



DARIAH-DE Repositorium (M4.3.3.3)

Version 30.11.2018
Cluster 4
Verantwortlicher Partner SUB Göttingen

Weiterentwickeltes Produktivsystem des DARIAH-DE Repositoriums

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.



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: Raisa Barthauer, Stefan E. Funk

Revisionsverlauf:

Datum	Autor	Kommentare
23.10.2018	Raisa Barthauer	Entwurf
30.11.2018	Stefan E. Funk	Finalisierung



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

1. Einleitung	4
2. Funktionale Weiterentwicklung des DARIAH-DE Repositorys	4
Einbindung von Metriken für ein Ressourcen-Monitoring	5
Zusammenführung der Puppet-Konfiguration in ein dhrep-Puppet-Modul	5
Portierung der Anwendungen Digilib, Wildfly und Mirador	6
Erweitern der Index- und Landing-Pages	7
Anpassen der HTML-Seiten des Repositorys an den neuen StyleGuide	9
Optimierung des Publikationsworkflows und angehängter Dienste	10
Erweiterung und Aktualisierung der API-Dokumentation und des User-Guides	10
Optimieren der Download-Funktion von DH-crud für Web-Streaming	11
RDAWG Research Data Repository Interoperability BagPack	11
3. Zusammenfassung	12

1. Einleitung

Forschungsdaten und digitale Werkzeuge und Dienste zu ihrer Nutzung sind wesentliche Grundlagen der geistes- und kulturwissenschaftlichen Forschung. Durch die Verknüpfung, Archivierung und Nachnutzung dieser Daten wird nicht allein eine vorher unbekannte Forschungsdynamik möglich, sondern es können auch völlig neue Forschungsfragen bearbeitet werden. DARIAH-DE als digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften trägt dieser Prämisse mit seinem Angebot Rechnung, indem es zum einen Forschungsdaten verfügbar macht, zum anderen digitale Werkzeuge zum Forschungsdatenmanagement entwickelt und bereitstellt.

Das DARIAH-DE Repository¹ ist in diesem Rahmen eine zentrale Komponente der DARIAH-DE Forschungsdaten-Förderungsarchitektur, die verschiedene Dienste und Anwendungen zusammenträgt und komfortabel nutzbar macht. Das DARIAH-DE Repository stellt ein digitales Langzeitarchiv für die Forschungsdaten von Geistes- und Kulturwissenschaftlern bereit. Die Daten können mit Hilfe des DARIAH-DE Publikators² eingespielt und mit Metadaten versehen werden. Dank der zugewiesenen Persistenten Identifikatoren sind die Daten nachhaltig referenzierbar, dank der Generischen Suche bleiben sie auffindbar.

Um das DARIAH-DE-Produktivsystem zu fördern, besser nutzbar zu machen und weiter voranzubringen, wurden und werden von DARIAH-DE zahlreiche Anstrengungen unternommen. Für die Weiterentwicklung des Produktivsystems wurden neue Dienste eingebunden, erweitert und nutzbar gemacht, Workflows optimiert und Systeme schneller und zentraler konfigurierbar gemacht.

Das vorliegende Dokument dient als Nachweis des zum 1.10.2018 im Rahmen von Cluster 4 „Wissenschaftliche Sammlungen“ erbrachten Meilensteins „M4.3.3.3. Weiterentwickeltes Produktivsystem des DARIAH-DE Repositoriums“. Es fasst die einzelnen Bestandteile der erfolgten Weiterentwicklung kurz zusammen. Die Dienste des DARIAH-DE Repository sind über folgende URL erreichbar: <https://de.dariah.eu/repository>

2. Funktionale Weiterentwicklung des DARIAH-DE Repositoriums

Bis November 2018 wurden mehrere entscheidende Ziele umgesetzt, die das DARIAH-DE Repository nachhaltig sowohl in der Umsetzung als auch in der Anwendung verbessern werden, die im Folgenden kurz beschrieben werden.

¹ DARIAH-DE Portal. DARIAH-DE Repository. <https://de.dariah.eu/repository>

² DARIAH-DE Publikator. <https://repository.de.dariah.eu/publikator>

Einbindung von Metriken für ein Ressourcen-Monitoring

Gemeinsam mit den in M 4.3.3.2 (TextGrid-Produktivsystem)³ beschriebenen Maßnahmen wurde das Metrics-Monitoring⁴ auch für DARIAH-DE Repository-Dienste eingerichtet und stetig erweitert. Quantitative Aussagen über die Ressourcennutzung und die Auslastung der einzelnen Dienste können hier beobachtet und auch für einen gewissen Zeitraum analysiert werden. Das ist sehr hilfreich für die Weiterentwicklung der Dienste und zum Beispiel auch für das Feintuning der Speicherkonfiguration von beispielsweise den Tomcat Web-Application-Server.

