

## **Daten publizieren, auffinden, nachnutzen: Die DARIAH-DE Data Federation Architecture**

### **Jander, Melina**

jander[at]sub.uni-goettingen.de

Niedersächsische Staats- und Universitätsbibliothek Göttingen,  
Deutschland

ORCID-iD: 0000-0003-1646-6836

### **Weimer, Lukas**

weimer[at]sub.uni-goettingen.de

Niedersächsische Staats- und Universitätsbibliothek Göttingen,  
Deutschland

ORCID-iD: 0000-0001-6919-3646

**Zusammenfassung.** Die DARIAH-DE Data Federation Architecture (DFA) adressiert fünf Aspekte des Research Data Lifecycles und deckt unter Berücksichtigung der FAIR-Prinzipien sowohl Datenpublikation und -speicherung als auch Beschreibung, Mapping und Auffindbarkeit ab. Dadurch geht die DFA weit über die Funktionalität eines reinen Repositoriums hinaus. Eingebettet in die mit BMBF-Förderung aufgebaute Architektur von DARIAH-DE und CLARIAH-DE garantiert die DFA Nachhaltigkeit und Support – bei gleichzeitig niedrigen Nutzungshürden. Denn für deren Verwendung ist lediglich ein DARIAH- oder Föderationsaccount nötig und alle Dienste sind kostenlos. Somit richtet sich das Angebot der DFA an Individualforschende, Forschungsinstitutionen und weitere Datengeber gleichermaßen. Alle drei Gruppen wird der 90-minütige Grundlagenworkshop adressieren, indem er auf anwendungsbezogener Ebene die Funktionen des DARIAH-DE Publikators vorstellt. Der Publikator ist das in die DFA integrierte Dateneingabewerkzeug, mit dem Forschende ihre Daten für den Import ins DARIAH-DE Repository mit Metadaten anreichern, verwalten und final ins Repository einspielen können. Außerdem werden die Möglichkeiten der Collection Registry und der Generischen Suche sowie deren Einbettung in die DFA auf praktische Weise – im Idealfall mit dem Einspielen eigener Daten – ausprobiert und erläutert.

## 1 Einleitung

Nicht erst im Zuge der Empfehlungen des Rates für Informationsinfrastrukturen<sup>1</sup> streben Forschungsdatenrepositorien die möglichst FAIRe Bereitstellung ihrer Daten fernab von Individuallösungen an. Diese Forderung nach FAIRen Forschungsprinzipien und -methoden wurde lange auf den Weg gebracht<sup>2</sup> und ist im Wandel hin zum digitalisierten wissenschaftlichen Arbeiten so wichtig wie nie. So ist es nur folgerichtig, dass sich Forschungsdatenrepositorien an dieser Forderung orientieren. Um dabei die Nutzbarkeit zu erleichtern, versuchen Forschungsdateninfrastrukturen verstärkt, möglichst viele Schritte im Datenpublikations- und -archivierungsprozess in einem einzelnen Angebot verfügbar zu machen.

Darum vereint die DARIAH-DE Data Federation Architecture (DFA) insgesamt fünf Dienste in sich, die eine nachhaltige Publikation, aussagekräftige Beschreibung, Mapping und Auffindbarkeit der Forschungsdaten adressieren. Dabei stellt sie schrittweise jene Dienste bereit, die wichtige Aspekte des Research Data Lifecycles<sup>3</sup> abdecken. Der Workshop will diese Schritte gemeinsam mit den Teilnehmenden durchlaufen und somit zur Publikation eigener Forschungsdaten animieren. Dabei bietet er Raum für konstruktives Feedback, welches durch enge Zusammenarbeit der Workshop-Leitung mit den Entwicklern der DFA in deren technische Optimierung einfließt. Perspektivisch kann dies dazu beitragen, den Wandel innerhalb der akademischen Publikationskultur voranzutreiben.

## 2 Repositorien in den Geistes- und Kulturwissenschaften

Das Angebot an Forschungsdatenrepositorien ist vielfältig. Manche Repositorien beschränken sich auf einzelne Fachdisziplinen (z.B. AMAD), andere auf bestimmte Formate und Communities (bspw. Bildarchiv Foto Marburg oder DTA) oder Publikationstypen (Forschungs(roh)daten vs. Publikationen). Ferner bieten manche nur eine Archivfunktion ohne Veröffentlichung, viele können ihre Dienste nur

---

<sup>1</sup> RfII 2019.

<sup>2</sup> Schirnbacher & Müller 2009.

<sup>3</sup> Ein Überblick über den Research Data Lifecycle ist bspw. auf dem deutschsprachigen Informationsportal zu Forschungsdatenmanagement (FDM) *forschungsdaten.info* zu finden: [forschungsdateninfo.de](http://forschungsdateninfo.de) 2021. In Bezug auf DARIAH-DE bzw. die Geisteswissenschaften vgl. Puhl et al. 2015.

für eine sehr kurze Zeit aufrechterhalten, andere sind kostenpflichtig (z.B. RADAR).

