Digital Humanities

Mirjam Blümm*, Stefan E. Funk und Sibylle Söring

Die Infrastruktur-Angebote von DARIAH-DE und TextGrid

DOI 10.1515/iwp-2015-0055

Zusammenfassung: DARIAH-DE stellt eine digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften zur Verfügung, die Ressourcen, Tools und Services für Forschung und Lehre umfasst. TextGrid, eine virtuelle Forschungsumgebung für die text- und quellenbasierten Geisteswissenschaften, unterstützt Forschende u. a. bei der Erstellung digitaler Editionen. Die beiden BMBF-geförderten Projekte ergänzen sich in vielen Bereichen und arbeiten eng zusammen. Ein Beispiel dafür ist das TextGrid- und DARIAH-DE Repository, ein digitales Langzeitarchiv für geistes- und kulturwissenschaftliche Forschungsdaten.

Deskriptoren: Geisteswissenschaften, Kultur, Forschung, Hochschulausbildung, Langzeitarchivierung, Wissenschaftliches Arbeiten, Infrastruktur, Digital

The digital infrastructures DARIAH-DE and TextGrid

Abstract: DARIAH-DE provides a digital research infrastructure for the arts and humanities with resources, tools and services for research and teaching. TextGrid is a Virtual Research Environment for scholars in the text-based humanities. Among other features, it supports the creation of digital editions. The projects complement each other in many areas and cooperate closely. One example is the TextGrid and DARIAH-DE repository, which is a digital long-term archive for research data.

***Kontaktperson: Dr. Mirjam Blümm**, Georg-August-Universität Göttingen, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Abteilung Forschung & Entwicklung, Platz der Göttinger Sieben 1, 37070 Göttingen, Deutschland, E-Mail: bluemm@sub.uni-goettingen.de

Dipl.-Inform. Stefan E. Funk, Georg-August-Universität Göttingen, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Abteilung Forschung & Entwicklung, Platz der Göttinger Sieben 1, 37070 Göttingen, Deutschland, E-Mail: funk@sub.uni-goettingen.de

Sibylle Söring, M. A., Georg-August-Universität Göttingen, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Abteilung Forschung & Entwicklung, Platz der Göttinger Sieben 1, 37070 Göttingen, Deutschland, E-Mail: soering@sub.uni-goettingen.de

Descriptors: Humanities, Culture, Research, University education, Long term archiving, Academic approach, Infrastructure, Digital

Les offres d'infrastructure de DARIAH-DE et de TextGrid

Résumé: DARIAH-DE propose une infrastructure de recherche numérique pour les sciences humaines et culturelles qui comprend des ressources, des outils et des services pour la recherche et l'enseignement. TextGrid, un environnement de recherche virtuel pour les sciences humaines basées sur du texte et des sources, soutient les chercheurs, entre autre, lors de la création d'éditions numériques. Les deux projets subsideés par le BMBF (ministère allemand de l'éducation et de la recherche) se complètent mutuellement dans de nombreux domaines et travaillent en étroite collaboration. Un bon exemple est le référentiel TextGrid et DARIAH-DE, une archive numérique à long terme pour les données de recherche dans les domaines des sciences humaines et culturelles.

Descripteurs: Sciences humaines, Culture, Recherche, Enseignement supérieur, Archivage à long terme, Infrastructure, Numérique

Arbeiten mit digitalen Forschungsinfrastrukturen

Digitale Werkzeuge und Services, die im Rahmen einer *Digitalen Forschungsumgebung* angeboten und in einer *Digitalen Forschungsinfrastruktur* implementiert sind, können genutzt werden, um vielfältigste Forschungsfragen zu beantworten. Anders als z.B. individuelle und/oder lokale (Einzel-)Lösungen bieten derartige Angebote verschiedene Lösungsmodelle, wenn es darum geht, Forschungsdaten so aufzubereiten, dass sie such- und recherchierbar, nach internationalen Daten- und Metadatenstandards gespeichert und – auch interoperabel – nachgenutzt werden können. Auch unabhängig von der jeweils spezifischen fachwissenschaftlichen Fragestellung sollen bestimmte Methoden, Verfahren und Tech-

Abbildung 1: Der TextGrid Text-Bild-Link-Editor.

nologien auf unterschiedlichstes Datenmaterial angewandt werden können.

