1. Abstract zu: „Forschungsdaten – Versuch einer Definition"

Peter Andorfer, HAB Wolfenbüttel

Schon 1998 empfahl die Deutsche Forschungsgemeinschaft in ihren „Vorschlägen zur Sicherung guter wissenschaftlicher Praxis",[1] dass: „Primärdaten als Grundlage für Veröffentlichungen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, zehn Jahre lang aufbewahrt werden [sollen]."[2] Dieser Passus findet sich unverändert auch in der 2013 veröffentlichten „ergänzten Auflage" wieder. Als Primärdaten[3] nennt die DFG dabei Daten, die in allen „experimentellen Wissenschaften" aus „Einzelbeobachtungen", „Experimenten" und „numerischen Rechnungen" gewonnen würden. In den „Sozialwissenschaften" wäre es außerdem mehr und mehr üblich, „Primärdaten nach Abschluss ihrer Auswertung durch die Gruppe, die die Erhebung verantwortet, bei einer unabhängigen Stelle zu hinterlegen." Weitere Primärdaten wären darüber hinaus noch: „Messergebnisse, Sammlungen, Studienerhebungen, Zellkulturen, Materialproben, archäologische Funde, Fragebögen". Wie zu sehen ist, bleibt der große Bereich der Geistes- und Kulturwissenschaften bei diesen Überlegungen aber weitgehend ausgespart. Dies dürfte nicht zuletzt auch daran liegen, dass Begriffe wie Primär- oder Forschungsdaten in den geistes- und kulturwissenschaftlichen Disziplin schlichtweg nicht gebräuchlich sind. Womit gearbeitet, woran geforscht wird, sind im Selbstverständnis wohl in erster Linie Quellen. Quellen, die in Publikationen dann im Regelfall auch im Fußnotenapparat bzw. im Quellen- und Literaturverzeichnis in Form von Verweisen nachgewiesen werden. Als besonders wichtig erachtete oder nur schwer zugängliche Quellen können darüber hinaus noch meist in einem Anhang als Reproduktion (z. B. Faksimile oder Transkript) der Publikation beigefügt werden. Vor dem Hintergrund dieser fest etablierten Tradition des Publizierens geistes- und kulturwissenschaftlicher Ergebnisse ist es wenig verwunderlich, dass sich die Frage nach dem Umgang mit Primär- oder Forschungsdaten und das Problem der Aufbewahrung dieser Materialen nicht stellt.

---

[1] DFG, Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft", Bonn 1998, S. 21f, http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf.

[2] DFG, Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Ergänzte Auflage. Bonn 2013, S. 21f, http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf.

[3] Kritisch zum Begriff Primärdaten siehe Jens Klump, Digitale Forschungsdaten, in: Heike Neuroth u.a. (Hg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3, 2010, S. 523-535, hier v.a. S. 524, http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949.

Für den Nachweis und Beleg der eigenen Forschungsergebnisse mag dieses System in den allermeisten Fällen ausreichen. Die effiziente Weiterarbeit an den in Fußnotenapparat und Quellenverzeichnis benannten Materialen wird dadurch aber kaum unterstützt, höchstens vielleicht dadurch, dass anderen Forschern vielleicht das mühevolle Suchen der genannten Quellen etwas erleichtert wird. Der Gang ins Archiv, das Ausheben des gesuchten Materials, die Anfertigung einer Reproduktion oder das Transkribieren relevanter Passagen bleibt nicht erspart, und dass, obwohl ein Fachkollege genau dieselbe Quelle schon selbst im Archiv gesucht, davon eine Reproduktion angefertigt, eine Kopie, eine Fotographie oder einen Scan, und womöglich gar den gesamten Inhalt der Quelle in Form eines Regestes oder einer Transkription erfasst hat. In die Publikation fließt davon aber vielleicht nur eine knappe Paraphrase eines Teiles der Quelle samt der entsprechenden Archivsignatur in der Fußnote. Die Faksimiles und Transkripte hingegen liegen auf der privaten Festplatte des Forschers, wodurch diese Forschungsdaten zu einer bestimmten Quelle faktisch für Andere nicht mehr zugänglich sind.

Geht man wie eben von einem konkreten Beispiel aus, so gestaltet sich die Unterscheidung von Quellen und Forschungsdaten als einfach und nachvollziehbar. Als ungleich schwieriger, vor allem vor dem Hintergrund der großen Heterogenität geistes- und kulturwissenschaftlichen Arbeitens, erweist sich hingegen die Formulierung einer möglichst generischen Definition des Begriffs Forschungsdaten. Insbesondere dann, wenn diese Definition nicht nur der erwähnten Diversität der einzelnen Disziplinen gerecht werden soll, sondern auch die praktisch-technischen Rahmenbedingungen für den Aufbau eines Forschungsdatenrepositoriums vorzugeben hat.