Mit seinen Diensten geht das DARIAH-DE Repository, welches sich aktuell in der Antragsphase für die Zertifizierung mit dem CoreTrustSeal<sup>4</sup> befindet, über reine Archivfunktionen hinaus. Es ist durch die DFA und seine Anbindung an DARIAH- und CLARIAH-DE Teil größerer Infrastrukturen, die Nachhaltigkeit garantieren und an geisteswissenschaftliche NFDI-Konsortien angeschlossen sind. 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 nicht rein institutionell und bietet neben zitierfähigen Links bspw. auch die Vergabe von Persistenten Identifikatoren (DataCite DOI und ePIC Handle). Technisch von der GWVG und SUB Göttingen betrieben, existiert gleichzeitig via CLARIAH-DE ein Helpdesk,<sup>5</sup> über den Fragen gestellt werden können. Darüber hinaus sind die Verwendung des Repositoriums sowie der gesamten DFA, die Speicherung von Daten und alle zusätzlichen Services kostenlos.<sup>6</sup>

### **3 Die DARIAH-DE Data Federation Architecture (DFA)**

Im Rahmen digitalen Forschens gelten die FAIR-Prinzipien zunehmend als Leitlinien. Oftmals beziehen sich diese allerdings ausschließlich auf die Erzeugnisse von Forschung, bspw. digitalisierte Sammlungsbestände oder auch Forschungsergebnisse. Für Forschende ist es daher wichtig, dass schon der Publikationsprozess auf eine FAIRe Publikation hin ausgerichtet ist.<sup>7</sup>

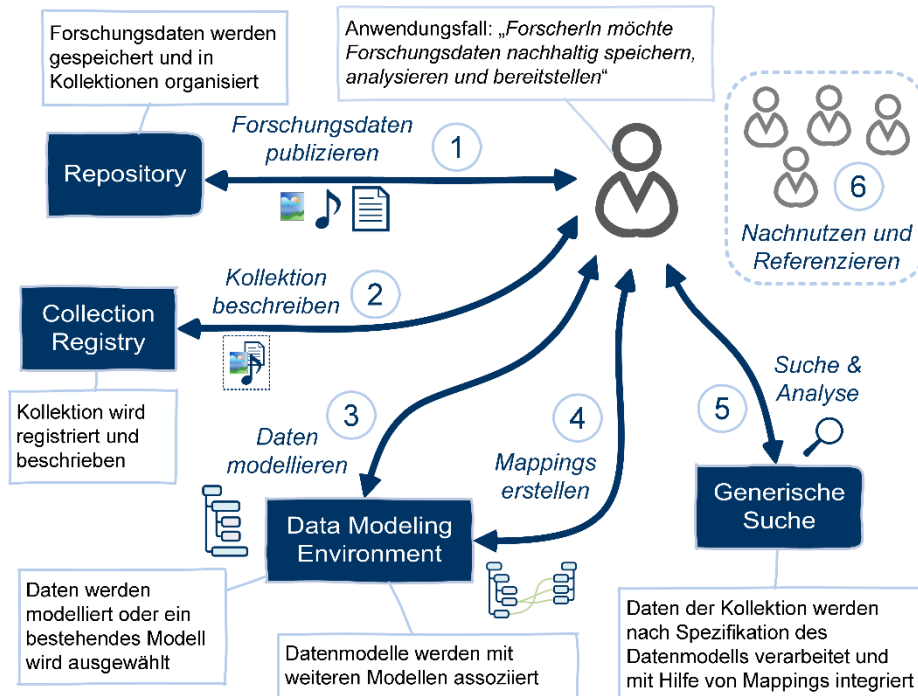
---

<sup>4</sup> Siehe: <https://www.coretrustseal.org/>.

<sup>5</sup> CLARIAH-DE 2021.

<sup>6</sup> Für eine ausführliche Dokumentation des DARIAH-DE Repository siehe DARIAH-DE 2020.

<sup>7</sup> Ivanović et al. 2019.



