

FAIRes Datenmanagement mit dem DARIAH-DE Repository

Jander, Melina

jander@sub.uni-goettingen.de
SUB Göttingen, Germany

Weimer, Lukas

weimer@sub.uni-goettingen.de
SUB Göttingen, Germany

Einleitung

Forschende, Lehrende und Studierende produzieren bei ihrer Arbeit kontinuierlich Daten. Im Zuge des kulturellen Wandels bleibt die Frage danach, wie mit diesen Daten den FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable; Wilkinson et al. 2016) entsprechend umgegangen werden kann, stets aktuell und nach wie vor ungelöst. Im direkten Austausch mit der Community hat CLARIAH-DE (durch qualitative Interviews mit verschiedenen Zielgruppen)¹ unterschiedliche Bedarfe ermittelt, u.a. die Möglichkeit, Forschungsdaten auffindbar, zugänglich und nachnutzbar zu speichern. Das DARIAH-DE Repository wird diesen Anforderungen gerecht, indem sein Publikationswerkzeug, der DARIAH-DE Publikator, mit grafischer Nutzeroberfläche und detaillierten (mehrsprachigen) Beschreibungs- und Hilfetexten auch jene Nutzende abholt, die keine Vorerfahrung im Forschungsdatenmanagement mitbringen. Um diese Qualitätsmerkmale zertifiziert zu belegen, befindet sich das DARIAH-DE Repository derzeit im Beantragungsprozess des CoreTrustSeals (<https://www.coretrustseal.org/>). Das Repository ist Teil des von der SUB Göttingen für das NFDI-Konsortium Text + (<https://www.text-plus.org/>) angebotenen Dienstes der DARIAH-DE Data Federation Architecture (DFA), die insgesamt fünf Dienste in sich vereint. Diese adressieren eine nachhaltige Publikation, aussagekräftige Beschreibung, Mapping und Auffindbarkeit der Forschungsdaten. Dabei stellt die DFA schrittweise jene Dienste bereit, die wichtige Aspekte des Research Data Lifecycles (Puhl et al. 2015) abdecken.

Im Workshop soll das DARIAH-DE Repository mit seinen Funktionalitäten sowohl im übergeordneten Kontext der FAIR-Prinzipien als auch im technischen Kontext der DFA vorgestellt werden. Mit den Teilnehmenden werden die einzelnen Schritte des Publikationsprozesses durchlaufen, um mit diesem vertraut zu werden und mögliche Hemmschwellen bei der Datenpublikation zu senken. Gleichzeitig bietet der Workshop Raum für konstruktives Feedback. Perspektivisch kann dies dazu beitragen, den Wandel innerhalb der akademischen Publikationskultur hin zu einem grundlegenden Bewusstsein von Open Access voranzutreiben und somit die Kartierung der Forschungsdatenlandschaft positiv zu beeinflussen.

Der Workshop richtet sich an Geisteswissenschaftler*innen in allen Stufen der akademischen Laufbahn und unabhängig von ihrer zugehörigen Institution, ihrer Arbeit oder ihren Forschungsinteressen, da der Bedarf an nachhaltiger Datenpublikation in all

diesen Bereichen gleichermaßen hoch ist. Das Repository wird sowohl von Einzelforscher*innen als auch kollaborativ in Forschungsprojekten genutzt. Es beinhaltet aktuell 267 Kollektionen mit mehr als 1.700 Dokumenten.

Repositorien in den Geistes- und Kulturwissenschaften

Das Angebot an Forschungsdatenrepositorien ist vielfältig. Manche Repositorien beschränken sich auf einzelne Fachdisziplinen (z.B. AMAD für Mittelalterstudien), andere auf bestimmte Formate und Communities (bspw. das Bildarchiv Foto Marburg oder das Deutsche Textarchiv) oder Publikationstypen (Forschungs(roh)daten vs. Publikationen). Ferner bieten manche nur Archivfunktionen ohne Veröffentlichung an und viele können ihre Dienste nur für eine kurze Zeit aufrechterhalten. Andere wiederum sind kostenpflichtig (z.B. RADAR).

