Stefan Schmunk* und Stefan E. Funk

Das DARIAH-DE- und das TextGrid-Repositorium: Geistes- und kulturwissenschaftliche Forschungsdaten persistent und referenzierbar langzeitspeichern

DOI 10.1515/bfp-2016-0020

Zusammenfassung: DARIAH-DE entwickelt seit 2011 eine modular auf Diensten basierende Forschungsdaten-Föderationsarchitektur. Eines der zentralen Kernelemente ist hierbei das DARIAH-DE-Repositorium, mit dem Geistes- und Kulturwissenschaftler, aber auch Forschungsprojekte, ihre erzeugten, angereicherten und auch erhobenen Daten speichern, mit persistenten IDs und dadurch referenzierbar in einer Repository-Umgebung ablegen und für die Nachnutzung bereit stellen können.[1]

Schlüsselwörter: DH-Curricula; Forschungsdaten; Repositorien

The DARIAH-DE Repository and the TextGrid Repository: Persistent and Long-Term Preservation of Research Data in the Arts and Humanities

Abstract: Since 2011, DARIAH-DE is developing federation architecture for research data that consists of module based (web) services. One of the core elements is the DARIAH-DE Repository that allows scholars and also research projects from the humanities to safely store their research data. Providing persistent identifiers the repository guarantees access to the data and so ensures re-use and enables sustainability.

Keywords: Digital humanities; research data; repositories

Inhalt

1 Einleitung

Im Rahmen von DARIAH-DE widmet sich das Cluster „Wissenschaftliche Sammlungen und Forschungsdaten“[2] nicht nur methodischen und konzeptionellen Fragen des Umgangs, der Generierung, der Nutzung[3] und der Anreicherung von digitalen Forschungsdaten, sondern ein zentraler Teil der Tätigkeiten besteht in der Entwicklung und Realisierung einer Repository-Lösung für geistes- und kulturwissenschaftliche Forschungsdaten.[4]

Das DARIAH-DE-Repositorium steht DARIAH-DE assoziierten Forschungsprojekten zur Verfügung, wie derzeit beispielsweise TextGrid[5] und darüber hinaus Forschern sowie Forschungsprojekten, die ihre Forschungsdaten persistent, referenzierbar und langzeitarchiviert speichern und Dritten zur Verfügung stellen wollen. Ebenfalls sind Wissenschaftler an Universitäten und Forschungseinrichtungen adressiert, die in Forschungsprojekten entstande-

1 Der Beitrag basiert auf einem im Jahr 2015 erschienenen Report der beiden Autoren zum DARIAH-DE-Repositorium, siehe: https://wiki.de.dariah.eu/download/attachments/14651583/M%204.3.2.1-DARIAH-Repositorium-Prototyp-final.pdf?version=1&modificationDate=1430220062670&i=v2.

*Kontaktperson: Dr. Stefan Schmunk,
schmunk@sub.uni-goettingen.de
Stefan E. Funk, funk@sub.uni-goettingen.de

2 Siehe: https://wiki.de.dariah.eu/display/publicde/Cluster+4%3A+Wissenschaftliche+Sammlungen.

3 Aber auch Nutzungsmöglichkeiten, wie z. B. lizenzrechtlichen Fragen, siehe: https://de.dariah.eu/lizenzen.

4 Vgl. Forschungsdaten in DARIAH-DE: https://de.dariah.eu/forschungsdaten.

5 Vgl. TextGrid: Digital edieren – forschen – archivieren: http://textgrid.de/.

ne, erhobene, erfasste und/oder generierte Forschungsdaten langfristig im Rahmen einer Repository-Lösung speichern wollen. Hierbei steht vor allem der einfache und nutzerorientierte Zugang (Usability) von Fachwissenschaftlern zu einer Langzeitspeicherung von Forschungsdaten im Vordergrund. Das DARIAH-DE-Repositorium ermöglicht es, Forschungsdaten zu speichern, mit Metadaten zu versehen, diese durch die Generische Suche aufzufinden und vor allem durch die Nutzung von EPIC-PIDs[6] eine permanente (maschinenlesbare) Referenzierung zu gewährleisten.

Um dies zu erreichen, arbeiten DARIAH-DE und TextGrid, das aus der Virtuellen Forschungsumgebung TextGrid Laboratory (TextGridLab)[7] und dem TextGrid Repository (TextGridRep)[8] besteht, zusammen. Das DARIAH-DE-Repositorium stützt sich auf die Codebasis des TextGrid Repository und wurde mit verschiedenen Service-Instanzen und unterschiedlichen an das DARIAH-DE-Repositorium angepassten Modulen mit weiteren Funktionen wie Speicher- und AAI-Zugriff implementiert.

