

## Poster: Der Kreislauf des (Daten-)Lebens. Zugangspunkte in die DARIAH-DE Datenföderationsarchitektur

Von Lisa Klaffki, Claudio Leone, Hanna-Lena Meiners

### Abstract:

Der Research Data Lifecycle oder Forschungsdatenzyklus ist ein bekanntes Konzept aus der Informationswissenschaft<sup>1</sup>. Es wird in vielen Forschungsinfrastrukturen<sup>2</sup> genutzt, um die verschiedenen Schritte aufzuzeigen, die Daten im Forschungsprozess durchlaufen und um zu verdeutlichen, welche Anforderungen an die Daten und die Werkzeuge oder Dienste gestellt werden. Dabei werden in den jeweiligen Fachbereichen unter Berücksichtigung ihrer heterogenen wissenschaftlichen Praxis und (Daten-)Standards verschiedene Modelle entworfen und diskutiert. Es gibt daher kein verbindliches, einheitliches Modell, die einzelnen Schemata sind aber letztlich Varianten des geisteswissenschaftlichen Forschungsprozesses und akzentuieren durch unterschiedliche Gewichtung einzelne Aufgaben. Sie repräsentieren also grundlegende Prozesse, die in allen Modellen enthalten sind.

Die basalen Prozesse eines Forschungsdatenzyklus bilden daher auch das intellektuelle Konzept<sup>3</sup> hinter der Infrastruktur von DARIAH-DE<sup>4</sup>, dem deutschen Beitrag zu DARIAH-EU<sup>5</sup>, einem digitalen Forschungsinfrastrukturprojekt für die Geistes- und Kulturwissenschaften. Dabei macht der dort vorgeschlagene generische Workflow<sup>6</sup> die Komplexität des Forschungsprozesses deutlich und insbesondere die Iteration der verschiedenen Schritte. Auf der praktischen Ebene wurde innerhalb von DARIAH-DE eine Data Federation Architecture (DFA)<sup>7</sup> entwickelt. Unter diesem Begriff sind mehrere modulare Komponenten gebündelt, die für sich alleine oder im Zusammenspiel genutzt werden können<sup>8</sup>: das

---

<sup>1</sup> S. für ein vereinfachtes, in Anlehnung an das DCC Curation Lifecycle Modell (<http://www.doc.ac.uk/resources/curation-lifecycle-model>) entwickeltes Modell Ludwig / Enke 2013, Leitfaden zum Forschungsdaten-Management, 14–16.

<sup>2</sup> z.B. IANUS <https://www.ianus-fdz.de/it-empfehlungen/lebenszyklus>; UK Data Archive: <http://www.data-archive.ac.uk/create-manage/life-cycle>

<sup>3</sup> J. Puhl et.al.: [Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften](#). *DARIAH-DE Working Papers* Nr. 11. Göttingen: DARIAH-DE, 2015 URN: <urn:nbn:de:gbv:7-dariah-2015-4-4>

<sup>4</sup> <http://de.dariah.eu>, <https://de.dariah.eu/dariah-de-in-kurze>

<sup>5</sup> <http://dariah.eu/>

<sup>6</sup> Puhl et. al. 2015, 42–44.

<sup>7</sup> <https://de.dariah.eu/data-federation-architecture>; Gradl, Tobias; Henrich, Andreas: „Die DARIAH-DE Föderationsarchitektur. Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen“. In: Neuroth, H. et al. (2016) S. 222-228. DOI:10.1515/bfp-2016-0027; T. Gradl, A. Henrich, Ch. Plutte: Heterogene Daten in den Digital Humanities: Eine Architektur zur forschungsorientierten Föderation von Kollektionen. In: Grenzen und Möglichkeiten der Digital Humanities. Hg. von Constanze Baum / Thomas Stäcker. 2015 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1). text/html Format. DOI: [10.17175/sb001\\_020](https://doi.org/10.17175/sb001_020)

<sup>8</sup> S. T. Gradl et. al.: Daten sammeln, modellieren und durchsuchen mit DARIAH-DE. Zenodo, 2017. DOI: <http://doi.org/10.5281/zenodo.582316>, Abb. 1.