Der modulare Aufbau solcher Forschungsumgebungen, die standortunabhängig und fächerübergreifend für verteiltes, kollaboratives Arbeiten genutzt werden können, ermöglicht idealerweise zudem die projektspezifische Anpassung und Erweiterung von Werkzeugen und Diensten. *Digitale Forschungsinfrastrukturen* unterstützen die Geistes-, Kultur- und Sozialwissenschaften darin, digitale Ressourcen, Forschungsdaten und Forschungsergebnisse langfristig, nachhaltig, zitierfähig und interoperabel zur Verfügung zu stellen.

Den Aufbau und Betrieb digitaler Infrastrukturen für die Geistes- und Kulturwissenschaften fördert das Bundesministerium für Bildung und Forschung seit rund zehn Jahren, darunter auch die beiden Verbundprojekte DARIAH-DE¹ und TextGrid². DARIAH-DE unterstützt als digitale Forschungsinfrastruktur die digital arbeitenden Geistes- und Kulturwissenschaften mit entsprechenden Ressourcen, Tools und Services in Forschung und Lehre. TextGrid, eine Virtuelle Forschungsumgebung (VFU), bie-

tet Tools und Dienste sowie ein digitales Langzeitarchiv für die mit digitalen Texten und Quellen arbeitenden Geisteswissenschaften, wie etwa die Editionsphilologie. Zielanwendung ist u. a. die Erstellung digitaler Editionen.

Die gemeinsame Ausrichtung beider Projekte, digitales Arbeiten in den Geistes- und Kulturwissenschaften zu unterstützen, bietet die Grundlage für zahlreiche Synergieeffekte, die – auf technologischer ebenso wie auf informations- bzw. fachwissenschaftlicher Basis – seit mehreren Jahren erfolgreich umgesetzt werden. Ziel ist dabei, Geistes- und KulturwissenschaftlerInnen ein umfassendes, dauerhaftes und aufeinander abgestimmtes Angebot liefern zu können, das den *Digital Humanities* eine fächerübergreifende Perspektive für die Generierung, Bearbeitung und nachhaltige Bereitstellung von Forschungsdaten bietet.

TextGrid

TextGrid, 2006 im Rahmen der D-Grid-Initiative³ (Herpay, Neweling & Schwiegelshohn, 2009; Neuroth, Kerzel &

1 <https://de.dariah.eu/>; Hinweis: sämtliche in diesem Text genannten Internetquellen wurden letztmalig am 18.08.2015 besucht.

2 <https://textgrid.de/>

3 <http://www.d-grid.de/>

Abbildung 2: Vom digitalen Ausgangsmaterial zu zitierfähigen publizierten Forschungsdaten: TextGrid-Workflow für Editionsprojekte.

Gentzsch, 2007) gestartet, hat drei Förderphasen durchlaufen:

- 2006 bis 2009: TextGrid I – Modulare Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung. Erstellung eines Community-Grid für die Geisteswissenschaften
- 2009 bis 2012: TextGrid II: Vernetzte Forschungsumgebung in den e-Humanities
- 2012 bis 2015: TextGrid III: Institutionalisierung einer Virtuellen Forschungsumgebung in den Geisteswissenschaften

Die jeweiligen Projekttitle der Teilphasen entsprechen dabei der stufenweisen Entwicklung der VFU TextGrid vom Aufbau (2006 bis 2009) über den Ausbau (2009 bis 2012) bis zur Verstetigung (2012 bis 2015).⁴ Möglich wurde dies durch das Zusammenspiel breit gefächerter Kompetenzen im Konsortium und den aktiven Austausch mit der Community. In den knapp zehn Jahren der Förderung entwickelten fünfzehn institutionelle Partner – darunter Hochschulen und Forschungseinrichtungen, kommerzielle Partner, aber auch ein Rechenzentrum, eine Akademie und eine Universitätsbibliothek – in nur geringfügig wechselnder Zusammensetzung die VFU TextGrid mit ihren drei Kernbereichen⁵:

1. Das *TextGrid Laboratory*, die eigentliche Software, liefert den Einstiegspunkt in die VFU und stellt unterschiedliche Open-Source-Werkzeuge und -Services für den gesamten Forschungsprozess zur Verfügung, z.B. einen Text-Bild-Link Editor, mit dem etwa das Digitalisat einer Handschrift mit der Transkription verknüpft werden kann (vgl. Abb. 1).