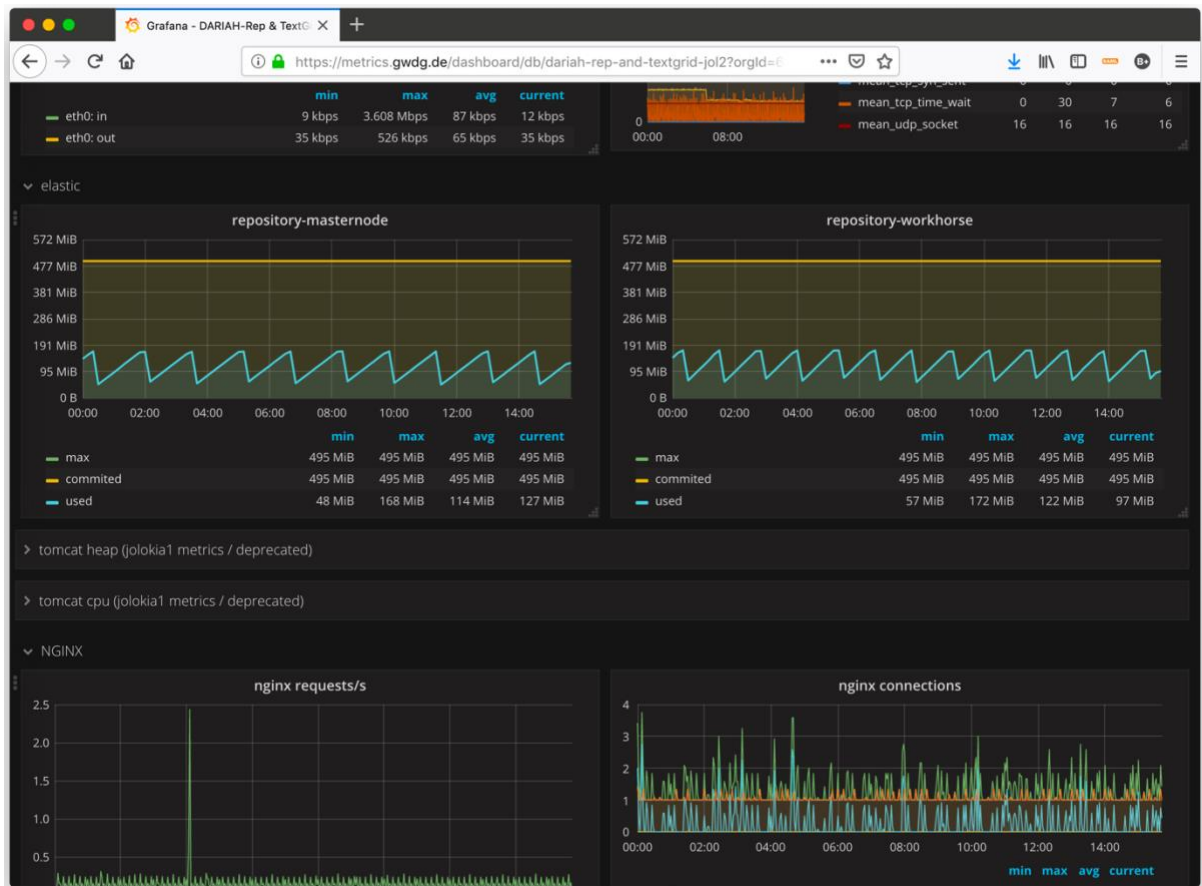


Abbildung 1: Ressourcenanalyse per Grafana

Zusammenführung der Puppet-Konfiguration in ein dhrep-Puppet-Modul

Auch im DARIAH-DE Repository wurde die zentrale Betreuung durch das Konfigurationstool Puppet erweitert und mit der Konfiguration des TextGrid Repositories zusammengeführt. Zu diesem Zweck wurde ein Puppet-Modul „dhrep“⁵ erstellt, mit dem mittels virtueller Maschinen und der Puppet-Konfiguration jeweils einfach TextGrid- und DARIAH-DE Repository-Server erstellt und hochgefahren werden können. Gemeinsam mit dem DARIAH-DE Puppetmodul⁶

³ Vgl. M4.3.3.2. TextGrid-Produktivsystem – Weiterentwickeltes Produktivsystem des TextGrid Repository.