Innerhalb DARIAH-DEs wurde an einer solchen Definition gearbeitet. Ein weit fortgeschrittener Entwurf einer Definition von „digitalen geistes- und kulturwissenschaftlichen Forschungsdaten innerhalb DARIAHs" wird im Rahmen des Vortrages vor- und natürlich auch zur Diskussion gestellt. Bei dieser Vorstellung werden sukzessive die einzelnen Bauteile der Definition näher beleuchtet und die Entscheidungsschritte, die zur letztendlich vorliegenden Formulierung geführt haben, beschrieben.

In einem zweiten Schritt wird diese Definition einem Praxistext unterzogen. Ausgehend von einem konkreten Forschungsprojekt werden jene Materialen, die sich im Zuge des Forschungsvorhabens auf der Festplatte angehäuft haben, dahingehend untersucht, ob es sich darum um Forschungsdaten im Sinne der DARIAH-Definition handelt. Die Ergebnisse dieses Praxistestest sollen dabei einerseits zur schärferen Abgrenzung von Begriffen wie „Quelle", „Primärdaten", „Rohdaten", „Forschungsdaten" und „Publikation" dienen. Andererseits soll anhand dieses Fallbeispiels auch die Problematik der Vielfalt an verwendeten Programmen respektive Dateiformaten reflektiert werden. Behandelt werden Transkripte von archivalischen Quellen im docx-Format, die Auswertung eines Steuerkatastars als

Excel-Dokument im xlsx-Format, Fotographien eines Manuskriptes als jpg, eine in Zotero erstellte Bibliographie, Bilddateien die METS/MODS beschrieben sind, wie auch die Edition dreier Briefe aus dem 18. Jahrhunderts nach den Regeln der TEI.

Vor dieser Vielfalt an Daten und Dateiformaten gilt es dann Konzepte und Workflows zu entwickeln und erproben,[4] um aus den wenig strukturierten und stark idiosynkratrischen Datenmengen, die sich im Laufe eines Forschungsprojektes auf einer Festplatte ansammeln, Forschungsdaten im Sinne der DARIAH-Definition zu generieren. Forschungsdaten, die dann in die ebenfalls von DARIAH-DE entwickelten Infrastruktur, dem DARIAH-Repository und der DARIAH-Collection Registry, eingespeist werden können um, ganz im Sinne des Research Data Lifecycles, Ausgangspunkt für andere Forschungen werden können. Wie ein solcher Workflow aussehen könnte, wird in der Vorstellung von Repository und Collection Registry anhand der in diesem Vortrag konkret benannten Daten demonstriert.

2. Abstract zu: Der Forschungsdatenzyklus in DARIAH-DE. Automatische Erkenntnisse durch Automatisierung von Methoden?

Johann Puhl, HKI Köln

Auf dem großen Feld wissenschaftlicher (Teil- und Unter-) Disziplinen existiert eine breite Vielfalt an Definitionen für einen Forschungsdatenzyklus. Dabei kommen viele dieser Definitionen aus dem Bereich der Informations- und Bibliothekswissenschaften[5] und beziehen sich ganz generisch auf Forschungsdaten und ihre Bereitstellung ohne dabei diese näher gemäß ihrer Zugehörigkeit zu einer Disziplin zu untersuchen. Andere Ansätze stammen eher aus spezifischen (oft naturwissenschaftlichen) Fachbereichen, wie den Lebenswissenschaften[6] oder der Physik.[7]

In DARIAH-DE wird eine Infrastruktur mit einem digitalen Repository als Kernbestandteil implementiert, die einen solchen Lebenszyklus speziell für Forschungsdaten und Fragestellungen aus den Geistes- und Kulturwissenschaften konzeptionell ermöglichen und mithin automatisiert unterstützen soll. Das entscheidende Motiv bei der Konzeption von Forschungsdatenzyklen ist die

[4] Für einen systematischen Überblick zu den "im Feld" verwendeten Dateiformaten samt deren Evaluierung in Bezug auf Archivierbarkeit siehe: IANUS, IT-Empfehlungen für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften, http://www.ianus-fdz.de/it-empfehlungen/.

[5] http://www.lib.ua.edu/wiki/sura/index.php/Data_Life_Cycle_Models.

[6] Joyce M. Ray (Hg.), Research Data Management: Practical Strategies for Information Professionals, USA 2014.