DARIAH-DE Repository,<sup>9</sup> die Collection Registry,<sup>10</sup> das Data Modelling Environment<sup>11</sup> und die Generic Search<sup>12</sup> sowie einen EPIC-PID-Service.

Für viele der in den verschiedenen Schritten eines Forschungsdatenzklus anfallenden Aufgaben können Komponenten der Data Federation Architecture zum Einsatz kommen. Hierunter fallen die Indizierung und Anzeige von Forschungsdaten, die Bereitstellung von Beschreibungsschemata für Sammlungsbeschreibungen und deren langfristige Speicherung, sowie eine umfassende Suchfunktionalität für heterogene strukturierte Datensammlungen und Archive. Zudem bietet die DFA spezifische hinterlegte Metadatenstandards und Crosswalks zwischen Metadaten-Schemata, um Hilfestellung beim Mapping von Forschungsdaten unterschiedlicher Herkunft und Beschaffenheit zu ermöglichen.

Die DFA deckt nicht alle Schritte eines Forschungsprozesses und damit letztlich auch des Forschungsdatenzklus ab. So bleibt der Abschnitt der Analyse (je nach Einteilung / Benennung der Etappen im Datenzyklus auch der der Visualisierung) offen. Dies stellt aber kein Problem dar, denn es gibt in den digitalen Geisteswissenschaften hinreichend andere Tools, die diese Lücke zielgerichtet füllen können<sup>13</sup>.

Das Poster möchte den Zusammenhang dieser beiden Konzepte, Forschungsdatenzklus und Datenföderationsarchitektur deutlich machen bzw. die Einstiegsstellen in den Forschungsdatenzklus mit Hilfe der DFA-Module aufzeigen. Dabei lassen sich diese Modelle nicht Eins zu Eins aufeinander beziehen, sie sind vielmehr durch "one-to-many"-Beziehungen miteinander verbunden. Auch wurde der Forschungsdatenzklus für die visuelle Darstellung auf einem Poster vereinfacht, da der oben genannte generische Workflow relativ komplex ist, und stellt daher ein vereinfachtes Modell und nicht die Realität dar. Es wurden folgende Bereiche abgegrenzt: Erstellung, Verarbeitung, Analyse, Archivierung, Zugang und Nachnutzung.

Die modulare Struktur der DFA erlaubt eine ständige Weiterentwicklung und Anpassung an sich verändernde Nutzungsszenarien und neue Entwicklungen im Forschungsdatenmanagement. Sie befindet sich zurzeit noch in einem "work in progress"-Stadium, das Poster wird daher den aktuellen Stand der Entwicklung vorstellen.

Das Poster wird im Oktober mit dem dann aktuellen Stand im Rahmen der DINI-Jahrestagung 2017 in Göttingen präsentiert. Da die Überschneidung im Publikum der beiden Veranstaltungen als eher gering eingeschätzt wird, möchten wir das Poster gerne auch in Innsbruck präsentieren und diesen Ansatz zu DH und Forschungsinfrastruktur vorstellen.

---

<sup>9</sup> <https://repository.de.dariah.eu/publikator/>

<sup>10</sup> <https://colreg.de.dariah.eu/colreg-ui/>

<sup>11</sup> <https://dme.de.dariah.eu/dme/registry/>

In älterer Literatur, Präsentationen etc. getrennt unter "Schema Registry" und "Crosswalk Registry" vorgestellt, wurden die beiden Komponenten im Zuge ihrer Weiterentwicklung im Sommer 2017 zusammengelegt. Die Dokumentation ist zur Zeit noch in Bearbeitung, sodass die Begriffe parallel zu finden sind.

<sup>12</sup> <https://search.de.dariah.eu/search/>

<sup>13</sup> Eine aktuelle Abfrage des DiRT-Verzeichnisses etwa liefert 94 Tools für "Analysis" (<http://dirtdirectory.org/tadirah/analysis>), 105 für "Visualization" (<http://dirtdirectory.org/tadirah/visualization>).