2 Forschungsdaten-Föderationsarchitektur

Im Projekt DARIAH-DE wurde in den vergangenen Jahren u. a. eine Authentifizierungs- und Autorisierungsinfrastruktur (AAI)[9] und die DARIAH-DE Storage API für die Speicherung von Forschungsdaten auf Bit Preservation Level aufgebaut, so dass Forschungsdaten zwischen den beteiligten Rechenzentren repliziert werden können. Dadurch ist sichergestellt, dass die digitale Forschungsinfrastruktur nicht nur als Speicherort für statische Daten verwendet werden kann, sondern über mehrere Standorte verteilt gespeichert werden kann. Diese sind auf diese Weise öffentlich zugänglich, zitierfähig und langzeitarchiviert. Darüber hinaus besteht ebenso die Möglichkeit, 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 (maschinenlesbar) zugegriffen werden und zugleich werden alle Forschungsdaten mit EPIC-PIDs versehen, so dass andere Tools und Services diese nachnutzen können.[10] Zu diesen Tools gehört beispielsweise die DARIAH-DE Collection Registry.[11] Sie enthält Informationen über beliebige Forschungsdaten-Repositorien und deren Sammlungsbeschreibungen. Die in DARIAH-DE entwickelte Generische Suche[12] indiziert die Metainformationen der Sammlungen der Collection Registry und bietet so einen nutzerfreundlichen 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.

Die DARIAH-DE Forschungsdaten-Föderationsarchitektur weist einen modularen Aufbau auf. Alle vorhandenen Tools und Services sind auch einzeln nutzbar, beispielsweise in anderen Projektkontexten oder in anderen Architekturumgebungen. Die Implementierung und der Betrieb weiterer Instanzen des Repositoriums oder auch der Collection Registry durch Dritte ist technologisch möglich und mit dem Ziel verbunden, mehrere betriebene Instanzen miteinander zu verknüpfen. Zugleich – und dies ist die Stärke dieses architektonischen Ansatzes – können die über die Generische Suche such- und findbaren Forschungsdaten auch aus anderen Registries bzw. Repositorien stammen. Auf diese Weise sollen perspektivisch beispielsweise Europeana,[13] die bibliographischen Informationen der Deutschen Digitalen Bibliothek (DDB)[14] oder Repositorien von CLARIN[15] eingebunden werden. Dieser föderative Ansatz bietet so die Möglichkeit, sowohl die einzelnen von DARIAH-DE entwickelten und betriebenen Komponenten als „Gesamtsystem“ zu nutzen, zugleich haben Projekte und Einrichtungen über definierte APIs und Metadatenstandards auch die Möglichkeit, ihre eige-

6 Vgl. Nachhaltige Referenzierung von Digitalen Objekten mit Hilfe von persistenten Identifikatoren (PID): https://de.dariah.eu/pid-service. Darüber hinaus ist geplant, in einer zweiten Ausbauphase eine Referenzierung mittels DataCite DOIs umzusetzen: https://www.datacite.org.

7 Vgl. TextGrid – Download und Installation: https://www.textgrid.de/registrierungdownload/download-und-installation/.

8 Vgl. TextGrid Repository: http://www.textgridrep.de/.

9 Vgl. DARIAH-DE Autorisierungs- und Authentifizierungs-Infrastruktur: https://de.dariah.eu/aai.

10 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.