2. Im *TextGrid Repository*, einem Langzeitarchiv für geisteswissenschaftliche Forschungsdaten, können XML/TEI-kodierte Texte, Bilder und Datenbanken sicher gespeichert, publiziert und durchsucht werden.

3. Die *TextGrid Community* ist in den letzten Jahren beständig gewachsen. Neben individuellen Schulungen werden regelmäßige Nutzertreffen angeboten, die neben Einführungen in die Arbeit mit dem TextGridLab und themen- bzw. anwendungsspezifischen Workshops nicht zuletzt auch den Austausch zwischen digitalen Forschungsvorhaben aus den Geisteswissenschaften befördern.

Neben EinzelforscherInnen in kleineren nutzen zahlreiche auch langfristige Forschungsverbünde und Editionsprojekte TextGrid für ihre Arbeit, wie etwa die Projekte *Genetisch-kritische Hybrid-Edition von Theodor Fontanes Notizbüchern*⁶ (Universität Göttingen), *Bibliothek der Neologie*⁷ (Universität Münster) oder *Textdatenbank und Wörterbuch des Klassischen Maya*⁸ (Universität Bonn).

Dabei unterstützt TextGrid mit seinen beiden Hauptkomponenten *TextGrid Laboratory* und *TextGrid Repository* den gesamten wissenschaftlichen Arbeitsprozess im Rahmen der Erstellung digitaler Editionen: Das digitale Ausgangsmaterial – in der Regel Text- und/oder Bilddateien / Faksimiles – wird zunächst in die (geschützte) Virtuelle Forschungsumgebung (TextGrid Laboratory) geladen, dort mit maschinenlesbaren Strukturdaten angereichert (XML/TEI) und texttechnologisch erschlossen. Dazu gehören z.B. Kommentare, Annotationen, Apparate, Register sowie die Verknüpfung von Text- und Bilddaten. Anschließend stehen für die Forschungsdaten drei Publikationswege bereit: die Veröffentlichung im Rahmen einer *Digitalen Edition* (Internet-Portal), die nachhaltige Archivierung

⁴ Einen umfassenden Überblick über die Entwicklungsgeschichte von TextGrid bietet der Sammelband von Neuroth, Rapp & Söring (2015).

⁵ <https://textgrid.de/>

⁶ <http://fontane-nb.dariah.eu/index.html>

⁷ <http://bdn-edition.de/>

⁸ <http://www.iae.uni-bonn.de/forschung/forschungsprojekte/lau-fende-projekte/idiom-dictionary-of-classic-mayan>

Abbildung 3: Die digitale Forschungsinfrastruktur DARIAH-DE.

der Forschungsdaten und ihrer verschiedenen Ausgabeformate im digitalen Langzeitarchiv TextGrid Repository und schließlich, nach Erzeugung von Druckdaten aus den entsprechenden XML/TEI-Anreicherungen im TextGrid Laboratory, die Publikation in Buchform (print).

DARIAH

DARIAH⁹ als europäische Initiative ist 2006 als eines von 48 Projekten auf die ESFRI Roadmap¹⁰ aufgenommen worden, von denen außer DARIAH jedoch nur noch ein weiteres Projekt, CLARIN¹¹, in den Geisteswissenschaften angesiedelt ist. ESFRI, das European Strategy Forum on Research Infrastructures, verfolgt das Ziel, die Wissen-

schaften in Europa insgesamt zu stärken und weltweit sichtbar zu machen. ESFRI-Projekte können mit dem European Research Infrastructure Consortium (ERIC) eine eigene Rechtsform gründen.¹² Seit August 2014 haben sich mit dem DARIAH-ERIC¹³ die 15 DARIAH-Mitgliedsländer, darunter Deutschland, vertraglich zum dauerhaften Betrieb der Forschungsinfrastruktur verpflichtet.

DARIAH-DE ist der deutsche Beitrag zu dieser europäischen Forschungsinfrastruktur. Das Projekt wird auf nationaler Ebene seit 2011 vom BMBF gefördert.

- 2011 bis 2016: DARIAH-DE: Aufbau von Forschungsinfrastrukturen für die eHumanities
- ab 2016: DARIAH-DE Überführung der digitalen Forschungsinfrastrukturen für die e-Humanities in die Operational Phase (in Planung)

⁹ <http://www.dariah.eu/>

¹⁰ http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

¹¹ <http://www.clarin.eu/>

¹² http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric

¹³ <http://dhd-blog.org/?p=3899>

Abbildung 4: Alle datierten Grabmale der Datenbank epidat¹⁴ im Geo-Browser¹⁵ visualisiert.