[7] UKOLN, I2S2 Idealised Scientific Research Activity Lifecycle Model, UK 2011, http://www.ukoln.ac.uk/projects/I2S2/documents/I2S2-ResearchActivityLifecycleModel-110407.pdf.

grundsätzliche Gewährleistung, dass Forschungsergebnisse reproduziert und damit überprüft werden können. Daneben soll speziell durch den in einem solchen Zyklus getriebenen Dokumentations- und Publikationsaufwand auch die Nachnutzbarkeit von Forschungsdaten erhöht werden, sodass einmal erhobene Daten oder digitalisierte Dokumente häufiger und in breiterem Kontext gefunden und genutzt werden können.[8]

Die Allianz der deutschen Wissenschaftsorganisationen verweist in ihrer Empfehlung dabei dezidiert darauf, dass „Formen und Bedingungen des Zugangs zu Forschungsdaten […] gesondert für die jeweiligen Fachdisziplinen unter Berücksichtigung der Art und Weise der Datenerhebung, des Umfangs und der Vernetzbarkeit des Datenmaterials sowie der praktischen Brauchbarkeit der Daten entwickelt werden [müssen]."[9]

Für die Bedarfserhebung in den digitalen Geisteswissenschaften und hier insbesondere für den Aufbau einer Infrastruktur in DARIAH-DE gelten daher folgende Modelle als besonders geeignet und können als Grundlage für ein geeignetes eigenes Modell dienen:

Geistes- und sozialwissenschaftliche Ansätze für Forschungs(daten)zyklen

Ein recht differenziertes Modell für einen Forschungsdatenzyklus stammt in den Sozialwissenschaften von der Data Documentation Alliance.[10] Dabei wird folgender Ablauf beschrieben: Aufbauend auf einem „Study Concept” erfolgt die Sammlung von Daten („Data Collection”). Die Sammlung der Daten wird gemäß dem „Study Concept” verarbeitet („Data Processing"), archiviert („Data Archiving") und verbreitet („Data Distribution"). Auf diesen verbreiteten und veröffentlichten Daten kann hernach die Suche und Nachnutzung („Data Discovery”) zur erneuten Analyse („Data Analysis”) und der damit einhergehenden Umwidmung („Repurposing”") für eine veränderte Forschungsfrage der Daten erfolgen. Die umgewidmeten Daten werden hernach erneut verarbeitet und der Forschungsdatenzyklus kann erneut beginnen.

Schon vor dem sozialwissenschaftlichen Ansatz wurden von John Unsworth eine Liste typischer Methoden ("primitives") eines Geisteswissenschaftlers veröffentlicht.[11] Auch diese lässt sich zyklisch (oder wie Unsworth es nennt: rekursiv) betrachten, sodass auf die Tätigkeit des „Representing”

[8] DFG, Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Ergänzte Auflage, Bonn 2013. S. 21f, http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf.

[9] Allianz der deutschen Wissenschaftsorganisationen, Grundsätze zum Umgang mit Forschungsdaten, 2010, http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaetze.htm

[10] DDI Structural Reform Group, DDI Version 3.0 Conceptual Model, DDI Alliance 2004, Figure: "Combined Life Cycle Model", S. 8, http://opendatafoundation.org/ddi/srg/Papers/DDIModel_v_4.pdf.

[11] John Unsworth, Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?, London 2000, http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html.

erneut „Discovering” folgt. Jedoch betont er in der dazugehörigen Veröffentlichung den nicht erschöpfenden Charakter seines Modells. Die Liste enthält folgende Tätigkeiten: Discovering, Annotating, Comparing, Referring, Sampling, Illustrating, Representing.

In den Niederlanden wurde speziell für die historischen Informationswissenschaften ein Zyklus beschrieben, der konkret den Fluss und die Nachnutzung von „historischer Information” darstellt.[12] Hier werden im Forschungsprozess sechs Schritte identifiziert, die zyklisch wiederholt werden können: „Creation", „Enrichment”, „Editing”, „Retrieval”, „Analysis” and „Presentation”. Dabei stellt „Creation” nicht nur den Erhebungs- sondern auch den Anreicherungsprozess dar. Darunter können also auch Tätigkeiten, wie Scannen oder die Verarbeitung eines Scans mit OCR-Methoden fallen.