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

Diesem Anspruch wird das DARIAH-DE Repository (vgl. Abb. 1) – unter Nutzung u.a. der DARIAH-DE Authentifizierungs- und Autorisierungsinfrastruktur (AAI) sowie der DARIAH-DE Storage API zur Speicherung von Forschungsdaten<sup>8</sup> – mit seinem modularen Aufbau gerecht. Die DFA setzt sich aus dem Publikator, dem Repository, der Collection Registry, dem Data Modeling Environment und der Generischen Suche zusammen. Während der Publikator es ermöglicht, Forschungsdaten online über eine grafische Benutzeroberfläche ins Repository einzuspielen und dabei mit Metadaten zu versehen, können in der Collection Registry sowohl jene eingespielten Daten als Kollektionen registriert werden als auch externe Sammlungen durch entsprechende Beschreibungen auffindbar gemacht werden. Mit der Generischen Suche beinhaltet die DFA außerdem ein Front-End für die in der Collection Registry und dem Repository verfügbar gemachten Daten und deren Metadaten. Ein besonderer Vorteil ist dabei die kostenfreie Vergabe von Persistenten Identifikatoren (DataCite DOI und ePIC Handle), denn dadurch wird die dauerhafte Referenzierbarkeit der eingespielten Daten sichergestellt. Das Data Modeling Environment ist

<sup>8</sup> Schmunk & Funk 2016.

schließlich eine Komponente, in der Datenmodelle verzeichnet sind. Es ist ein Werkzeug zur Modellierung und Assoziation von Daten, worauf integrative Dienste wie die Generische Suche zugreifen.

Während es zum Kerngeschäft aller Forschenden gehört, Forschungsdaten zu erzeugen und zu analysieren, erfährt das nachträgliche Publizieren der Daten häufig nur wenig Aufmerksamkeit. Die Frage danach, wie mit möglichst geringem Aufwand und technischem Vorwissen die eigenen Forschungsdaten langfristig nachnutzbar gemacht werden können, ist deshalb von zentraler Bedeutung.

Auf diese Frage liefert die DFA mit dem Publikator eine nutzendenfreundliche und zeitsparende Antwort. Mit einem DARIAH- oder Förderationsaccount greifen Nutzende auf den Publikator zu und können mit wenigen Klicks mit dem Einspielen der Daten und Auszeichnung der Metadaten beginnen. Hierfür liegt das Dublin Core Metadatenchema als Standard zugrunde. Das Design von Eingabemaske und Fileupload wird von Forschenden als intuitiv wahrgenommen.<sup>9</sup> Durch die Einbindung des Publikators in die DFA ist dessen Funktionalität nicht auf den Datenupload beschränkt.

#### **4 Durchführung des Workshops**

Unser 90-minütiger Workshop wird den Teilnehmenden die Funktionen des Publikators sowie die Möglichkeiten der Collection Registry und die Einbettung in die DFA auf praktische Weise – im Idealfall mit dem Einspielen eigener Daten – nahebringen. Dabei wird nur kurz zum Thema hingeführt, um möglichst viel Zeit für die praktische Anwendung und den Nutzen der DFA zu haben. Er richtet sich an alle am Publikations- und Archivierungsprozess interessierten Forschenden in den Geistes- und Kulturwissenschaften. Technisches Know-How ist nicht von Nöten; der Workshop konzentriert sich auf die praxisnahe Durchführung und streift die zugrundeliegende Technik nur. 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. Eine optionale Voraussetzung sind außerdem mitgebrachte eigene Forschungsdaten, die dann verwendet werden können, wenn wir gemeinsam den Publikationsprozess durchspielen.

---

<sup>9</sup> Cremer 2018.

Alternativ wird die Workshop-Leitung einen Beispieldatensatz bereitstellen.

Um eine optimale Durchführung zu gewährleisten und möglichst individuell Feedback geben zu können, ist die Anzahl der Workshopteilnehmenden auf 20 beschränkt.

## **Bibliografie**

CoreTrustSeal. Zugriff am 2. Juli 2021.  
<https://www.coretrustseal.org/>

Cremer, Fabian. "DARIAH-DE Repository: Notizen zum Nutzen jenseits der Nutzung." *DHd-Blog*, August 2018. Zugriff am 7. April 2021. <https://dhd-blog.org/?p=10368>.

CLARIAH-DE. "Support". Zugriff am 7. April 2021.  
<https://www.clariah.de/support>.

DARIAH-DE. *DARIAH-DE Repository Documentation: Release 2020-06-25* (2020).  
[https://repository.de.dariah.eu/doc/services/dhrep\\_doc.pdf](https://repository.de.dariah.eu/doc/services/dhrep_doc.pdf).

Forschungsdateninfo.de. "Der Datenlebenszyklus. Stationen des Forschungsdatenmanagements," Zugriff am 12. April 2021.  
<https://www.forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/>.

Ivanović, Dragan, Birgit Schmidt, Rob Grim, and Alastair Dunning. "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>.

Puhl, Johanna, Peter Andorfer, Mareike Höckendorff, Stefan Schmunk, Juliane Stiller, and Klaus Thoden. "Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften." *DARIAH-DE Working Papers* 11 (2015). urn:nbn:de:gbv:7-dariah-2015-4-4.

RfII – Rat für Informationsinfrastrukturen. *Herausforderung Datenqualität. Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel*. Göttingen 2019.

Schirnbacher, Peter, and Uwe Müller. "Das wissenschaftliche Publizieren – Stand und Perspektiven." *cms-journal* 32 (2009): 7–12.

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