Wie bei TextGrid, so standen auch bei DARIAH-DE die ersten drei Jahre des Projekts im Zeichen des Aufbaus, die beiden folgenden Jahre wurden und werden zum Ausbau der digitalen Forschungsinfrastruktur genutzt; die Verstärkung wird Ziel des Neuantrags ab 2016 sein.

Das DARIAH-DE-Konsortium setzt sich aus 20 institutionellen Partnern zusammen, darunter sieben Universitäten, vier Forschungseinrichtungen, vier Rechenzentren, zwei Bibliotheken, eine Akademie, ein kommerzieller Part-

ner und eine Non-Profit-Organisation, die gemeinsam mit und für die Forschungscommunity die digitale Forschungsinfrastruktur DARIAH-DE (vgl. Abb. 3) aufgebaut haben. Diese besteht im Wesentlichen aus vier Kernbereichen:

1. Der Bereich *Lehre* umfasst eine umfangreiche Sammlung von DH-Schulungsmaterialien¹⁶, die Bibliographie *Doing Digital Humanities*¹⁷, die das Spektrum der DH-Literatur von Einführungen bis zu vertiefender Forschungsliteratur verzeichnet, sowie ein reiches Angebot an zielgrup-

¹⁴ <http://steinheim-institut.de/cgi-bin/epidat>

¹⁵ <http://geobrowser.de.dariah.eu>

¹⁶ <https://de.dariah.eu/schulungsmaterial-sammlung>

¹⁷ <https://de.dariah.eu/bibliographie>

penspezifischen Workshops,¹⁸ darunter Summer Schools und Experten-Kolloquien z. B. zu kontrollierten Vokabularen, Stylometrie oder Annotationsverfahren.¹⁹

2. Eine weitere Säule von DARIAH-DE ist die *Forschung* zu und Weiterentwicklung von DH-Methoden und -Verfahren (Reiche et al., 2014), z. B. im Bereich Big Data in den Geisteswissenschaften oder zu Methodik und Verfahren fachwissenschaftlicher Annotationen. Außerdem entwickelt DARIAH-DE eine Reihe fachwissenschaftlicher Dienste (weiter), die über das DARIAH-DE-Portal zur Verfügung gestellt werden,²⁰ z. B. einen Geo-Browser zur Raum-Zeit-Visualisierung von Daten, den Noteneditor MEISE oder das quantitative Textanalyse-Tool Digivoy.

3. DARIAH-DE ist zudem im Bereich *Forschungsdaten* aktiv. Hier bietet DARIAH verschiedene Werkzeuge und Dienste²¹ wie etwa die Collection Registry, mit der Forschungsdaten-Sammlungen in DARIAH-DE nachgewiesen und neu registriert werden können, die Schema-Registry und die Crosswalk Registry, mit denen spezifische Metadatenstandards hinterlegt bzw. miteinander verbunden werden können, sowie ein Forschungsdaten-Repositorium für wissenschaftliche Sammlungen. Außerdem gibt DARIAH-DE Empfehlungen für die Verwendung von Daten- und Metadatenformaten sowie lizenzrechtliche Informationen zum Umgang mit Forschungsdaten heraus.²²

4. Die *technische Infrastruktur* von DARIAH-DE bietet neben Hosting Services wie der Bereitstellung von virtuellen Maschinen und Speicherplatz auch operative Dienste an, darunter eine Authentifizierungs- und Autorisierungsinfrastruktur (AAI), Monitoring und einen Helpdesk. Zusätzlich stehen kollaborative Arbeitsumgebungen zur Verfügung, darunter ein Wiki, der kollaborative Editor Etherpad oder ein Developer Portal,²³ das verschiedene Dienste für Entwickler zur Verfügung stellt, um ein komfortables gemeinsames Arbeiten im Rahmen der Software-Entwicklung zu ermöglichen (z. B. Projektmanagementsysteme, Bugtracker, Code-Repositorien, Continuous Integration-Server). Der PID-Service kann genutzt werden, um Daten mit *Persistenten Identifikatoren (PIDs)* dauerhaft zu referenzieren und zitierfähig zu halten.