Zur Bezeichnung „Enrichment” zählen Funktionen wie die Verwendung von Annotationen und die Extension der Daten mit Metadaten. „Editing” hingegen beschreibt als Erweiterung von Enrichment komplexere Tätigkeiten, wie Markup und Erweiterung um intellektuelle Inhalte. „Retrieval” schließlich macht die erweiterten und annotierten Daten such- und nutzbar. Darauf erfolgt der Schritt der „Analysis”, welcher sich eher auf den qualitativen Vergleich von Daten aber auch die quantitative Analyse von Datensätzen bezieht. Die Funktion der „Presentation” ist nun als Abschluss der Schritt, der der Veröffentlichung und Zugänglichmachung der Forschung dient und zur Nachnutzung einladen soll.

Das DARIAH-DE Modell

Eine besondere Herausforderung besteht in DARIAH-DE darin, dass der erarbeitete Forschungsdatenzyklus auch tatsächlich in einer technischen Infrastruktur umgesetzt werden soll. Zusätzlich zu den oben geschilderten geisteswissenschaftlichen Modellen für einen Forschungsdatenzyklus ist hier explizit sowohl die Veröffentlichung der Daten als auch ihre Langzeitarchivierung und Kuration vorgesehen, sodass der nachhaltige Zugang nicht nur zu einer Publikation sondern auch zu den hierfür verwendeten geisteswissenschaftlichen Forschungsdaten sicher gestellt werden kann. Auf diese Weise soll sowohl die Nachnutzung der Forschungsdaten als solche erhöht werden als auch der zu einer Publikation führende Forschungsprozess transparent und nachvollziehbar werden.

[12] Boonstra, Breure und Doorn, Past, present and future of historical information science, Amsterdam 2006, Kapitel 2.2: The life cycle of historical information, S.21 f.