<https://wiki.de.dariah.eu/display/publicde/Reports+and+Milestones?preview=%2F14651583%2F64974692%2FM+4.3.3.2+TextGrid-Produktivsystem.pdf>

ist es nun ebenfalls möglich, die Test, Entwicklungs- und Produktivinstanzen des DARIAH-DE Repositorys zentral und automatisiert zu konfigurieren und zu aktualisieren.

Dies betrifft ebenso die technische Realisation der Verbesserung des Release-Managements unter Aspekten der Continuous Integration (CI). Das Build- und Deployment-Management wurde größtenteils auf Jenkinsfiles umgestellt. In den meisten Fällen wird automatisiert nach dem Einchecken des Quellcodes der Buildprozess angestoßen und Debian.Pakete (DEB) erstellt, die dann bei jedem Puppet-Lauf automatisch deployed werden, abhängig von der Versionsnummer jeweils auf den Produktiv- bzw. Entwicklungsservern. Auch wurde der Build-Prozess größtenteils auf einen einheitlichen Gitflow-Workflow umgestellt, der die Entwicklung besser kanalisiert und Neuentwicklungen, Hotfixes und neue Features übersichtlicher gestaltet.

Die Puppet-Konfiguration der DARIAH-DE Repository- und TextGridRep-Server wurde in ein dhrep-Puppet-Modul zusammengeführt. Hierfür wurde ein Puppet-Modul „dhrep“⁴ erstellt, mit dem mittels virtueller Maschinen und der Puppet-Konfiguration jeweils einfach DARIAH-DE Repository- und TextGridRep-Server erstellt und hochgefahren werden können. Gemeinsam mit dem TextGridRep Puppetmodul⁵ ist es nun möglich, die Test-, Entwicklungs- und Produktivinstanzen des DARIAH-DE Repositorys zentral und automatisiert zu konfigurieren und zu aktualisieren.⁶

Portierung der Anwendungen Digilib, Wildfly und Mirador

Die Services Digilib, Wildfly und Mirador wurden vom TextGrid Repository für das DARIAH-DE Repository portiert und an die Spezifika desselben angepasst. So können nun auch Kollektionen im DARIAH-DE Repository, die Bilder enthalten, bequem über den Mirador-Viewer angezeigt werden. Die Bilder des DARIAH-DE Repositorys, die mit Digilib ausgeliefert werden können, werden per IIIF-Protokoll exportiert und können im Mirador-Viewer⁷ als Edition oder Kollektion visualisiert werden.

⁴ DARIAH-DE/puppetmodule-dhrep. Github. <https://github.com/DARIAH-DE/puppetmodule-dhrep>

⁵ Dariah-de-puppet. GitLab. <https://gitlab.gwdg.de/dariah-de-puppet> (intern)

⁶ Vgl. M4.3.3.2. TextGrid-Produktivsystem.

⁷ DH-rep. Mirador-Viewer. <https://dev.textgridlab.org/1.0/iiif/mirador?uri=hdl:21.T11991%2F0000-0011-B8C6-E>