DARIAH verfolgt mit dieser breiten Palette unterschiedlichster Angebote die Idee, eine „Architecture of Participation“ bereit zu stellen, die möglichst viele Einzel-forschende, Projektverbünde und Initiativen aus dem Be-

reich der Digital Humanities weltweit nutzen und anreichern können. Für DARIAH-DE geht dieses Konzept auf. Eine Reihe von Forschenden und Projekten greifen bereits auf verschiedene Komponenten der digitalen Infrastruktur zu: Das Projekt *ePoetics*²⁴ erprobt informationstechnologische Methoden wie Textmining, Visualisierung, Text- und Corpus-Viewing, wobei die DARIAH-DE Schema-Registry, die Collection-Registry, das Developer Portal und der EPIC-PID-Service zum Einsatz kommen.

Das Projekt *eCodicology*²⁵ beschäftigt sich mit der automatischen Erkennung von Layoutelementen in Handschriften und knüpft damit an das DARIAH-Cluster „fachwissenschaftliche Annotationen“ an. Das Projekt *Relationen im Raum*²⁶ analysiert und visualisiert räumliche Relationen zwischen Grabmalen jüdischer Friedhöfe zwischen dem 11. und 20. Jahrhundert. Hierbei stützt sich das Projekt nicht nur auf verschiedene Komponenten der technischen Infrastruktur, sondern entwickelt auch den Geo-Browser zusammen mit DARIAH-DE weiter. Zudem ist geplant, ein im Projekt entwickeltes generisches Visualisierungs-Tool in die DARIAH-Infrastruktur zu integrieren und öffentlich zugänglich zu machen.

Das Zusammenspiel von TextGrid und DARIAH-DE

Die drei letztgenannten Projekte kooperieren zugleich auch mit TextGrid – ein Trend mit Tendenz nach oben. Nicht zuletzt aus diesem Grund arbeiten TextGrid und DARIAH-DE eng zusammen. Die Vorhaben weisen trotz ihrer unterschiedlichen Grundausrichtung eine gemeinsame Schnittmenge auf und ergänzen das jeweilige Angebot. Im Rahmen der Öffentlichkeitsarbeit und Außendarstellung betreiben sie gemeinsam das DHd-Blog,²⁷ das sich als Kommunikationsplattform in den Digital Humanities etabliert hat, sowie einen eigenen DHd-Kanal auf YouTube²⁸ mit Videos, Mitschnitten und Tutorials rund um das Thema Digital Humanities. Darüber hinaus erfolgen Kooperationen auf vier weiteren Ebenen:

– *Lehre*: Gemeinsame Workshops, zuletzt beispielsweise im Mai 2015 der DARIAH-Workshop „Store it, share it, use it. Lizenzierung digitaler Forschungsdaten in den Sprach- und Literaturwissenschaften“ und das Text-

¹⁸ <https://de.dariah.eu/veranstaltungen>

¹⁹ <https://de.dariah.eu/expertenkolloquien>

²⁰ <https://de.dariah.eu/fachwissenschaftliche-dienste>

²¹ <https://de.dariah.eu/data-federation-architecture>

²² <https://de.dariah.eu/working-papers-beitraege>

²³ <https://de.dariah.eu/developer-portal>

²⁴ <https://de.dariah.eu/epoetics>

²⁵ <http://www.ecodicology.org/>

²⁶ <https://de.dariah.eu/relationen-im-raum>

²⁷ <http://dhd-blog.org/>

²⁸ <https://www.youtube.com/user/dhdkanal>

Grid Nutzertreffen „Archivieren und Publizieren mit TextGrid“,²⁹ mit gemeinsamem Zielpublikum.