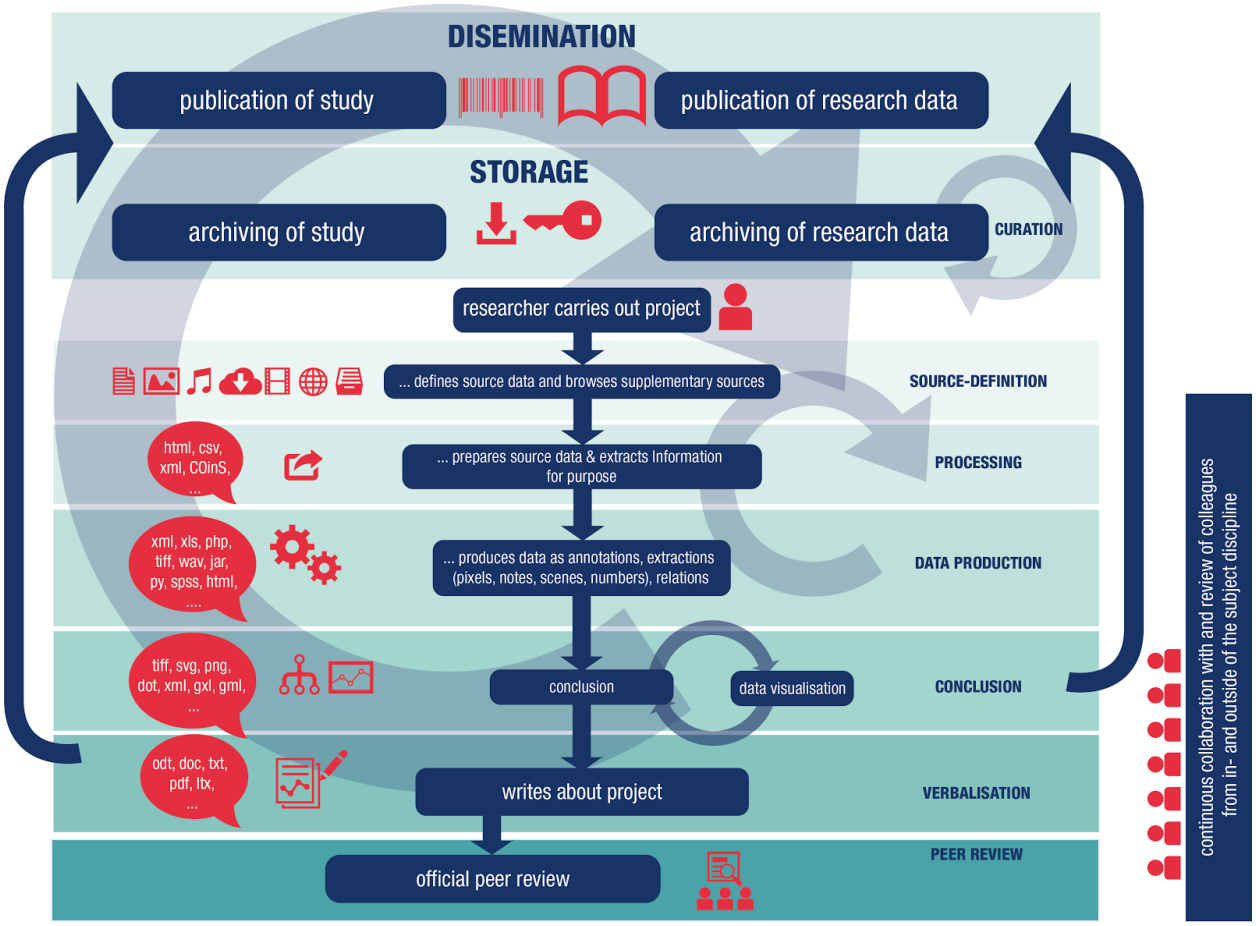

Abb. 1: Referenzmodell für einen Research Data LifeCycle in DARIAH

Bei der Entwicklung eines Referenzmodells in DARIAH-DE konnten eine ganze Reihe zyklischer Prozesse beobachtet werden, die sich nicht in einem linearen System abbilden lassen:
Insbesondere der Peer-Review als auch die Kuration der Daten im Archiv sind Aufgaben, die unbedingt in einer Infrastruktur implementiert werden sollten. Auch kleine und teilweise recht technische Zwischenschritte[13], welche in den großen abstrakten Modellen häufig undefiniert bleiben, bedürfen hier der konkreten Spezifikation.
Eine direkte Folge aus solchen Überlegungen ist der Bedarf nach einem Metadatenmodell, welches in der Lage ist, komplexe Arbeitsflüsse abzubilden, zu jedem Arbeitsschritt Informationen zu speichern oder zu erweitern und so Datei- und Projektversionierung und damit Transparenz über einen Forschungsprozess zu ermöglichen.

[13] Z.B. die Vergabe von persistenten Identifiern, die Spezifikation und Standardisierung von fachspezifischen Dateiformaten, welche in einem Forschungsdatenzyklus verwendet werden, die Angabe akzeptierter auslesbarer Metadatenformate und -felder etc.

Bedingt durch die Wahl eines Datenmodells oder Modellierung eines eigenen Schemas können eine ganze Reihe von Optionen eröffnet werden, die Auswirkungen auch auf die zukünftige Anwendung von Forschungsmethoden in den digitalen Geisteswissenschaften haben können.
Durch die Möglichkeit, Akteure und Ereignisse[14] in den Metadaten einer Infrastruktur zu modellieren, können Tools nicht nur im Routinebetrieb verwendet werden (z.B. Tools für OCR-Scans von Bildern, automatische Lemmatisierung anhand standardisierter Wörterbücher) sondern diese Tools auch als Akteure spezifiziert und in den Metadaten verankert werden. Dabei können die Ergebnisse dieser Prozesse in standardisierten Dateiformaten exportiert, referenziert und später sogar miteinander verglichen werden. Je exakter und gleichzeitig komplexer nun ein solches Metadatenmodell aufgebaut und in einer Infrastruktur verankert ist, desto mannigfaltiger sind die Fragen, die sich mithilfe einer solchen Infrastruktur automatisiert anhand einer Forschungsdatensammlung beantworten lassen.
Entscheidende Kriterien bei der Implementation eines Research Data LifeCycle sind also die Spezifikation der in den Geistes- und Kulturwissenschaften verwendeten Dateiformate und die Beschreibbarkeit und Automatisierbarkeit der darauf anwendbaren Methoden. Hier leistet DARIAH-DE mit der Arbeit an einem Referenzmodell für einen Forschungsdatenzyklus der digitalen Geisteswissenschaften Pionierarbeit, welche sich insbesondere in der praktischen Implementation in einer Infrastruktur auszahlen wird. Der Vortrag soll die hier geschilderten Fragestellungen, insbesondere einzelne Definitionen und Spezifikationen beleuchten und das daraus resultierende Referenzmodell für einen Forschungsdatenzyklus in den digitalen Geisteswissenschaften einer interessierten Öffentlichkeit vorstellen.

3. Abstract zu: „Die Nutzung von Geistes- und kulturwissenschaftlichen Forschungsdaten – Das DARIAH-DE Repository"

Dr. Stefan Schmunk, SUB Göttingen

Im Rahmen von DARIAH-DE widmet sich das Cluster „Wissenschaftliche Sammlungen und Forschungsdaten" nicht nur methodischen und konzeptionellen Fragen des Umgangs, der

[14] Das im Bibliothekswesen eingesetzte Metadatenformat für die Langzeitarchivierung, PREMIS, hält eine sehr komplexe Struktur für „events" und „agents" vor. Vgl. Library of Congress: PREMIS Data Dictionary for Preservation Metadata, Version 2.2, USA 2012, http://www.loc.gov/standards/premis/v2/premis-2-2.pdf.