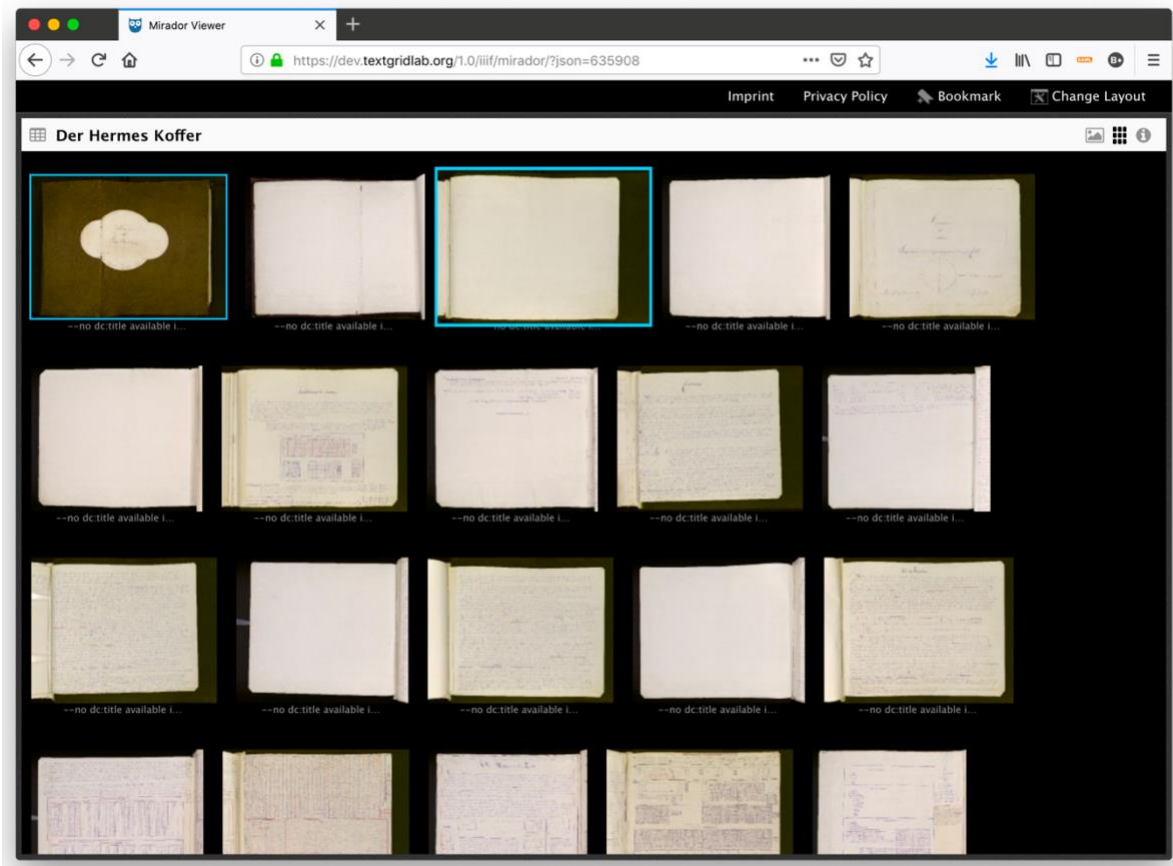


Abbildung 2: Darstellung einer Monographie im Mirador-Viewer

Erweitern der Index- und Landing-Pages

Die Index- und Landing-Pages der Objekte des DARIAH-DE Repositorys wurden überarbeitet und sind nun um neue Links und die neuen angebotenen Dienste wie beispielsweise Digilib und Mirador-Viewer erweitert.