- *Forschung*: Die fachwissenschaftliche Tool-Palette *Digivoy*³⁰ ermöglicht die unmittelbare Nutzung der Voyant-Tools³¹ aus dem TextGrid-Repository heraus durch eine Präprozessierung der dort gespeicherten Daten. Hiermit können die Inhalte des TextGrid-Repository, insbesondere die umfangreiche Digitale Bibliothek³² mit den Werken von rund 600 Autoren, mit den vielfältigen Voyant-Tools quantitativ untersucht, analysiert und visualisiert werden.
- *Forschungsdaten*: Für den DARIAH Use Case *Narrative Techniken und Untergattungen im deutschen Roman* (Aurast et al., 2015) dienen u. a. 400 deutschsprachigen Romane aus der Digitalen Bibliothek von TextGrid als Basis. Ziel ist, mit Hilfe quantitativer Verfahren die Entwicklung narrativer Techniken und literarischer Kategorien zu untersuchen und Best Practices, Beispiel-Workflows und Tutorials zu erstellen, die es TextwissenschaftlerInnen ermöglichen, bereits vorhandene Werkzeuge auf ihre eigenen Daten anzuwenden.
- *Technische Infrastruktur*: Zum Förderende von TextGrid im Mai 2015 wurden zahlreiche technologische Kernkomponenten wie die Nutzerverwaltung und die Repository-Technologien zu DARIAH-DE migriert.

TextGrid/DARIAH-DE Repository

Ein wesentlicher Bereich der gemeinsamen Weiterentwicklung von TextGrid und DARIAH-DE ist das *TextGrid Repository*, das als wissenschaftliches Forschungsdatenarchiv von TextGrid entwickelt und bereitgestellt wurde. Forschungsdaten können im TextGridLab kooperativ bearbeitet, angereichert und schließlich publiziert, oder auch direkt in das Repository eingespielt werden, so dass sämtliche Daten dauerhaft und referenzierbar verfügbar sind.

Die Daten im TextGrid Repository werden mit einem spezifischen TextGrid-Metadatenchema³³ abgebildet. Im TextGrid Repository existieren grundsätzlich zwei Speicherbereiche: ein dynamischer für die Bearbeitung der nicht-öffentlichen Daten, der per Rechteverwaltung abgesichert ist, und ein statischer für publizierte und damit öffentliche Daten. Im statischen Bereich erhält jede Datei einen persistenten Identifikator (PID) und kann so dauer-

haft referenziert werden. Alle öffentlichen Daten sind über die TextGrid Repository Website³⁴ recherchierbar.

- Die drei Säulen der TextGrid Repository-Infrastruktur
- TG-auth* – für Authentifizierung und Autorisierung,
 - TG-search – für Indizierung und Suche, und schließlich
 - TG-crud – für basale Speicheroperationen wie *create*, *retrieve*, *update* und *delete*

basieren auf Technologien wie dem Apache Tomcat-Applikationsserver sowie NOID und EPIC PIDs (TG-crud), OpenLDAP³⁵ und OpenRBAC³⁶ (TG-auth*) sowie ElasticSearch und Sesame RDF (TG-search), um nur einige zu nennen.

Die Anforderungen für ein *DARIAH-DE-Repository* resultieren daraus, dass mit einem überschaubaren Satz an deskriptiven Metadaten Forschungsdaten persistent und referenzierbar gespeichert werden sollen. So ist zum jetzigen Entwicklungsstand ein einfacher Satz von Dublin Core-Metadaten³⁷ erforderlich, um Forschungsdatenobjekte zu beschreiben und in das Repository einzuspielen. Wie auch im TextGrid Repository wird eine Gruppe von Objekten in einer übergeordneten Einheit zusammengefasst: die Kollektion. Über eine grafische Oberfläche im DARIAH-DE Portal (Publish GUI) ist es möglich, Kollektionen zu definieren und zu beschreiben, unter diesen Dateien zu gruppieren, und die Kollektion sowie jede einzelne Datei mit Dublin Core-Metadaten (DC Simple) auszuzeichnen. Ist dieser Prozess abgeschlossen, kann die Kollektion samt zugehörigen Dateien in das DARIAH-DE-Repository eingespielt werden. Auch hier bekommt jedes Objekt einen PID und wird sicher im Repository gespeichert. Nachdem die Kollektion in der DARIAH-DE Collection Registry registriert wurde, werden alle zur Kollektion gehörigen Objekte von der DARIAH-DE generischen Suche³⁸ indexiert und können hierüber recherchiert und verfügbar gemacht werden.

Die technische Infrastruktur des DARIAH-DE-Repositorys wurde weitestgehend von der des TextGrid Repository übernommen, so dass eine gemeinsame Codebasis entstand, deren modularer Charakter weiter entwickelt und ausgebaut wurde (siehe Abb. 5):

- Die Authentifizierung und Autorisierung von TextGrid wurde bereits vor einiger Zeit in die Nutzerverwaltung von DARIAH-DE migriert – DARIAH-DE-NutzerInnen können mit ihrem DARIAH-DE Account auch TextGrid nutzen.