Das DARIAH-DE Repository kombiniert die oben genannten Elemente und geht somit über reine Archivfunktionen hinaus. Dennoch kann es als generisches Repository nicht alle Bedarfe einzelner Fachcommunities adressieren (z.B. Helling et al. 2020). Seit 2017 ist es durch die DFA (vgl. Abb. 1) und seine Anbindung an DARIAH- und CLARIAH-DE Teil größerer Infrastrukturen, die Nachhaltigkeit garantieren und an geisteswissenschaftliche NFDI-Konsortien angeschlossen sind (Brünger-Weilandt et al. 2020). Nach Ablauf des Förderzeitraums von DARIAH-DE wurde es im Rahmen der DARIAH-DE Betriebskooperationsvereinbarung weitergeführt und war außerdem Teil des Angebots von CLARIAH-DE, das nun als Angebot der SUB Göttingen in Text + übergegangen ist. Via CLARIAH-DE existiert ein Helpdesk², über den Fragen gestellt werden können. Technisch wird das DARIAH-DE Repository von der GWDG und SUB Göttingen betrieben. Um die Fachwissenschaft gezielt anzusprechen und einen thematischen Rahmen für die Forschungsdaten zu bieten, ist es zwar geistes- und kulturwissenschaftlich ausgerichtet, hierbei aber nicht an Einzelwissenschaften gebunden. Es ist ferner durchsuchbar, nicht rein institutionell und bietet neben zitierfähigen Links bspw. auch die Vergabe von Persistenten Identifikatoren (DataCite DOI und ePIC Handle). Darüber hinaus sind die Verwendung des Repositoriums sowie der gesamten DFA, die Speicherung von Daten und alle zusätzlichen Services nicht mit Folgekosten für die Nutzenden verbunden.³

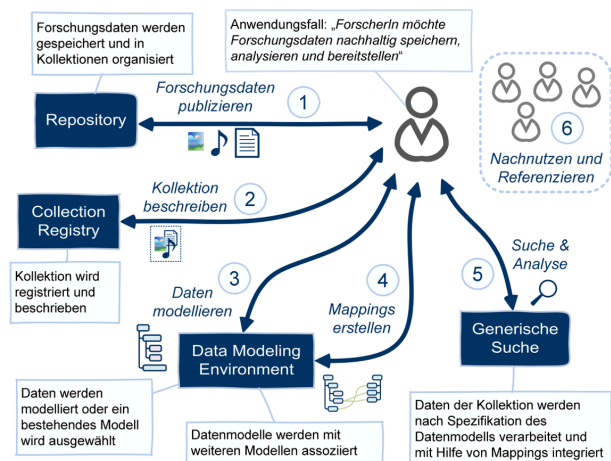


Abb. 1: Schematische Darstellung der DARIAH-DE DFA.

Das DARIAH-DE Repository in der Forschungsdaten-Föderationsarchitektur

Im Rahmen digitalen Forschens gelten die FAIR-Prinzipien zunehmend als Leitlinien. Für Forschende ist es daher wichtig, dass schon der Publikationsprozess auf eine FAIRe Publikation hin ausgerichtet ist (Ivanović et al. 2019). Diesem Anspruch wird das DARIAH-DE Repository auf nutzerfreundliche Weise gerecht: Mit einem DARIAH- oder Föderationsaccount greifen Nutzende auf den Publikator in einem grafischen Interface zu und können mit wenigen Klicks mit dem Einspielen der Daten und der Auszeichnung der Metadaten beginnen. Das Dublin Core Metadaten-Schema liegt hier als Standard zugrunde und ermöglicht es auch jenen Forschenden, die nur wenige Metadaten eingeben möchten, ohne großen Aufwand ihre Daten zu beschreiben. Das Design von Eingabemaske und Fileupload wird dabei als eines wahrgenommen, mit dem intuitiv gearbeitet werden kann (Cremer 2018).

Die im Publikator erstellten Kollektionen können in einem nächsten Schritt in der Collection Registry eingetragen und mit weiteren deskriptiven Metadaten ausgezeichnet werden. Mit der Generischen Suche beinhaltet die DFA außerdem ein Front-End für die in der Collection Registry verfügbar gemachten Daten und deren Metadaten. Durch den modularen Aufbau der DFA können die in ihr vereinten Werkzeuge und Dienste – unter Nutzung u.a. der DARIAH-DE Authentifizierungs- und Autorisierungsinfrastruktur (AAI) sowie der DARIAH-DE Storage API zur Speicherung von Forschungsdaten (Schmunk / Funk 2018) – somit sowohl kombiniert in einem Workflow als auch individuell genutzt werden. Das technische Workaround für einen Datenimport ins Repository lässt sich dabei folgendermaßen beschreiben (vgl. Abb. 2):

