zum Beispiel von Gedächtniseinrichtungen oder anderen Einrichtungen anzugeben. Sofern die entsprechenden (offenen) Schnittstellen bei der Datenaufnahme ebenfalls angegeben werden, werden diese Daten ebenfalls in der Generischen Suche indexiert und damit durchsuchbar.
3. In der Forschungsrealität in den Geistes- und Kulturwissenschaft ist es, trotz der Etablierung der Nutzung von digitalen Werkzeugen zur Analysen von Daten, Praxis – und wird dies auch auf Dauer bleiben –, mit analogen Daten und Quellen bzw. mit digital-analogen Sammlungen zu arbeiten und diese als Grundlage der Forschung zu verwenden. In diesem Kontext entstehen eine Vielzahl von wissenschaftlich interessanten und relevanten Sammlungsbeschreibungen, die trotz ihrer desperate medialen Heterogenität durch eine Verzeichnung in der Collection Registry öffentlich bekannt gemacht werden können und somit auch nutzbar sind.

Im Rahmen des Milestones wurde das User Interface der Collection Registry grundlegend überarbeitet und zugleich das genutzte technologische Framework an das der Generischen Suche und der Schema Registry angeglichen.

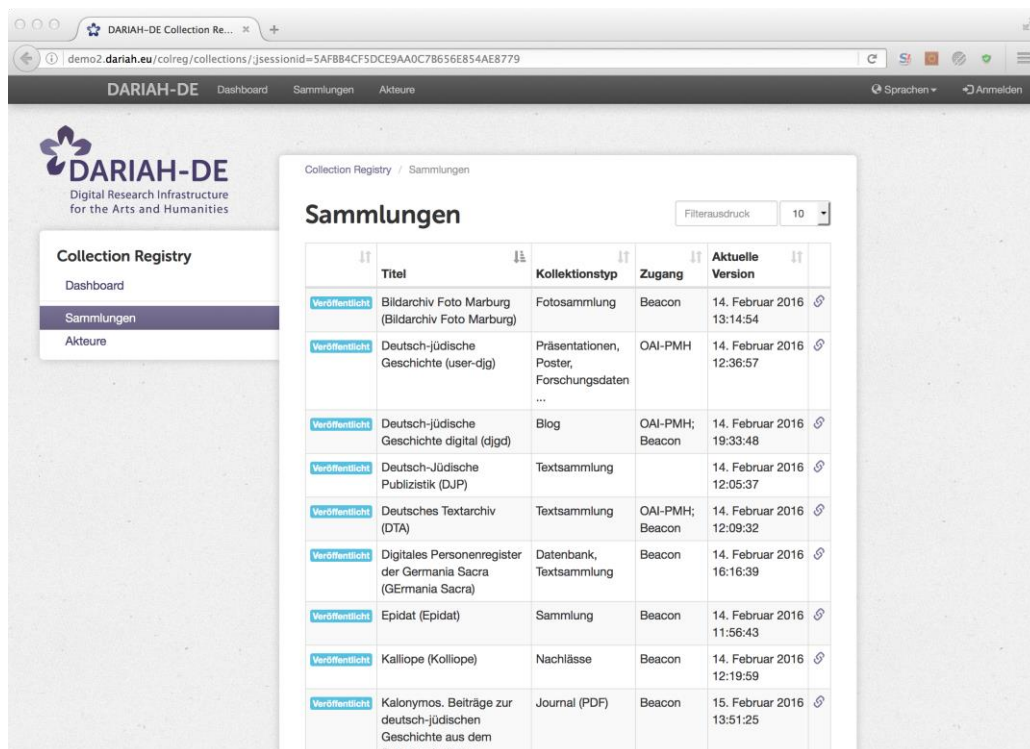


Abbildung 1: Neues Look and Feel der Collection Registry und Integration in das Corporate Design von DARIAH-DE.

Im Zuge der Entwicklungsarbeiten der zweiten Förderphase von DARIAH-DE wurde die Collection Registry auf Spring, Bootstrap und Java umgestellt. Auf diese Weise kann die Pflege und Administration mit den anderen Werkzeugen abgestimmt werden. Ebenfalls wurde im Rahmen des neuen Releases das User Interface überarbeitet und das Corporate Design von DARIAH-DE³ implementiert.

³ <https://github.com/tgradl/colreg/tree/master/colreg-ui/src/main/webapp/resources>.