Generierung, der Nutzung[15] und des Enrichments von digitalen Forschungsdaten, ein zentraler Teil der Tätigkeiten besteht insbesondere auch in der Entwicklung und Realisierung einer Repository-Lösung für geistes- und kulturwissenschaftliche Forschungsdaten.[16]

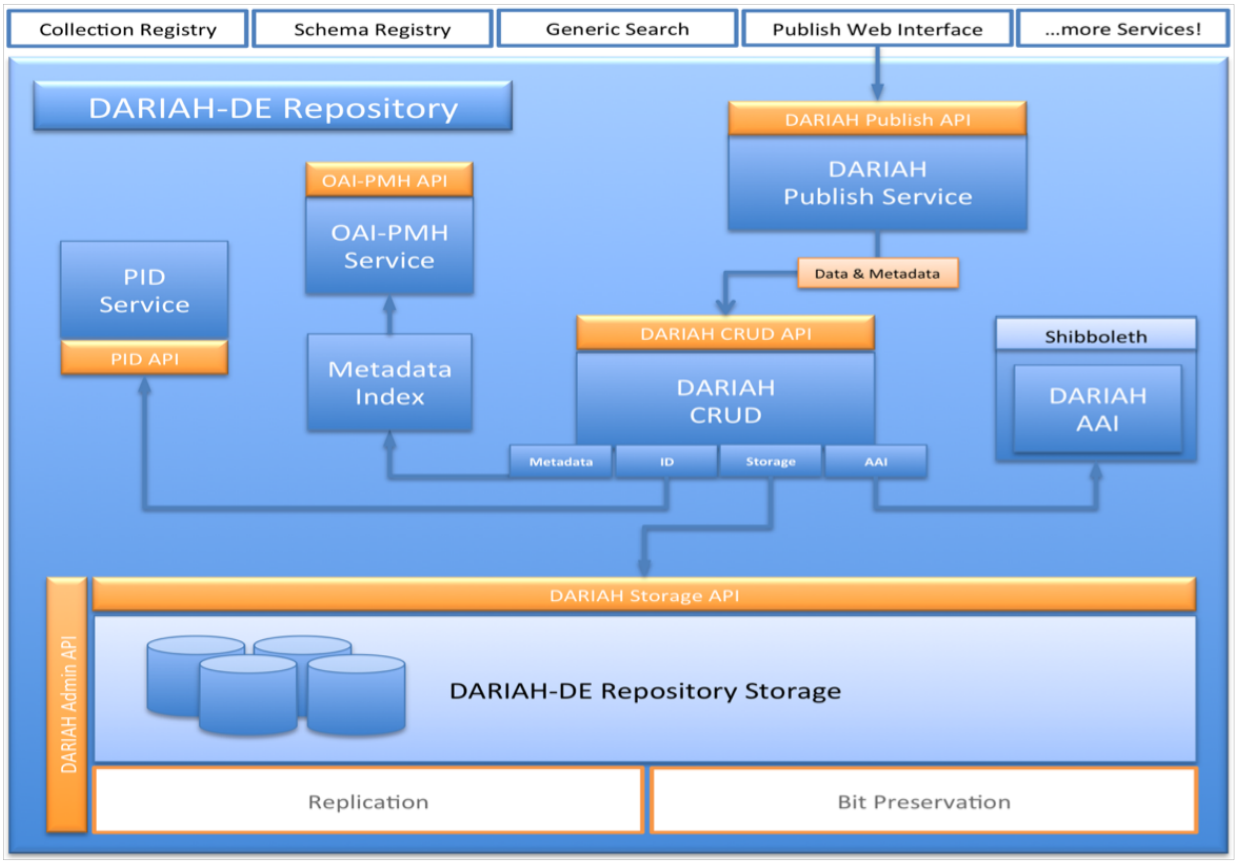

Abb. 2: DARIAH-Repository

Das DARIAH-DE Repositorium wird zukünftig nicht nur assoziierten Forschungsprojekten zur Verfügung stehen, wie derzeit beispielsweise TextGrid,[17] sondern auch EinzelforscherInnen und Forschungsprojekten, die ihre Forschungsdaten persistent, referenzierbar und langzeitarchiviert speichern und Dritten zur Verfügung stellen wollen.