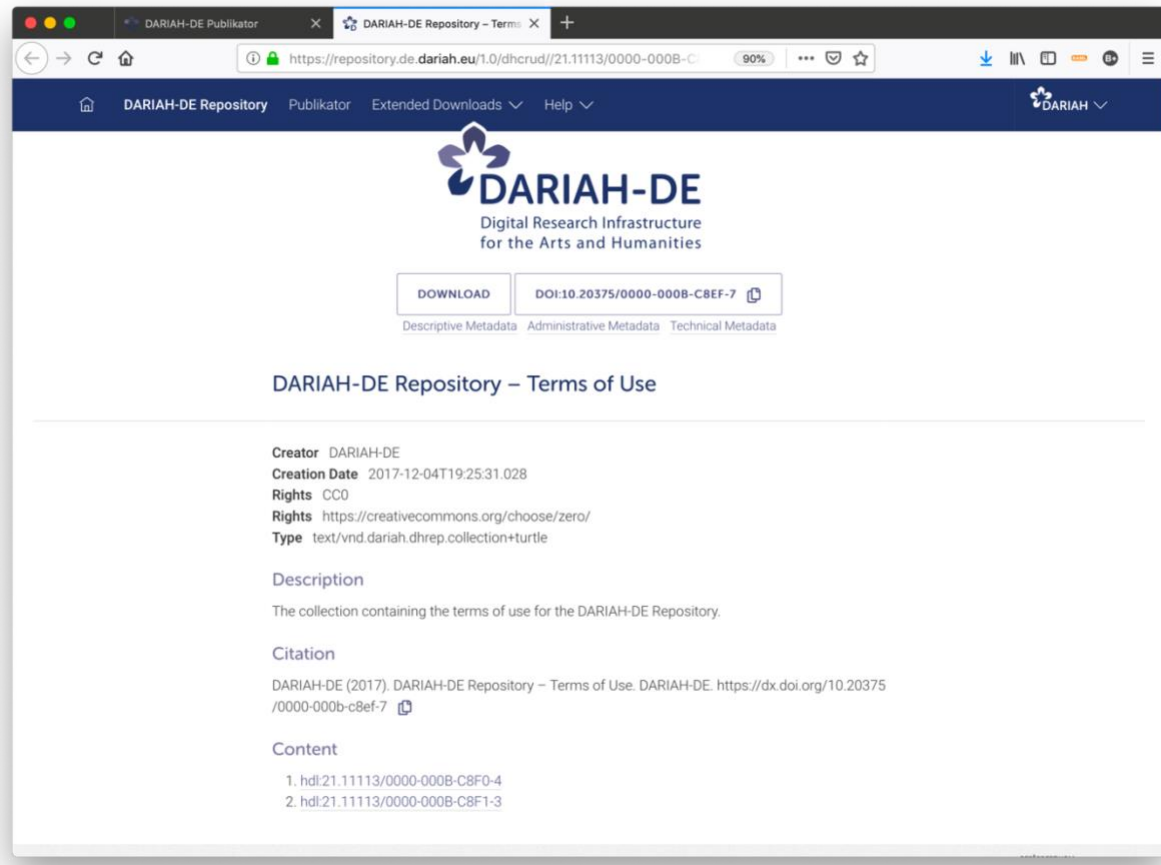


Abbildung 3: Die Landing-Page eines DARIAH-DE Repository-Objekts

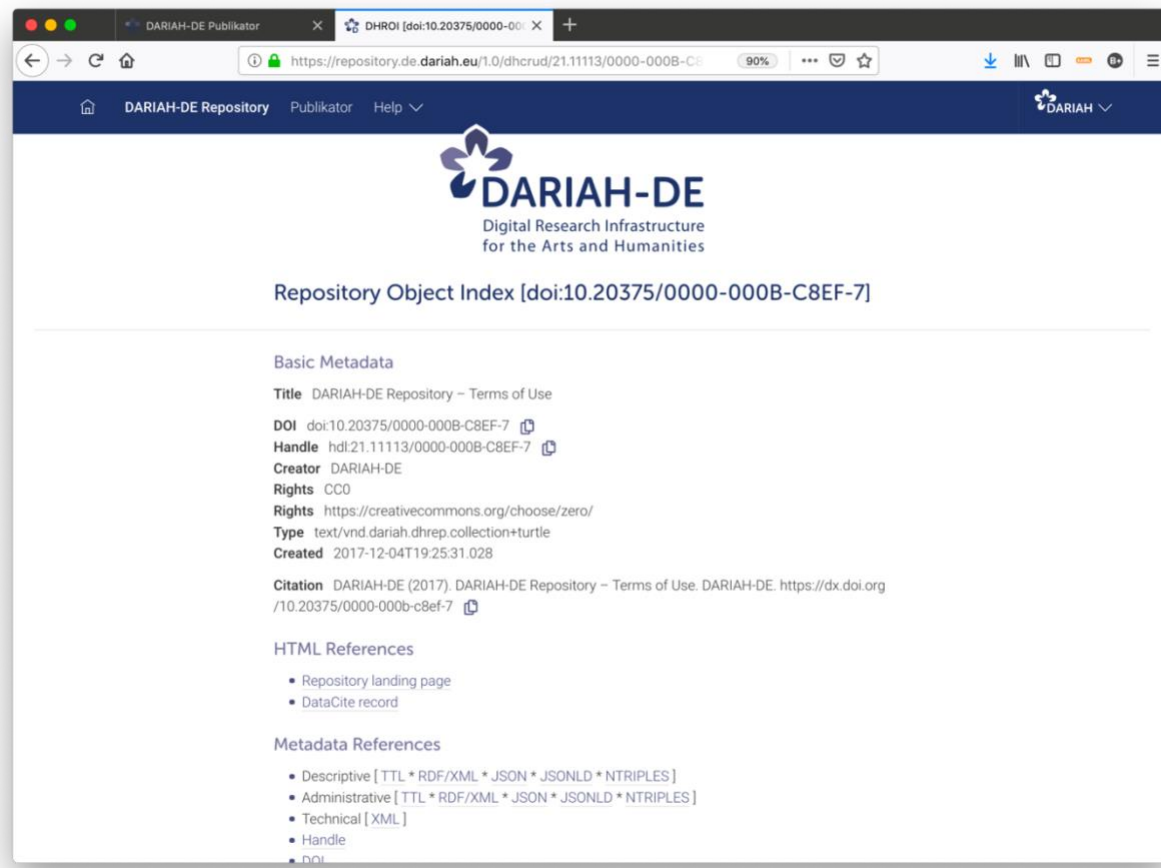


Abbildung 4: Die Index-Seite eines DARIAH-DE Repository-Objekts

Anpassen der HTML-Seiten des Repositorys an den neuen StyleGuide

Die HTML-Seiten des DARIAH-DE Repositorys wurden an den neuen DARIAH-DE Styleguide⁸ angepasst, darunter der Publikator, die Index-Seiten sowie die Landing-Pages und Fehlerseiten.