11 Vgl. DARIAH-DE – Collection Registry: https://de.dariah.eu/collection-registry.

12 Vgl. DARIAH-DE – Generische Suche: https://de.dariah.eu/generische-suche.

13 Vgl. Europeana Collections: http://www.europeana.eu/portal/.

14 Vgl. Deutsche Digitale Bibliothek: https://www.deutsche-digitale-bibliothek.de/.

15 Vgl. Clarin-D – Forschungsinfrastruktur für die Geistes- und Sozialwissenschaften: http://www.clarin-d.de/de/.

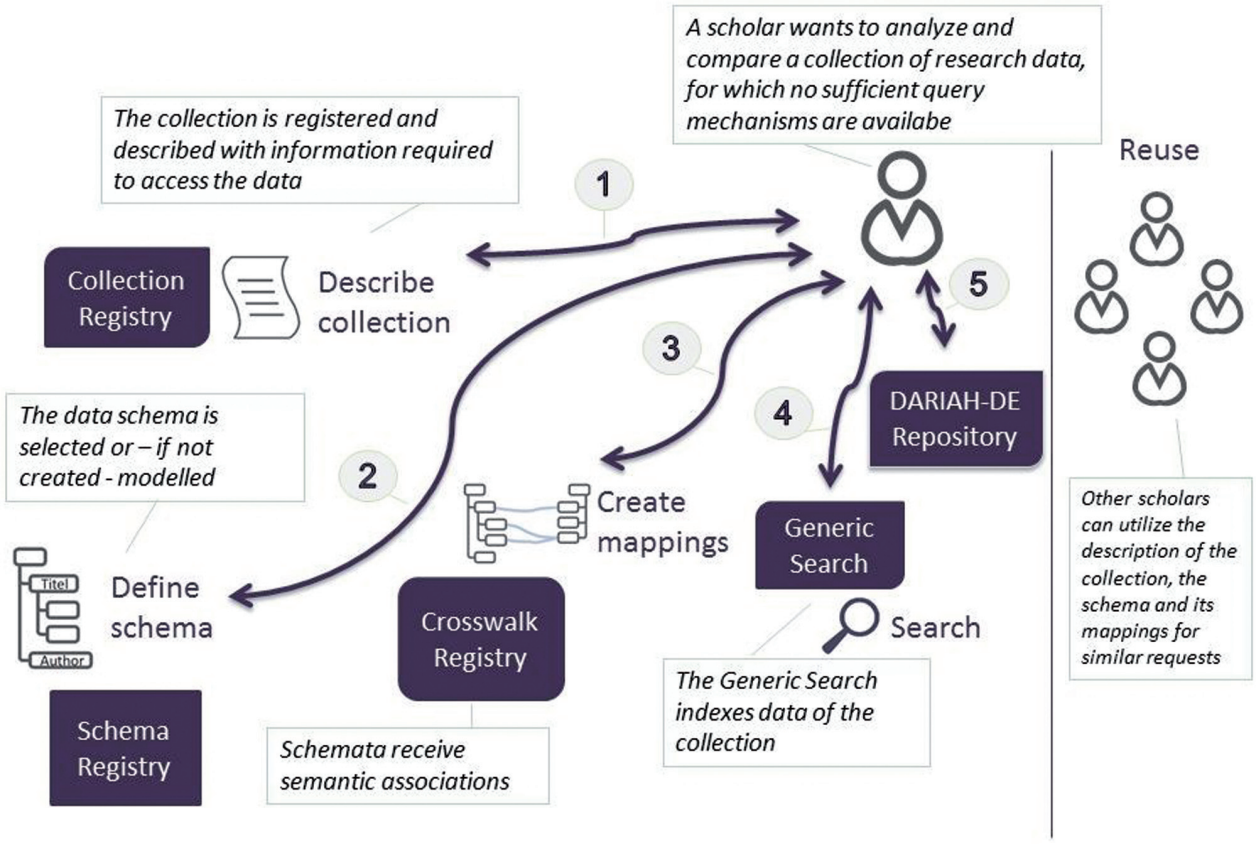

Abb. 1: Aufbau der DARIAH-DE Forschungsdaten-Föderationsarchitektur und Zusammenspiel der einzelnen technologischen Komponenten[16]

nen Sammlungsbeschreibungen und Forschungsdaten für Dritte nutzbar einzubinden. Hierbei zielt die Forschungsdaten-Föderationsarchitektur auf folgende Punkte: [17]

- Indizierung und Verzeichnung von Forschungsdaten und Sammlungsinformationen
- Ermöglichen eines nachhaltigen und persistenten Zugriffs, insbesondere unter der Perspektive von Nachnutzungsmöglichkeiten
- Entwicklung von Werkzeugen und Diensten, die das Suchen, Finden und Vergleichen ermöglicht
- Vergleichende Suchfunktionalitäten für heterogen strukturierte Metadaten, Sammlungen und digitale Archive

Im Rahmen dieses Artikels werden die Architektur und die verwendeten Technologien des DARIAH-DE-Repositoriums beschrieben. Darüber hinaus wird auf den Publikationsprozess von Forschungsdaten und des der Sammlungsbeschreibung zugrunde liegenden Datenmodells eingegangen. Die gesamten zweijährigen Entwicklungstätigkeiten basieren einerseits auf umfangreichen und engen Abstimmungen mit Fachwissenschaftlern, deren Anforderungen die Basis der Funktionalitäten und Umsetzungen sind, andererseits auf umfangreichen Vorarbeiten, die im Rahmen von TextGrid durchgeführt wurden.

3 Das DARIAH-DE-Repositorium

Die Kernkomponenten des TextGrid Repository sind die drei Dienste TG-auth*,[18] TG-crud[19] und TG-search.[20] Diese sind für die Authentifizierung und Autorisierung der Nutzer verantwortlich, für grundlegende Speicheroperationen

16 Vgl. DARIAH-DE Data Federation Architecture: https://de.dariah.eu/data-federation-architecture. Eine ausführliche Beschreibung und Darstellung dieses Konzepts und der einzelnen zugrundeliegenden Ansätze findet sich im Beitrag von Tobias Gradl und Andreas Henrich Beitrag in diesem Heft.