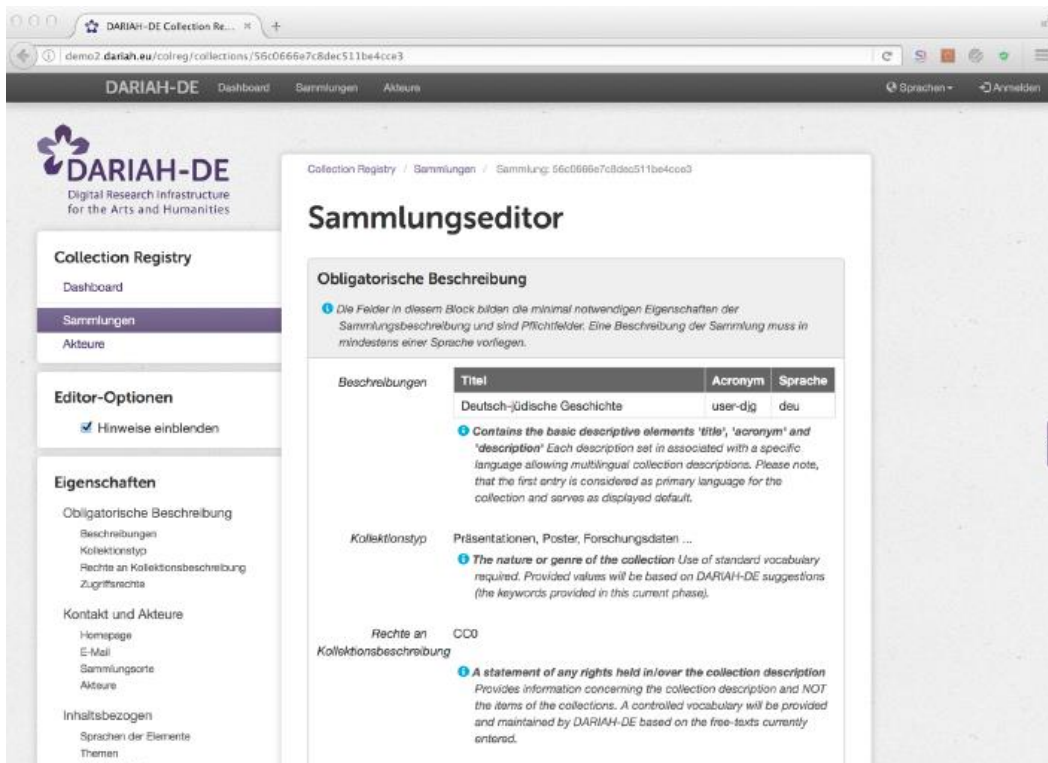


Abbildung 2: Sammlungseditor der Collection Registry, bei der neben deskriptiven Aspekten auch entsprechende Rechte an den Kollektionsbeschreibungen hinterlegt werden können.

Darüber hinaus lag die Hauptaufgabe in der Integration der Collection Registry in das Zusammenspiel zwischen Generische Suche, Schema Registry und dem DARIAH-DE-Repository, das in DARIAH-DE als Forschungsdaten-Föderationsarchitektur bezeichnet wird. Hierfür mussten Schnittstellen entwickelt werden, die im Folgenden beschrieben werden.

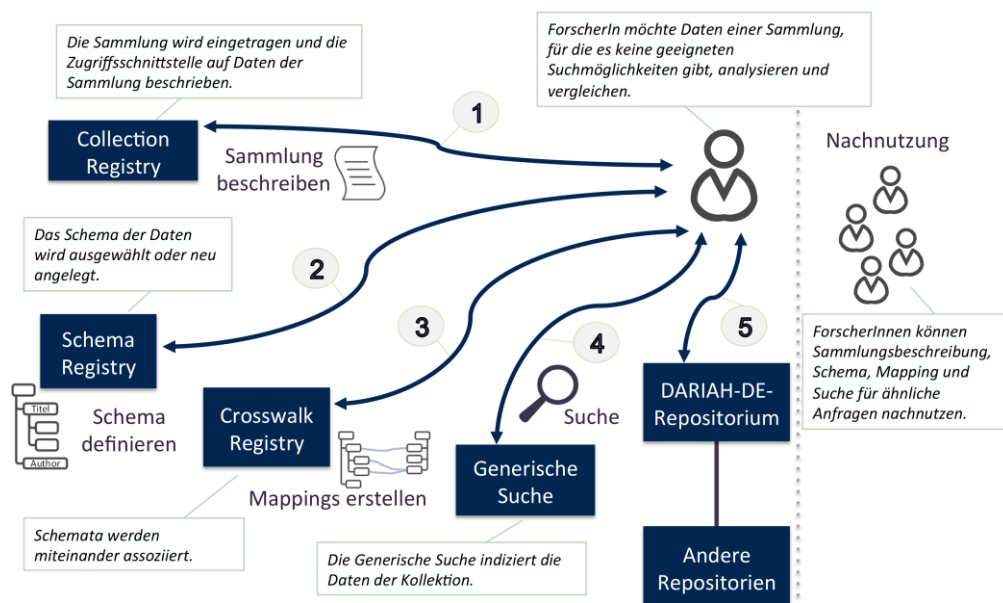


Abbildung 3: Die DARIAH-DE Forschungsdaten-Föderationsarchitektur.