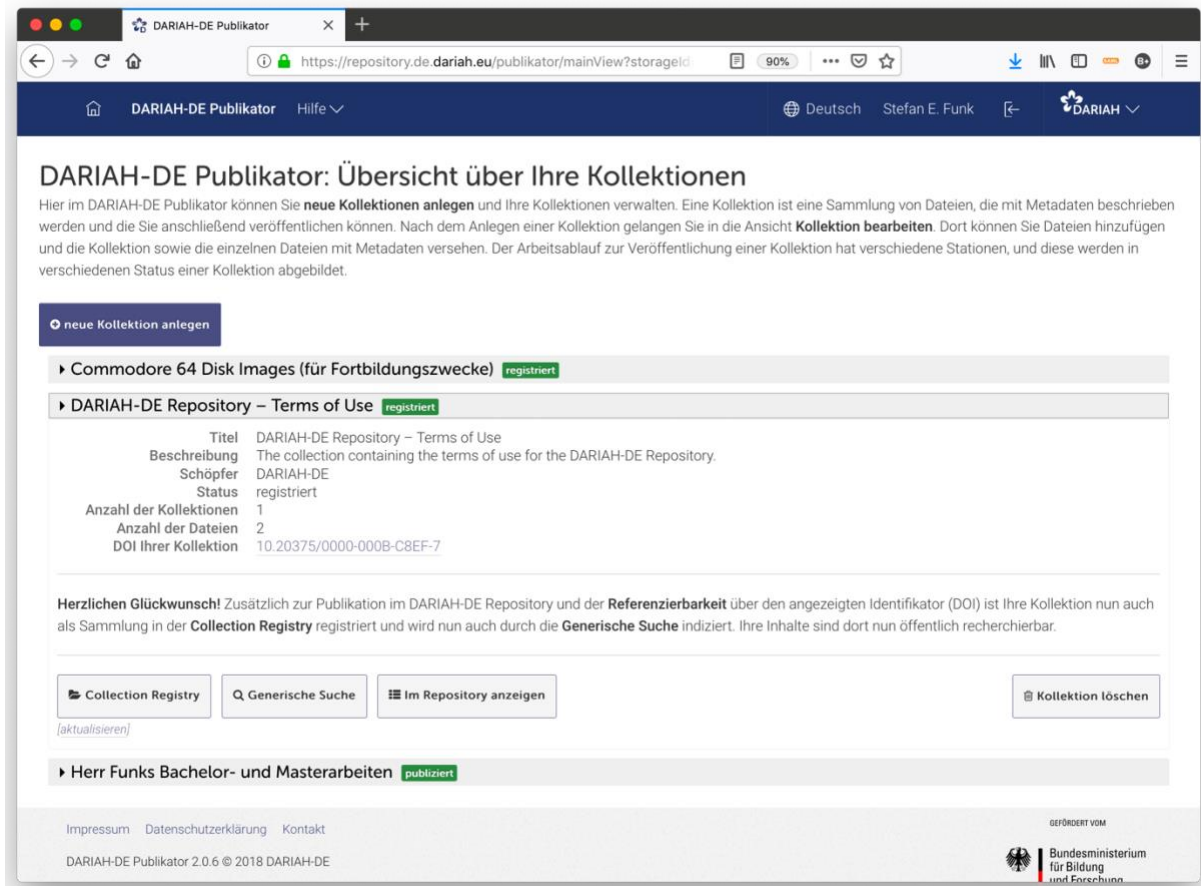


Abbildung 5: Der Publikator im neuen Gewand

⁸ DARIAH-DE Style Guide Template. <https://github.com/DARIAH-DE/StyleGuideTemplate>