17 Vergleiche für eine ausführliche Beschreibung den Aufsatz in diesem Sonderheft von Tobias Gradl und Andreas Henrich.

18 Vgl. TG-auth*: http://textgridlab.org/doc/services/submodules/tg-auth/docs/index.html.

19 Vgl. TG-crud: http://textgridlab.org/doc/services/submodules/tg-crud/docs/index.html.

20 Vgl. TG-search: http://textgridlab.org/doc/services/submodules/tg-search/docs/index.html.

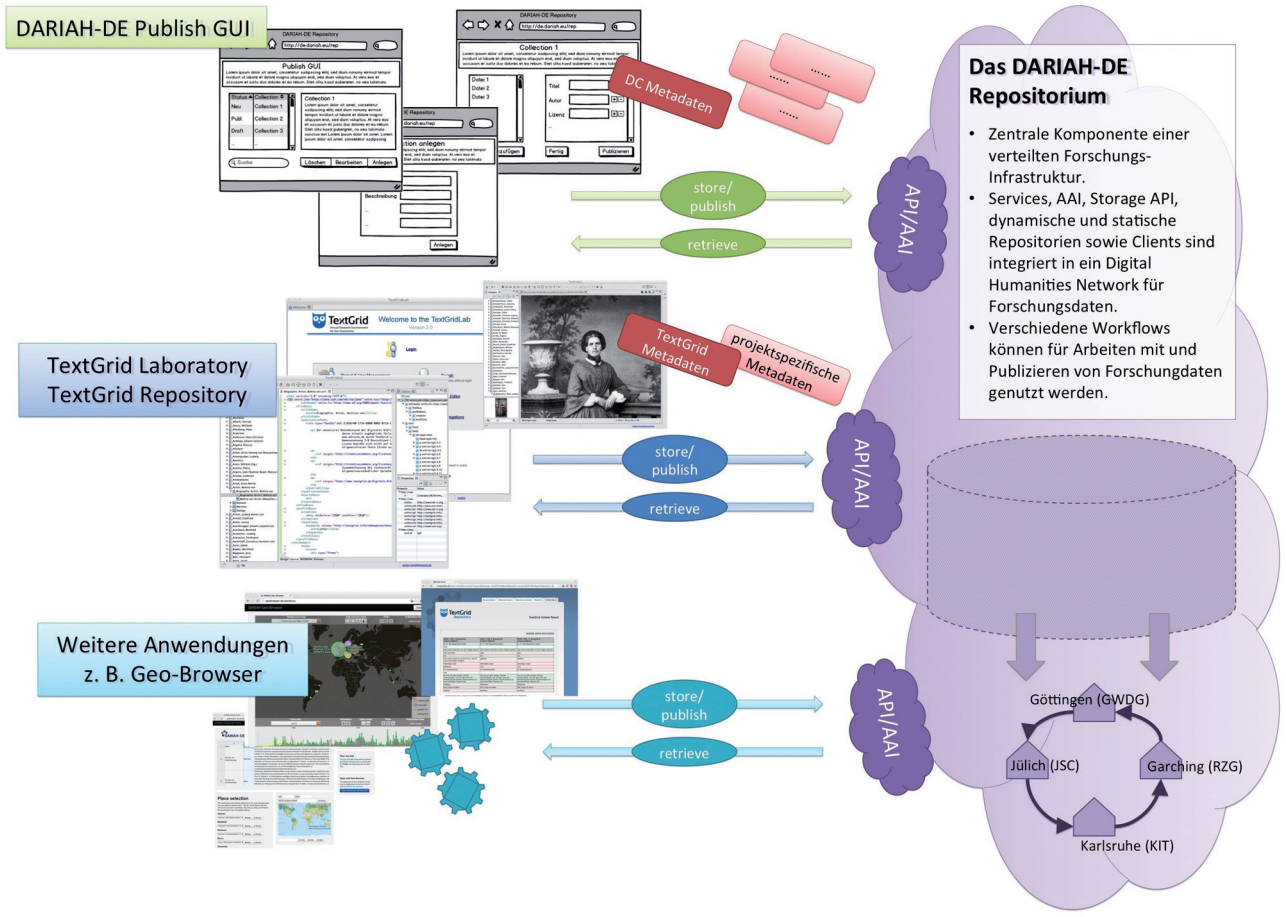

Abb. 2: Das DARIAH-DE-Repositorium und damit verbundene Dienste

sowie für die Indizierung und Suche über die Daten des Repository. Weitere Dienste sind TG-noid,[21] eine Implementierung zum Erzeugen und Verwalten von Identifiern der internen TextGrid URIs sowie TG-publish,[22] der für die Publikation und deren Workflows verantwortlich ist.

Die Dienste TG-crud und TG-publish wurden so erweitert und modularisiert, dass im DARIAH-DE-Repositorium derselbe Programmcode verwendet werden kann. So konnte eine neue Instanz für das Repositorium aufgesetzt werden und ist entsprechend konfigurierbar. Die Authentifizierung erfolgt seit längerem schon über die DARIAH AAI, hier kommt Shibboleth zum Einsatz. Für die Autorisierung wird im TextGrid Repository ein rollenbasiertes Zugriffssystem genutzt (RBAC),[23] das auch bald für das DARIAH-DE-Repositorium einsetzbar ist, hier wurde ebenfalls der TextGrid-Code nachgenutzt und erweitert. Die Dienste TG-pid[24] und TG-oaipmh[25] werden in verschiedenen Instanzen ebenfalls von beiden Repositorien genutzt, genauso wie der Metadaten-Index, hier wird bei beiden Repositorien ElasticSearch[26] für die Indizierung der Daten genutzt.