²⁹ <https://de.dariah.eu/ws>

³⁰ <https://de.dariah.eu/digivoy>

³¹ <http://voyant-tools.org/>

³² <https://textgrid.de/digitale-bibliothek>

³³ http://textgridlab.org/schema/textgrid-metadata_2010.xsd

³⁴ <http://textgridrep.de/>

³⁵ <http://www.openldap.org/>

³⁶ http://openrbac.de/en_startup.xml

³⁷ <http://www.dublincore.org/documents/dces/>

³⁸ <http://search.de.dariah.eu>

Abbildung 5: Die Architektur des TextGrid/DARIAH-DE Repositoriums.

- Die Kerndienste TG-crud und TG-publish wurden generalisiert und modularisiert, so dass nun zahlreiche Module von beiden Diensten genutzt werden können.
- Beide Metadaten-Indexe arbeiten mit Elasticsearch als Technologie für die TextGrid- und DARIAH-DE OAI-PMH Data Provider – es können nun Metadaten von TextGrid und DARIAH-DE mit dem OAI Protokoll für Metadata-Harvesting (OAI-PMH) bei den jeweiligen Diensten abgerufen werden.
- Als Identifier- und PID-Implementierung wurde ein PID-Modul entwickelt, das den EPIC2 Handle-Service der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen³⁹ (GWDG) nutzt.
- Sowohl TG-crud als auch sein Pendant für das DARIAH-DE Repository DH-crud kommunizieren über die DARIAH Storage API mit dem unterliegenden DARIAH-DE Repository Storage, dessen Implementation sich um die Replikation sowie die Bit Preservation aller Daten kümmert.

Das DARIAH-DE-Repositorium liegt als Prototyp vor, der sich zurzeit in der Testphase befindet (Funk & Schmunk, 2015). Eine produktive Version ist für das Frühjahr 2016 geplant. Somit ist sichergestellt, dass geistes- und kulturwissenschaftliche Forschungsdaten nun persistent und referenzierbar gespeichert werden können. Die Beschreibung mit einem deskriptiven Metadatenschema ermöglicht auf diese Weise nicht nur die langfristige Speicherung der Daten, sondern vor allem auch die Nachnutzung strukturiert erschlossener Forschungsdaten. Darüber hinaus liefert die Standardisierung und Dokumentation der Schnittstellen die Grundlage für ein Repository-Konzept, das eine Anbindung von DH-Analyse-Tools erlaubt, um Forschungsdaten unmittelbar aus diesen heraus in einem Repository zu speichern.

Ausblick

Ein Blick auf die Digital Humanities-Landschaft in Deutschland zeigt, wie rasant sie über die letzten Jahre hinweg gewachsen ist und wie lebendig und vielfältig sie

³⁹ <http://gwdg.de>

sich gestaltet. Neue Studiengänge wurden eingerichtet,⁴⁰ zahlreiche Projekte initiiert,⁴¹ DH-Konferenzen veranstaltet, der Verband „Digital Humanities im deutschsprachigen Raum“⁴² und die „Zeitschrift für digitale Geisteswissenschaften“⁴³ gegründet.

Eine zentrale Herausforderung wird zukünftig sein, wie die Ergebnisse aller dieser Aktivitäten – wissenschaftliche Erkenntnisse, Methodenkompetenz, Forschungsdaten, digitale Werkzeuge usw. – dauerhaft gesichert und verfügbar gehalten werden können. DARIAH-DE und TextGrid versuchen hier gemeinsam mit verschiedenen Stakeholdern, darunter VertreterInnen von Ministerien und Förderinstitutionen, Lösungen zu finden. Ziel ist, digitale Infrastrukturen nachhaltig aufzustellen und zu verstetigen, um Know-how zu vernetzen sowie Daten und Tools vorzuhalten und für die DH-Community auch in Zukunft bereitstellen zu können.