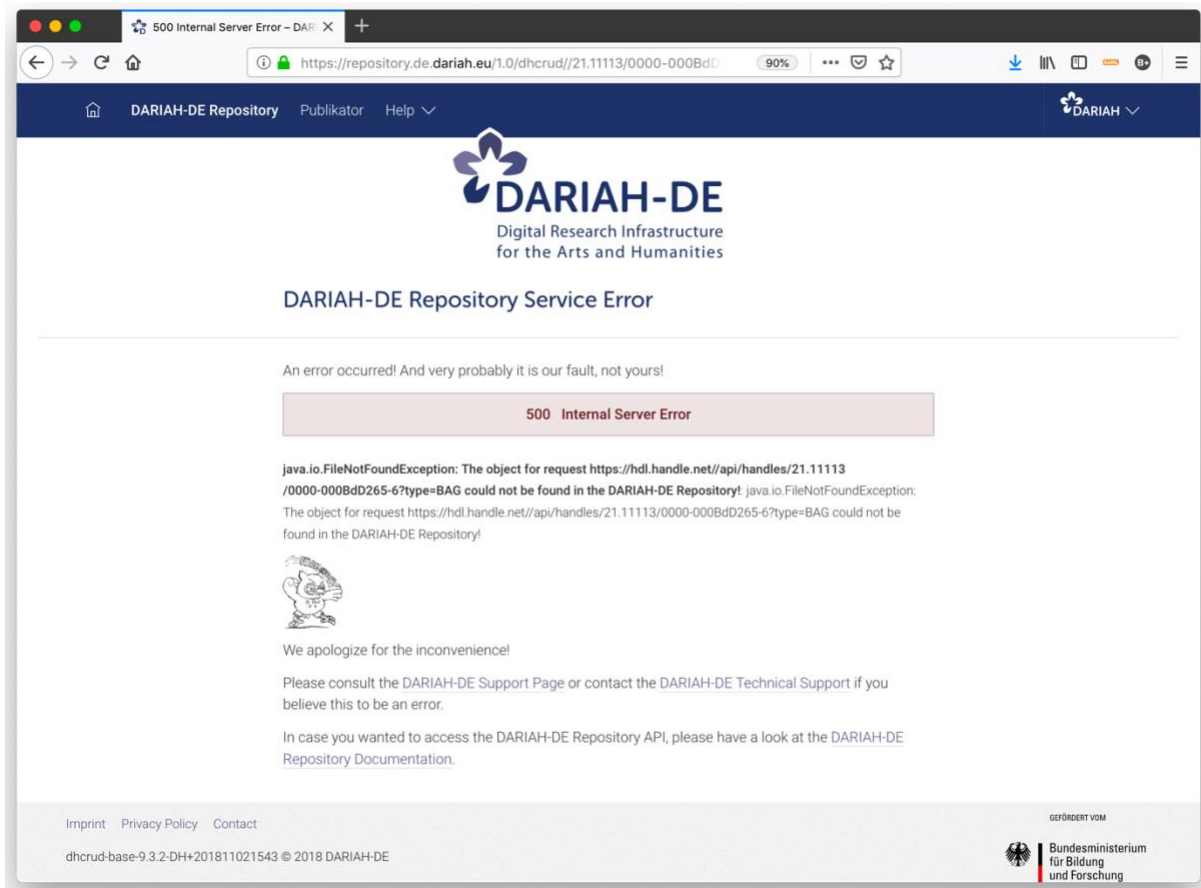


Abbildung 6: Ein absichtlich produzierter Fehler – die Fehlerseiten im neuen Gewand

Optimierung des Publikationsworkflows und angehängter Dienste

Der Publikationsworkflow und die entsprechenden angehängten Dienste, der Publikator, DH-publish, DH-crud, DH-pid, DH-oaimph und weitere konnten weiter optimiert werden. Hier wurden zahlreiche kleinere Verbesserungen und Bugfixes durchgeführt.

Erweiterung und Aktualisierung der API-Dokumentation und des User-Guides

Die Dokumentation des DARIAH-DE Repositorys⁹ wurde beständig aktualisiert und der User-Guide des Publikators samt DARIAH-DE Repository-Beschreibung wurde ebenfalls, wie die API-Dokumentation, auf die Dokumentations-Seite des Repositorys portiert. So kann nun im doc/-Ordner der einzelnen Dienste die Dokumentation jederzeit editiert werden, nach Einspielen in das Git-Repository wird die Dokumentation automatisch erstellt und deployed. Auch Release- und Snapshot-Zyklen können so gut verwaltet werden.

⁹ Vgl. The DARIAH-DE Repository Documentation. <https://repository.de.dariah.eu/doc/>