Parallel zu den Entwicklungen des DARIAH-DE-Repositoriums wurde die gesamte Architektur des TextGrid Repository auf die DARIAH-DE IT-Architektur (Storage, VMs, Monitoring, AAI, Liferay, Puppet usw.) umgezogen, so dass nun zwei Repository-Lösungen in ein und demselben technologischen Environment zur Verfügung stehen. Hierbei fokussiert das TextGrid Repository auf in XML ausgezeichnete (Text-)Daten und bietet entsprechende maschinenlesbare Schnittstellen an, wohingegen das DARIAH-DE-Repositorium vor allem auf andere nicht-XML

21 Vgl. NOID: https://metacpan.org/pod/distribution/Noid/noid.
22 Vgl. TG-publish: http://textgridlab.org/doc/services/submodules/kolibri/kolibri-tgpublish-service/docs/index.html.
23 Vgl. OpenRBAC: http://www.openrbac.de/en_startup.xml.
24 Vgl. TG-pid: http://textgridlab.org/doc/services/submoduls/tg-pid/docs/index.html.
25 Vgl. TG-oaipmh: http://textgridlab.org/doc/services/submodules/oai-pmh/docs/index.html.
26 Vgl. ElasticSearch: https://www.elastic.co/products/elasticsearch.

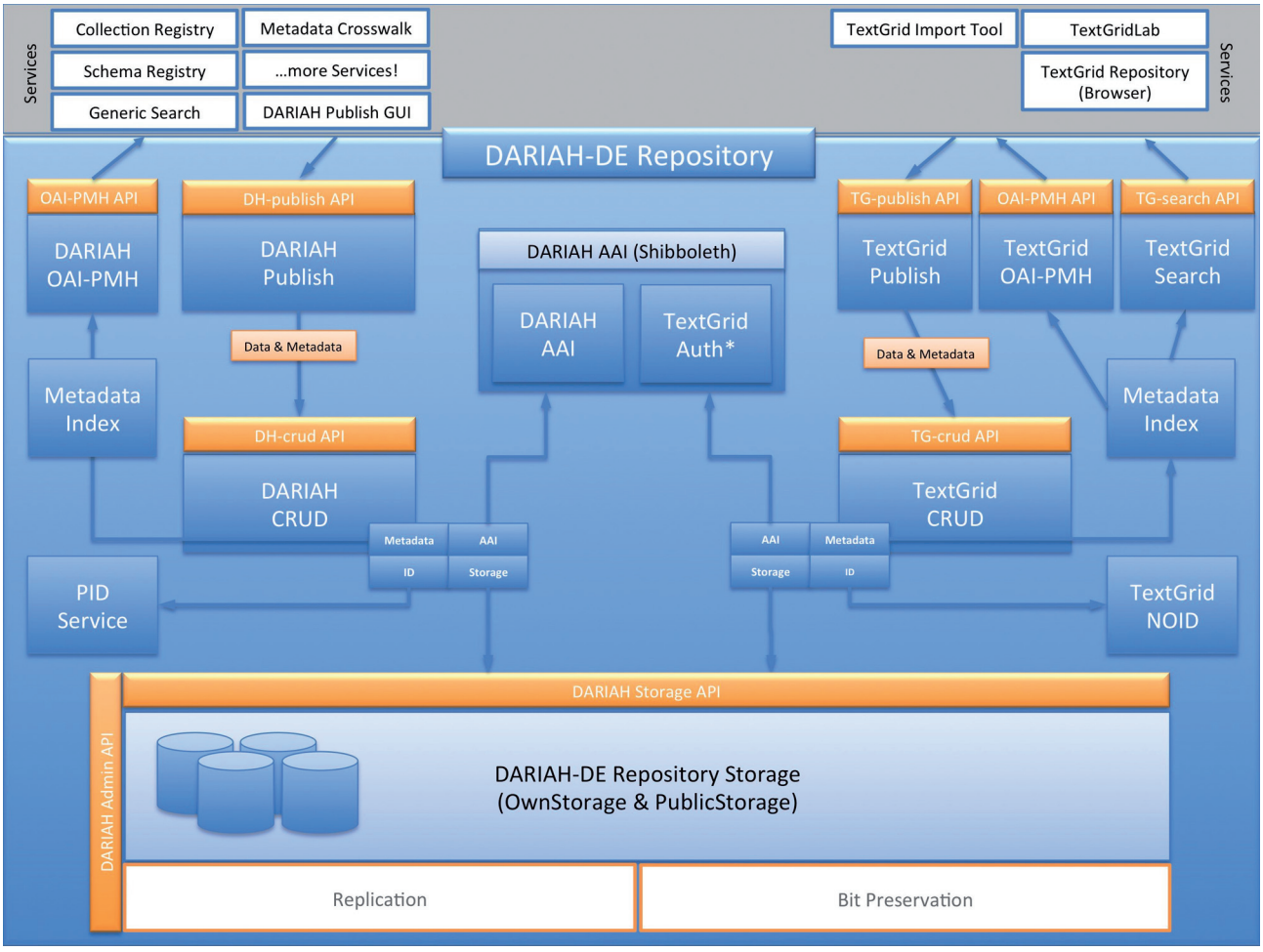

Abb. 3: Architekturvergleich DARIAH-DE und TextGrid Repository

Dateiformate abzielt. Diese Differenzierung ist vor allem deshalb notwendig, da basierend auf den in den Repositorien gespeicherten Daten- und Dateiformaten weitere technologische Aspekte abhängig sind. Hierzu zählen spezifische Schnittstellen, die genutzten Daten- und Dateiformate, aber auch Aspekte und Prozesse der Datenkuration und Langzeitarchivierung.