1. Publikator: Die Nutzenden laden ihre zur Publikation vorbereiteten Daten über die Eingabemaske ins Repository und versehen sie mit Metadaten entsprechend des Dublin Core Metadaten-Schemas. Sowohl die Objekte als auch ihre Metadaten werden als Kollektion per API an den DARIAH-publish Service übermittelt.
2. DARIAH-publish Service: Die Metadaten werden validiert, Referenzen auf Objekte innerhalb der einzuspielenden Kollektion werden von Dateipfaden auf Identifier umgeschrieben und ggf. werden weitere beschreibende Metadaten generiert. Die gesamte Kollektion wird an den DARIAH-crud Service weitergegeben.
3. DARIAH-crud Service: Die Kollektion wird im DARIAH-Storage gespeichert. Dabei werden die Metadaten in die Indexdatenbank eingetragen und es wird ein Persistenter Identifier (PID) erzeugt.
4. OAI-PMH Service: Der OAI-PMH Service kann öffentlich nach neuen Datensätzen des DARIAH-DE Repositorys angefragt werden. Dazu nutzt er den Metadatenindex des Repositorys. Über diesen Service indiziert auch die Generische Suche regelmäßig alle Daten der Collection Registry und kann sie so den Nutzerinnen und Nutzern komfortabel zur Verfügung stellen.

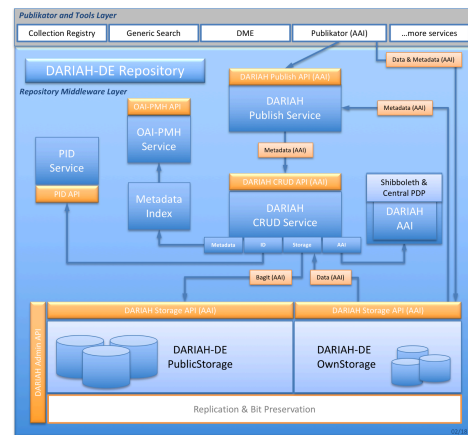


Abb. 2: Schematische Darstellung des Repositorys und des Workarounds im Publikator.

Ablauf der Workshops und Organisatorisches

Unser halbtägiger Workshop wird den Teilnehmenden die Funktionen des Publikators sowie die Möglichkeiten der Collection Registry und die Einbettung in die DFA auf praktische Weise nahebringen. Dies geschieht durch das im Forschungsdatenmanagement typische Szenario des Einspiels eigener Daten sowie der Vorbereitung und Durchführung von deren Publikation. Nach einer kurzen Vorstellungsrunde wird einleitend die DFA im Kontext der FAIR-Prinzipien vorgestellt, sodass möglichst viel Zeit für die praktische Anwendung der Werkzeuge in der DFA zur Verfügung steht. Der Workshop richtet sich an alle am FAIRen Publikations- und Archivierungsprozess interessierte Forschende in den Geistes- und Kulturwissenschaften. Technisches Know-How ist nicht von Nöten, da sich der Workshop auf die praxisnahe Durchführung konzentriert und die zugrunde liegende Technik dabei als Hintergrundwissen vermittelt. Voraussetzungen zur Teilnahme sind folglich nur ein funktionierender Rechner und eine stabile Internetverbindung. Um die Tools der DFA nutzen zu können, ist ein Login in die DARIAH AAI nötig, der mit einem DARIAH-Account oder über viele Forschungsorganisationen und -institutionen möglich ist. Um den Publikationsprozess einzuüben, stellt die Workshoporganisation einen Beispieldatensatz bereit. Gerne können aber auch eigene Daten mitgebracht und verwendet werden. Um eine optimale Durchführung zu gewährleisten und möglichst individuell Feedback geben zu können, ist die Anzahl der Workshopteilnehmenden auf 20 beschränkt.

Format

Halbtägiger Workshop bestehend aus einem kurzen Vortrag sowie einem deutlich längeren Praxisteil, in dem die Teilnehmenden durch den Publikationsprozess und die weiteren Funktionen des Repositorys geführt werden.

Zeitplan

Zeit	Inhalt
60 min	<ul style="list-style-type: none"> Kurzwortstellung der Workshopleitung u. Teilnehmenden Evaluation der Erwartungen Impulsvortrag: FAIRRes Datenmanagement – das DARIAH-DE Repository
15 min	<ul style="list-style-type: none"> Pause
75 min	<ul style="list-style-type: none"> Hands-On: Publizieren mit dem DARIAH-DE Repository (Teil 1)
15 min	<ul style="list-style-type: none"> Pause
75 min	<ul style="list-style-type: none"> Hands-On: Publizieren mit dem DARIAH-DE Repository (Teil 2) Ausblick: Weitere Services der DFA Workshopabschluss

Zielpublikum

- Geisteswissenschaftler*innen mit Interesse an FAIRer Datenspeicherung und –publikation
- Vorkenntnisse nicht notwendig