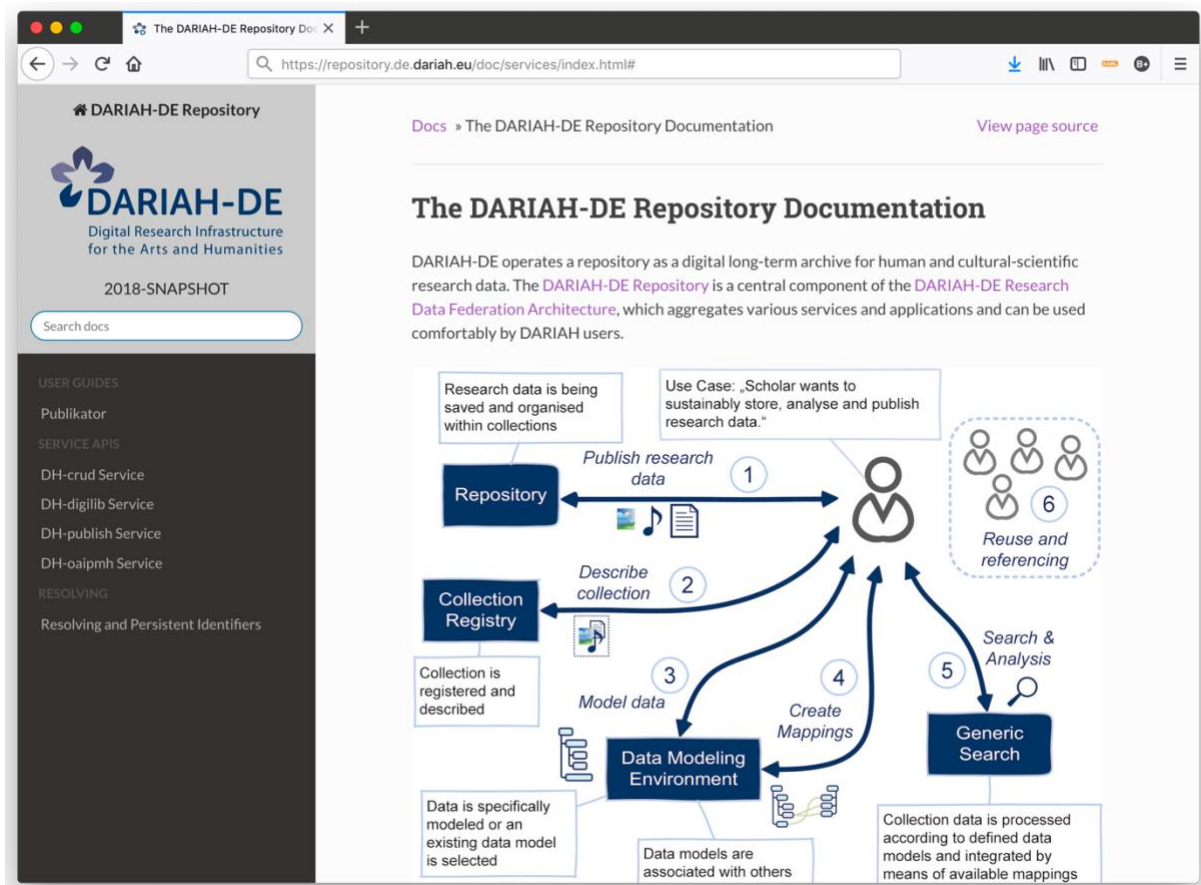


Abbildung 7: Die Dokumentation des DARIAH-DE Repositories

Optimieren der Download-Funktion von DH-crud für Web-Streaming

In DH-crud wurde darüber hinaus die Download-Funktion etwa für große Videodateien optimiert. Große Dateien werden nun als Stream zum Download zur Verfügung gestellt, so dass z. B. eine Videodatei schon während des Downloads angezeigt und abgespielt werden kann.

RDAWG Research Data Repository Interoperability BagPack

Die RDA Working Group Research Data Repository Interoperability¹⁰ hat es sich zur Aufgabe gemacht, eine API bzw. ein Austauschformat für Objekte aus Repositorien zu erarbeiten. Die Final Recommendations finden sich ebenfalls auf den Seiten der Working Group.¹¹

Das dort spezifizierte BagPack-Format wurde für den Export aus dem DARIAH-DE Repository für die Objekte des Repositoriums implementiert und die Nutzung ist in der API-

¹⁰ Vgl. Research Data Repository Interoperability WG. <https://www.rd-alliance.org/groups/research-data-repository-interoperability-wg.html>

¹¹ Vgl. Research Data Repository Interoperability WG Final Recommendations. <https://www.rd-alliance.org/group/research-data-repository-interoperability-wg/outcomes/research-data-repository-0>

Dokumentation des DH-crud beschrieben.¹² So ist das DARIAH-DE Repository für einen zukünftigen Austausch von Daten gerüstet. Die Implementierung eines Imports über dieses Format ist angedacht.

3. Zusammenfassung

In Bezug auf die technische Umsetzung wie im Hinblick auf angebotene Tools wurden wichtige Fortschritte gemacht. Diverse relevante Dienste und Funktionen des DARIAH-DE Repository konnten optimiert werden. An anderer Stelle wurden Anwendungen ausgebaut und Seiten um neue Dienste erweitert. Die Betreuung der Dienste wurde auf technischer Seite erleichtert, die Funktion der Dienste kann somit zuverlässiger sichergestellt werden.

¹² Vgl. DH-crud. <https://repository.de.dariah.eu/doc/services/submodules/tg-crud/service/dhcrud-webapp-public/docs/index.html>