Die DARIAH-DE Forschungsdaten-Föderationsarchitektur ist die Bezeichnung für

Dienste und Werkzeuge, mit deren Hilfe Forschungsdaten und Sammlungsbeschreibungen aus unterschiedlichen Quellen, wie beispielsweise von Kulturinstitutionen, Bibliotheken, Archiven, Forschungseinrichtungen und Rechenzentren registriert, modelliert und für Analysen verwendet werden können.⁴ Die Unterstützung von Informationsbedürfnissen in einem wissenschaftlichen Kontext erfordert einen hohen Grad an Flexibilität und Nachvollziehbarkeit von Suchwerkzeugen: so sollte es ForscherInnen möglich sein, ihre wissenschaftliche Recherche im digitalen Umfeld auf bestimmte Quellen zu beschränken und insbesondere nachvollziehen zu können, auf welche Art und Weise das Ranking der Suchergebnisse erfolgt. Bei föderierten Suchen besteht idealerweise zudem die Möglichkeit, iterativ auf die Verarbeitung und Zusammenführung von Daten einzuwirken und diese an spezifische Bedürfnisse anzupassen. Auf Ebene der bibliographisch erschlossenen Datensätze und Sammlungsbeschreibungen können auf diese Weise Sammlungen unterschiedlicher Provenienz durchsucht und dadurch die Interoperabilität heterogener Daten gewährleistet werden. Heterogene Daten und Metadatenquellen (so beispielsweise bezogen auf Daten- und Dateiformate, Medienformate etc.) können so trotz struktureller Differenziertheit von FachwissenschaftlerInnen über einen Zugriff genutzt werden.

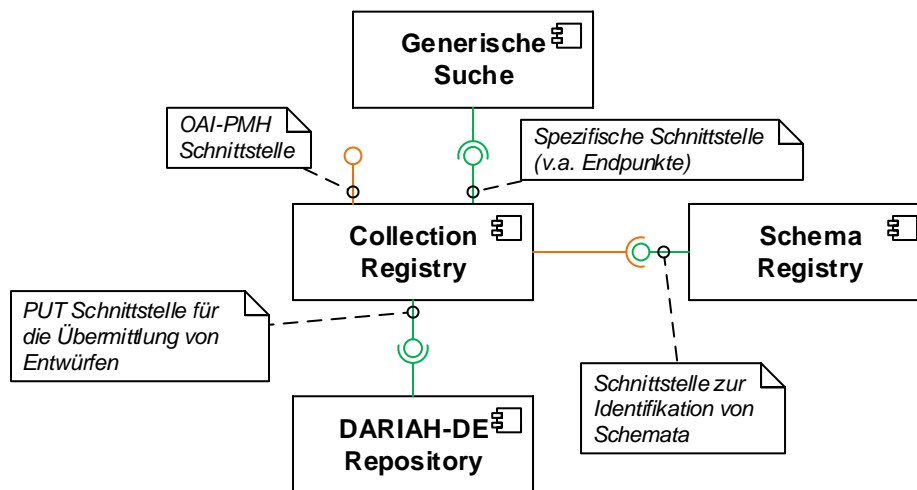


Abbildung 4: Zusammenhänge zwischen den Komponenten

Die Collection Registry steht als integraler Bestandteil dieser Föderationsarchitektur in einem technischen und funktionalen Zusammenhang mit weiteren DARIAH-DE Diensten. Das Diagramm in 4 skizziert die Schnittstellen und das Zusammenwirken wichtiger Komponenten der Föderationsarchitektur – die Farbe der Schnittstellen und Sockets visualisiert dabei den Stand der Implementierung (grün für „produktiv“, orange für „in Entwicklung“).

Die Collection Registry realisiert derzeit drei Schnittstellen:⁵

- Eine spezifische Schnittstelle für die generische Suche wurde realisiert, um nur die Informationen zu Sammlungen und technischen Zugriffsmöglichkeiten bereitzustellen, die für die generische Suche relevant sind. Aufgrund der per-

⁴ Siehe z. B. Tobias Gradl, Andreas Henrich, Christoph Plutte (2015): „Heterogene Daten in den Digital Humanities: Eine Architektur zur forschungsorientierten Föderation von Kollektionen. In: Grenzen und Möglichkeiten der Digital Humanities“. Hrsg. von Constanze Baum & Thomas Stäcker. 2015 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1). text/html Format. DOI: 10.17175/sb001_020.