Das DARIAH-DE-Repositorium ermöglicht es Forschenden, die sich bei der DARIAH-DE AAI authentifiziert haben und für das Repositorium autorisiert sind, ihre digitalen Objekte bzw. Datensammlungen und Kollektionen nachhaltig und sicher zu archivieren. Der Ingest-Prozess wird über ein Web-Interface, die DARIAH Publish GUI, vorgenommen und kann auf diese Weise durch die Nutzung eines beliebigen Browsers erfolgen. Hierzu muss zuerst eine Kollektion vom Forschenden über die Publish GUI angelegt und mit Metadaten ausgezeichnet werden. Dieser Kollektion kann in einem zweiten Schritt eine beliebige Anzahl an Dateien zugeordnet werden, die über die Publish GUI hochgeladen und ebenfalls mit Metadaten ausgezeichnet werden können. Der Vorteil dieser Vorgehensweise liegt auf der Hand. Hierdurch haben die Wissenschaftler die Möglichkeit, unmittelbar von ihrem Rechner aus – dem Ort, an dem zumeist geisteswissenschaftliche Forschung, wenn nicht ausschließlich durchgeführt, dann doch zumindest die Ergebnisse niedergeschrieben werden – ihre Forschungsdaten im Repositorium zu speichern. Die eigentliche Publikation findet allerdings nicht durch den Upload-Prozess statt, sondern erst zu einem späteren Zeitpunkt, an dem vom Forschenden eine bewusste Entscheidung zur Publikation der Sammlungen und der dazugehörigen Forschungsdaten getroffen wird. Die Daten werden dann unmittelbar während des Publikationsprozesses per Persistent Identifier (PID) referenziert, damit öffentlich zugänglich, und die Kollektion wird in der DARIAH-DE Collection Registry eingetragen und ist somit nachweisbar. Sobald die Kollektion selbst über die Collection Registry publiziert wurde, sind die Daten mit der Generischen Suche von DARIAH-DE recherchierbar. Auf diese Weise werden die Forschungsdaten nicht nur gespeichert