Um dies zu erreichen arbeiten TextGrid, das aus der Virtuellen Forschungsumgebung TextGrid Laboratory[18] und TextGrid Repository[19] besteht, und DARIAH-DE zusammen. Das DARIAH-DE Repositorium stützt sich auf die Codebasis des TextGrid Repository und wird mit verschiedenen

[15] Aber auch Nutzungsmöglichkeiten wie z.B. lizenzrechtlichen Fragen, siehe: https://de.dariah.eu/lizenzen
[16] https://de.dariah.eu/forschungsdaten
[17] https://www.textgrid.de
[18] https://www.textgrid.de/registrierungdownload/download-und-installation/
[19] http://www.textgridrep.de

Service-Instanzen und unterschiedlichen, an das DARIAH-DE Repositorium angepassten Modulen für Funktionen wie Speicher- und AAI-Zugriff, implementiert.

Im Projekt DARIAH-DE wurde beispielsweise in den vergangenen Jahren u.a. eine Authentifizierungs-und Autorisierungsinfrastruktur (AAI) und die DARIAH-DE Storage API für die Speicherung von Forschungsdaten auf Bit Preservation Level aufgebaut, sodass Forschungsdaten zwischen den beteiligten Rechenzentren repliziert werden können. Dadurch ist sichergestellt, dass die Infrastruktur nicht nur als Speicherort für statische Daten verwendet werden kann – diese also öffentlich zugänglich, zitierfähig und langzeitarchiviert sind – sondern ebenso die Möglichkeit gegeben ist, dynamische Daten – die gegebenenfalls durch eine AAI gesichert sind und die aufgrund andauernder aktiver Nutzung aktualisiert werden müssen – dort abzulegen.

Auf die Forschungsdaten kann mithilfe von APIs zugegriffen werden und zugleich werde alle Forschungsdaten mit EPIC-PIDs[20] versehen, sodass andere Tools und Service diese nachnutzen können. Zu diesen Tools gehört beispielsweise die DARIAH-DE Collection Registry.[21] Sie enthält Informationen über beliebige Forschungsdaten-Repositorien und deren Sammlungen. Die in DARIAH-DE entwickelte Generische Suche[22] indiziert die Sammlungen der Collection Registry und bietet so einen userfreundlichen und zudem konfigurierbaren Zugriff auf die Inhalte. Die dritte Komponente bildet die DARIAH-DE Schema Registry, die eng mit der Generischen Suche vernetzt ist und das Mapping unterschiedlichster Metadatenbeschreibungen von Sammlungen ermöglicht. Diese stellt die XML-Schemata für das Mapping und für Metadata Crosswalks zur Verfügung.

Neben dem technischen Aufbau und der Vorstellung des technischen Frameworks steht die Präsentation des Zusammenspiels der technischen Komponenten und deren modularer Struktur - und damit auch nachnutzbarer Integration in DH-Forschungsprojekte – im Vordergrund. Neben den technischen und administrativen Aspekten – wer kann, wie, unter welchen Voraussetzungen, ab wann und wie lange Repository und Collection Registry nutzen und Forschungsdaten nachnutzbar für Dritte speichern – soll anhand der bereits im ersten Vortrag vorgestellten Forschungsdaten der Vorgang des Dateningest in das Repository und der Registrierung von Sammlungsbeschreibungen in der Collection Registry exemplarisch vorgeführt werden. Hierbei sollen insbesondere die Nutzung, der Zugriff und die zugrunde liegenden Arbeits- und Forschungsprozesse beleuchtet werden, um exemplarisch die Möglichkeiten und zugleich die Grenzen eines Forschungsdaten-Repositorys aufzuzeigen.

[20] https://de.dariah.eu/pid-service
[21] Eine Übersicht der verzahnten Applikationen, die zur Speicherung, zur Suche und Recherche und den Zugang zu Forschungsdaten ermöglichen, findet sich hier: https://de.dariah.eu/forschungsdatensammlungen
[22] http://search.de.dariah.eu/search/

Forschungsdaten in Theorie und Praxis. Das DARIAH-DE Repository und die DARIAH-DE Collection-Registry

Sektionsvorschlag - DHd2015 „Von Daten zu Erkenntnissen", eingereicht von:
Peter Andorfer, Johanna Puhl, Stefan Schmunk

Abstract zur Sektion

Das Thema „Forschungsdaten" ist auch innerhalb DARIAH-DEs von zentraler Bedeutung. Dies gilt sowohl für die theoretisch-methodische Verortung dieses Begriffes als auch hinsichtlich des praktischen Umgangs mit Forschungsdaten in den kultur- und geisteswissenschaftlich arbeitenden Disziplinen. Die konkrete Arbeit kreist dabei vor allem um folgende Fragestellungen und Aufgabengebiete:

(1) Was sind Forschungsdaten in den Kultur- und Geisteswissenschaften? Kann angesichts der hohen Heterogenität in den einzelnen Disziplinen, deren vielfältigen Forschungsinteressen, -materialen und Methoden überhaupt eine allgemein verbindliche Definition des Begriffes Forschungsdaten gefunden werden und falls ja, was sind deren Kriterien und welche technisch-praktischen Konsequenzen lassen sich daraus wiederum für die Generierung, Sicherung und Distribution von Forschungsdaten ableiten.

(2) In engem Zusammenhang dazu stehen die Fragen zum Lebenszyklus von Forschungsdaten: So können Forschungsdaten in unterschiedlichen Phasen eines Projektes auf unterschiedliche Art und Weise erzeugt, gesammelt, aufbereitet und/oder analysiert werden. Forschungsdaten können dabei Ergebnis und/oder Quelle eines Forschungsprojektes sein. Diese Dynamik soll in einem eigenen, speziell auf die Eigenschaften kultur- und geisteswissenschaftlicher Forschungsdaten abgestimmten Modell abgebildet werden. Gleichzeit soll dieses Modell eines Forschungsdatenzyklus auch (technische) Anforderungen an Storage- und Publikationssysteme für digitale geistes- und kulturwissenschaftliche Forschungsdaten beschreiben können. Besonders interessant ist an dieser Stelle die Frage, inwiefern sich mithilfe von automatischen Prozessen in einer den Forschungsdatenzyklus unterstützenden Infrastruktur Erkenntnisse gewinnen lassen.

(3) Die in den Punkten eins und zwei ausgearbeiteten Anforderungen werden bei der Entwicklung des DARIAH-DE Repositoriums und der DARIAH-DE Collection Registry aufgegriffen und realisiert. Forschungsdaten, die in das Repository zur langfristigen Archivierung hochgeladen werden, werden in der Collection Registry kontextualisiert, als Sammlung von Forschungsdaten einem konkreten Forschungsprojekt zugeordnet und somit für die Nachnutzung durch andere aufbereitet.

Die Vorträge der Sektion „Forschungsdaten in Theorie und Praxis“ folgen diesen eben skizzierten Themenkomplexen. **Der erste Vortrag mit dem Titel „Forschungsdaten – Versuch einer Definition“ (Peter Andorfer, HAB Wolfenbüttel)** versucht in einem ersten Schritt eine generische Definition von „digitalen geistes- und kulturwissenschaftlichen Forschungsdaten“ und testet die Funktionalität dieser Definition anhand eines konkreten (geschichts)wissenschaftlichen Forschungsprojektes bzw. der darin gesammelten, beschrieben und/oder erzeugten (Forschungs?)Daten. **Im zweiten Vortrag „Definition *des* DARIAH Research Data LifeCycle“ (Johanna Puhl, HKI Köln)** werden der „DARIAH Research Data LifeCycle“ vorgestellt, die Besonderheiten und Spezifika gegenüber bereits bestehenden Modellen herausgearbeitet und die technische Anforderungen an eine Infrastruktur zur Speicherung und Publikation von Forschungsdaten formuliert. Eine solche von DARIAH-DE entwickelte Infrastruktur wird **im dritten Vortrag „Die Nutzung von Geistes- und kulturwissenschaftlichen Forschungsdaten - Das DARIAH-DE Repository“ (Stefan Schmunk, SUB Göttingen)** vorgestellt. Neben den technischen und administrativen Aspekten (wer kann, wie, unter welchen Voraussetzungen, ab wann und wie lange Repository und Collection Registry nutzen) soll anhand der bereits aus dem ersten Vortrag bekannten Forschungsdaten der Vorgang des Dateningest in das Repository und deren Registrierung in der Collection Registry exemplarisch vorgeführt werden.

Mit Hilfe der hier vorgestellten Sektionen sollen vornehmlich zwei Ziele erreicht werden. Einerseits geht es darum, die DARIAH-DE Collection Registry sowie das DARIAH-DE Repository in der DH-Community und über diese hinaus bekannt zu machen. Andererseits sollen die innerhalb von DARIAH-DE erarbeiten Konzepte und Definitionen zum Forschungsdatenbegriff und zum Research Data Lifecycle mit Vertretern der unterschiedlichen geistes- und kulturwissenschaftlichen Disziplinen diskutiert werden.