⁵ Siehe <https://github.com/tgradl/colreg/tree/master/colreg-ui/src/main/java/eu/dariah/de/colreg/controller/api>.

manenten Kommunikation zwischen diesen beiden Komponenten kann Traffic und Rechenaufwand reduziert werden

- Eine schreibende (PUT) Schnittstelle wurde realisiert, um eine automatisierte Übernahme von Entwürfen aus dem DARIAH-DE Repository einzurichten und so die Bedienbarkeit der Komponenten zu verbessern
- Die Umsetzung einer OAI-PMH⁶ Schnittstelle für den generischen und offenen Zugriff auf sämtliche Aspekte publizierter Sammlungsbeschreibung wird derzeit noch entwickelt

Zur Identifikation und Assoziierung von Kollektionen und den darin verwendeten Datenmodellen konsumiert⁷ die Collection Registry zudem eine Schnittstelle der Schema Registry. Erst auf Basis dieser Assoziation zwischen Kollektion und Schema werden weiterführende Dienste (wie z. B. die generische Suche) in die Lage versetzt, Daten aus verfügbaren Kollektionen nach den Spezifikationen von Domänenexperten zu verarbeiten, indexieren und verfügbar zu machen.⁸

⁶ Siehe Carl Lagoze, Herbert Van de Sompel, Michael Nelson, Warner Simeon (2002): "The Open Archives Initiative Protocol for Metadata Harvesting". Protocol Version 2.0, June 2002.

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

⁷ Siehe <https://github.com/tgradl/colreg/tree/master/colreg-ui/src/main/java/eu/dariah/de/colreg/service>.

⁸ Siehe z. B. Tobias Gradl, Andreas Henrich (2016): "Die DARIAH-DE-Föderationsarchitektur – Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen". In: Bibliothek Forschung und Praxis. Hrsg. v. Achim Bonte u.a. Band 40, Heft 2, Seiten 222–228, ISSN (Online) 1865-7648, ISSN (Print) 0341-4183, DOI: 10.1515/bfp-2016-0027, July 2016.

3. Literaturverzeichnis

Gradl, Tobias & Henrich, Andreas (2016): "Die DARIAH-DE-Föderationsarchitektur – Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen", S. 222–228. In: Bibliothek Forschung und Praxis. Hrsg. v. Bonte, Achim u.a. Band 40, Heft 2, ISSN (Online) 1865-7648, ISSN (Print) 0341-4183, DOI: 10.1515/bfp-2016-0027, July 2016.

Gradl, Tobias; Henrich, Andreas; Plutte, Christoph (2015): "Heterogene Daten in den Digital Humanities: Eine Architektur zur forschungsorientierten Förderung von Kollektionen". In: Grenzen und Möglichkeiten der Digital Humanities. Hrsg. von Baum, Constanze Baum & Stäcker, Thomas. 2015 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1). text/html Format. DOI: 10.17175/sb001_020

Lagoze, Carl; Van de Sompel, Herbert; Nelson, Michael; Warner, Simeon (2002): "The Open Archives Initiative Protocol for Metadata Harvesting". Protocol Version 2.0. <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

Plutte, Christoph & Harms, Patrick (2012): "DARIAH-DE Aufbau von Forschungsinfrastrukturen für die e-Humanities". Milestone.
https://wiki.de.dariah.eu/download/attachments/14651583/M1.2.2_Collection_Registry_incl_DCLAP.pdf.

Verwendete Webseiten

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

<https://github.com/tgradl/colreg/tree/master/colreg-ui/src/main/webapp/resources>

<https://github.com/tgradl/colreg/tree/master/colreg-ui/src/main/java/eu/dariah/de/colreg/controller/api>

<https://github.com/tgradl/colreg/tree/master/colreg-ui/src/main/java/eu/dariah/de/colreg/service>

4. Abbildungsverzeichnis

Abbildung 1: Neues Look and Feel der Collection Registry und Integration in das Corporate Design von DARIAH-DE.....	6
Abbildung 2: Sammlungseditor der Collection Registry, bei der neben deskriptiven Aspekten auch entsprechende Rechte an den Kollektionsbeschreibungen hinterlegt werden können.	7
Abbildung 3: Die DARIAH-DE Forschungsdaten-Föderationsarchitektur.	7
Abbildung 4: Zusammenhänge zwischen den Komponenten	8