Teilnehmerzahl

- max. 20

Techn. Anforderungen

Seitens der lokalen Organisation benötigen wir einen ausreichend großen Raum mit stabiler Internetverbindung und üblicher Präsentationstechnik (Beamer). Die Teilnehmenden müssen einen eigenen Laptop mitbringen und sich möglichst im Vorfeld des Workshops einen DARIAH- oder Föderationsaccount einrichten. Es wird von unserer Seite ein Beispieldatensatz für die Veröffentlichung von Forschungsdaten auf dem Testserver des DARIAH-DE Repository zur Verfügung gestellt. Selbstverständlich können auch eigene Forschungsdaten mitgebracht werden.

Fußnoten

- Im Rahmen einer Usability-Studie haben Mitarbeitende von CLARIAH-DE im Juni 2021 qualitative Interviews mit Geisteswissenschaftler*innen verschiedener Karrierestufen geführt. Dabei ging es primär darum, die Bedarfe der Community stichprobenartig kennenzulernen, um die Angebote von CLARIAH-DE dahingehend zu optimieren (Friedrichs / Jander / Reißler-Pipka, in Veröffentlichung).
- “Support,” CLARIAH-DE, Zugriff am 18. November 2021, <https://www.clariah.de/support>.
- Für eine ausführliche Dokumentation des DARIAH-DE Repositorys siehe DARIAH-DE, *DARIAH-DE Repository Documentation: Release 2020-06-25* (2020), https://repository.de.dariah.eu/doc/services/dhrep_doc.pdf.

Bibliographie

Brünger-Weilandt, Sabine / Bruhn, Kai-Christian / Busch, Alexandra W. / Hinrichs, Erhard / Maier, Gerald / Paulmann, Johannes / Rapp, Andrea / von Rummel, Philipp / Schlottheuber, Eva / Schmidt, Dörte / Schrader, Torsten / Simon, Holger /

Stein, Regine / Teich, Elke (2020): “Memorandum of Understanding by NFDI Initiatives from the Humanities and Cultural Studies.” Zugriff am 08. Juli 2021. <https://zenodo.org/record/4045000#.YORQuOgZPY>

CLARIAH-DE (2021): “Support”. Zugriff am 18. November 2021. URL: <https://www.clariah.de/support>.

CoreTrustSeal (2021): Zugriff am 12. Juli 2021. URL: <https://www.coretrustseal.org/>.

DARIAH-DE (2020): “DARIAH-DE Repository Documentation”: Release 2020-06-25 (2020). Zugriff am 08. Juli 2021. URL: https://repository.de.dariah.eu/doc/services/dhrep_doc.pdf.

Cremer, Fabian (2018): “DARIAH-DE Repository: Notizen zum Nutzen jenseits der Nutzung.” DHd-Blog. Zugriff am 08. Juli 2021. URL: <https://dhd-blog.org/?p=10368>.

Ivanović, Dragan / Schmidt, Birgit / Grim, Rob / Dunning, Alastair (2019): “FAIRness of Repositories & Their Data: A Report from LIBER’s Research Data Management Working Group.” Zugriff am 05. Juli 2021. <http://doi.org/10.5281/zenodo.3251593>.

Friedrichs, Sonja / Jander, Melina / Reißler-Pipka, Nanette (in Veröffentlichung): “User Studies zur digitalen Forschungsinfrastruktur von CLARIAH-DE: Konzept, Umsetzung, Erkenntnisse”, in: *DARIAH-DE Working Papers*.

Helling, Patrick / Jung, Kerstin / Pielström, Steffen (2020): “Standards and harmonized components of technical/structural infrastructures for long-term archiving and publishing of complex and heterogeneous data packages”. Text+ User Story. Zugriff am 08. November 2021. URL: <https://www.text-plus.org/en/research-data/user-story-337/>.

Puhl, Johanna / Andorfer, Peter / Höckendorff, Mareike / Schmunk, Stefan / Stiller, Juliane / Thoden, Klaus (2015): “Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften”, in: *DARIAH-DE Working Papers* 11.

Schmunk, Stefan / Funk, Stefan E. (2016): “Das DARIAH-DE- und das TextGrid-Repository: Geistes- und kulturwissenschaftliche Forschungsdaten persistent und referenzierbar langzeitspeichern.” *Bibliothek – Forschung und Praxis* 40, no. 2: 213–221.

Wilkinson, Marc D. / Dumontier, Michel / Aalbersberg, I. Jan et al. (2016): “The FAIR Guiding Principles for scientific data management and stewardship”, in: *Scientific Data* 3.