Literatur

- Aurast, A.; Baumgardt, F.; Held, M.; Pernes, S.; Pielström, S.; Piotrowski, M.; Schöch, C. (2015). Beschreibung der *Use Cases*. DARIAH-DE Projektbericht R 5.2.1. <https://dev2.dariah.eu/wiki/download/attachments/14651583/R%205.2.1%20%E2%80%93%20Beschreibung%20der%20Use%20Cases.pdf?version=1&modificationDate=1424423806974&api=v2> [18.08.2015].
- Funk, S. E.; Schmunk, S. (2015). DARIAH-DE Repositorium – Prototyp. DARIAH-DE Projektbericht M 4.3.2.1. <https://dev2.dariah.eu/wiki/download/attachments/14651583/M%204.3.2.1-DARIAH-Rep%20ositorium-Prototyp-final.pdf?api=v2> [18.08.2015].
- Hervay, L.; Neweling, S.; Schwiigelshohn, U. (Hrsg.) (2009). D-Grid. Die Deutsche Grid-Initiative. Vorstellung der Projekte. Veröffentlichung im Rahmen des „D-Grid All Hands Meeting“, März 2009. <http://www.d-grid-ggmbh.de/downloads/BroschuereAHM2009.pdf> [18.08.2015].
- Neuroth, H.; Rapp, A.; Söring, S. (Hrsg.) (2015). TextGrid: Von der Community – für die Community. Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften. Universitätsverlag Göttingen & Verlag Werner Hülsbusch, Glückstadt. http://univlag.uni-goettingen.de/handle/3/Neuroth_TextGrid [18.08.2015].
- Neuroth, H.; Kerzel, M.; Gentzsch, W. (Hrsg.) (2007). Die D-Grid Initiative. Universitätsverlag Göttingen. http://webdoc.sub.gwdg.de/univlag/2007/D-Grid_de.pdf [18.08.2015].
- Reiche, R.; Becker, R.; Bender, M.; Munson, M.; Schmunk, S.; Schöch, C. (2014). Verfahren der Digital Humanities in den Geistes- und Kulturwissenschaften. *DARIAH-DE Working Papers* Nr. 4, Göttingen. URN: urn:nbn:de:gbv:7-dariah-2014-2-6 [18.08.2015].

Dr. Mirjam Blümm

Georg-August-Universität Göttingen
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Abteilung Forschung Entwicklung
Platz der Göttinger Sieben 1
37070 Göttingen
Deutschland
bluemm@sub.uni-goettingen.de

Mirjam Blümm ist Programmleiterin für Digital Humanities in der Abteilung Forschung & Entwicklung der Niedersächsischen Staats- und Universitätsbibliothek Göttingen. Sie ist promovierte Germanistin (2009, Universität Würzburg) und wissenschaftliche Bibliothekarin. Zu ihren Arbeitsschwerpunkten zählen: Forschungsinfrastrukturen (DARIAH-DE) und Virtuelle Forschungsumgebungen (TextGrid).

Dipl.-Inform. Stefan E. Funk

Georg-August-Universität Göttingen
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Abteilung Forschung Entwicklung
Platz der Göttinger Sieben 1
37070 Göttingen
Deutschland
funk@sub.uni-goettingen.de

Stefan E. Funk arbeitet seit 2004 als Entwickler in der Niedersächsischen Staats- und Universitätsbibliothek Göttingen und war unter anderem als Mitarbeiter der Firma DAASI International GmbH maßgeblich an der Entwicklung des TextGrid Repository beteiligt. Im Rahmen von DARIAH-DE ist er nun verantwortlich für die Architektur des DARIAH-DE Repositoriums. Er ist studierter Informatiker (2004, Uni Göttingen) und erwirbt momentan berufsbegleitend den Master in Bibliotheks- und Informationswissenschaft (MALIS) an der Fachhochschule Köln.

Sibylle Söring, M. A.

Georg-August-Universität Göttingen
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Abteilung Forschung Entwicklung
Platz der Göttinger Sieben 1
37070 Göttingen
Deutschland
soering@sub.uni-goettingen.de

Sibylle Söring ist seit 2011 an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen in der Konsortialleitung der Forschungsinfrastruktur-Projekte TextGrid und DARIAH-DE tätig. Arbeitsschwerpunkte der Germanistin, die verschiedene, auch internationale geisteswissenschaftliche Forschungsvorhaben im digitalen Medium betreut hat und neue Projekte beim Start in die Nutzung digitaler Werkzeuge und Services begleitet, bilden u. a. Digitale Editionen.

⁴⁰ <https://de.dariah.eu/dh-standorte-in-d>

⁴¹ <https://de.dariah.eu/projekte>

⁴² <http://www.dig-hum.de/>

⁴³ <http://www.zfdg.de/>