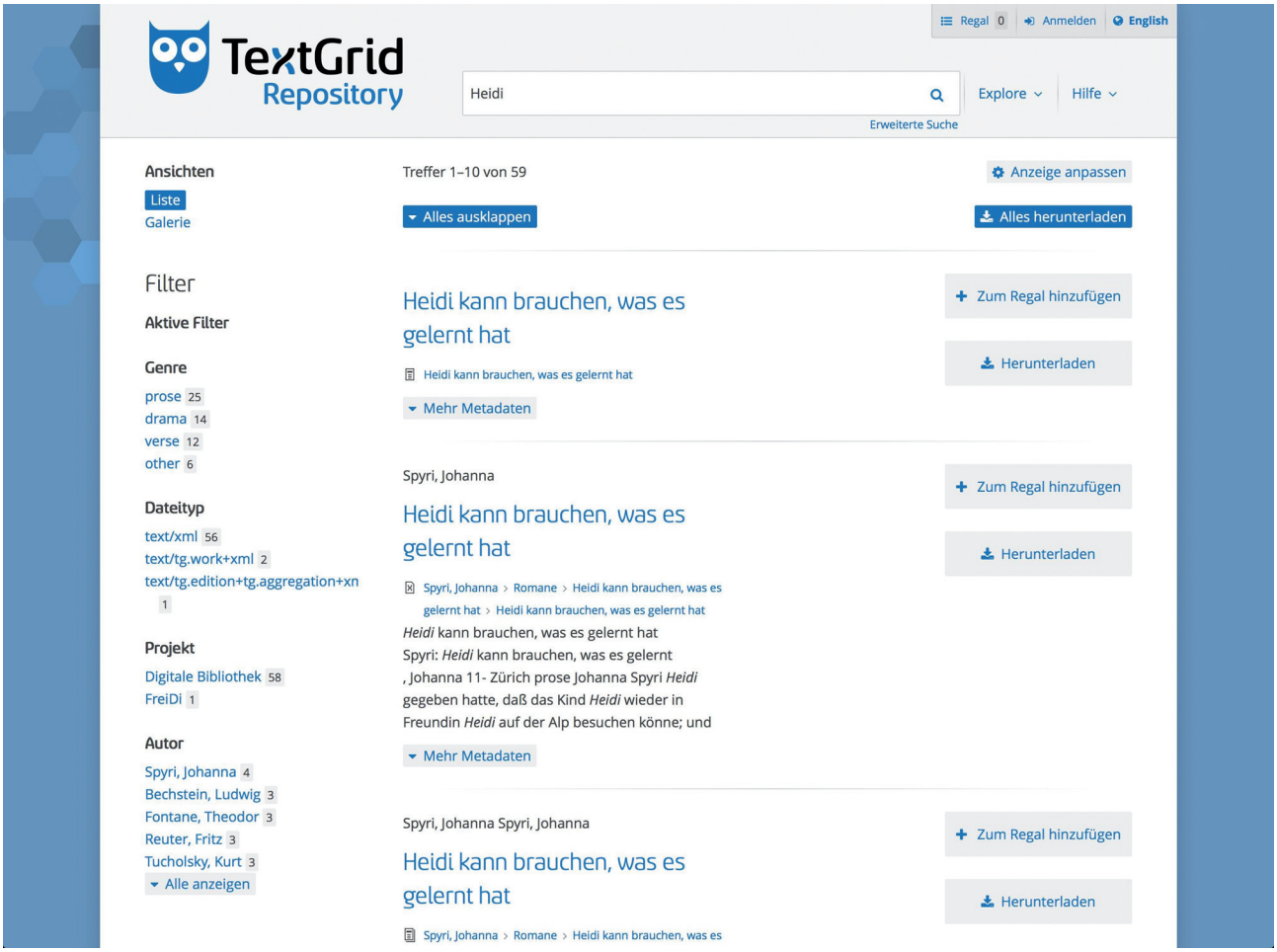

Abb. 4: Das TextGrid Repository nach dem Relaunch im Februar 2016

und archiviert, sondern auch die dazugehörigen Forschungskontexte in Form von Kollektionen angelegt und persistent und referenzierbar gespeichert.

Dieser konzeptionelle Ansatz, Forschungsdaten spezifischen Kollektionen zuzuordnen, hat noch einen weiteren Vorteil: Auf diese Weise können einzelne Daten – beispielsweise eine Publikation von Goethes unterschiedlichen Kollektionen – in diesem Sinne Forschungsprojekten – zugeordnet werden, obwohl sie zugleich nur einmal gespeichert werden müssen. Neben diesem physikalischen Vorteil können sich Nutzer zudem anzeigen lassen, in welchen Kollektionen die Daten bereits genutzt wurden und auf diese Weise sehen, in welchen Forschungskontexten diese bislang verwendet wurden. Forschungskontexte werden dadurch digital dargestellt, so dass auch eine Überprüfung der Validität und die Reliabilität von Forschungsergebnissen möglich sind.

Eine zentrale Anforderung der Forschenden war zudem, dass durch das Konzept von Kollektionen Beziehungen zwischen digitalen Objekten – als digitales Objekt wird hier eine Datei samt ihrer zugehörigen beschreibenden Metadaten verstanden – abgebildet werden können, um auf diese Weise kontextualisierende Informationen zur Entstehung und insbesondere zur Nutzung der Daten abbilden zu können. Aus diesem Kontext kann auch auf eine Metadaten-Eingabe und Metadaten-Validierung nicht verzichtet werden.

Der Workflow für einen Import in das Repository wird im Folgenden kurz dargestellt. Die Authentifizierung erfolgt über die DARIAH AAI[27] und muss von allen Services bedient werden. Auf diese Weise können alle Nutzer, die über einen DARIAH-DE-Account verfügen und für die Nutzung des DARIAH-DE-Repositoriums freigeschaltet wurden, das Repositorium nutzen und Daten speichern.

27 Vgl. DARIAH Authorization and Authentication Infrastructure: https://dev2.dariah.eu/wiki/download/attachments/6783645/DARIAH-AAI-Concept-v0.3a.pdf?version=1&modificationDate=1328883126670&api=v2.

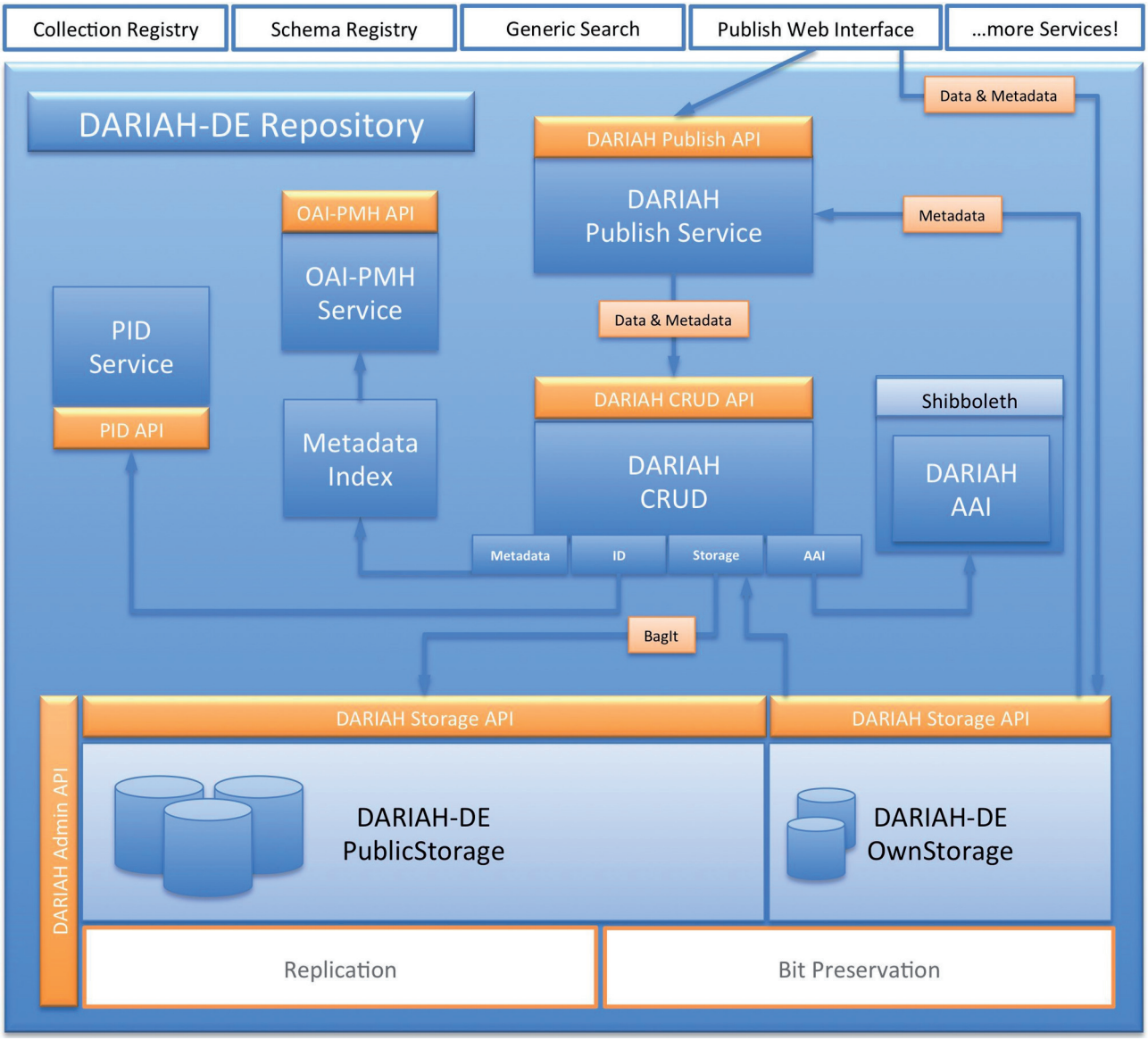

Abb. 5: Architektur des DARIAH-DE-Repositoriums

4 Publish Web Interface

Der Nutzer erzeugt über das DARIAH-publish-Web-Interface (Publish GUI) – implementiert als Liferay-Portlet – eine Kollektion, wählt einzuspielende Daten aus und versieht jedes einzelne Objekt, auch oder gerade die Kollektion selbst, mit DC-Metadaten.[28] Die Publish GUI liefert die Objekte samt Metadaten per API an den DARIAH-publish-Service. Die Dateien werden von der Publish GUI in den OwnStorage – eine Implementierung der DARIAH Storage API[29] – von DARIAH gespeichert, auf den zunächst nur der jeweilige Forschende Zugriff hat. Eine Datei mit allen nötigen Daten und Metadaten wird von der Publish GUI an den Publish-Service weitergegeben. Die einzuspielenden Dateien können lokal vom Rechner des Forschenden stammen. Der Rückgabewert des Publish-Service gibt Aufschluss über den Status des Publikationsvorgangs. Hierfür sind folgende Status möglich:

28 Im DARIAH-DE Repositorium werden zunächst DC-Simple Metadaten verwendet, die aus 15 Feldern bestehen. Vgl. Dublin Core Metadata Element Set, Version 1.1: http://www.dublincore.org/documents/dces/.

29 Vgl. DARIAH Storage API – A Basic Storage Service API on Bit Preservation Level: http://hdl.handle.net/11858/00-1734-0000-0009-FEA1-D.

- DRAFT – neu angelegt bzw. in Bearbeitung innerhalb der Publish GUI,
- RUNNING – Publikation ist gerade in Arbeit,
- ERROR – Fehler beim Publikationsprozess,
- PUBLISHED – im DARIAH-DE-Repositorium publiziert,
- REGISTERED – in der Collection Registry registriert und von der Generischen Suche indiziert.

Die Publish GUI liefert nach erfolgreicher Publikation einen direkten Link auf die Kollektion in der Collection Registry, sowie auf den PID der Kollektion.

5 DARIAH-publish-Service

DARIAH-publish ist ein Workflow-Service, der verschiedene Schritte im Rahmen der Publikation ausführt. Es werden u. a. die Metadaten validiert, Referenzen auf Objekte innerhalb der einzuspielenden Kollektion von Dateipfaden auf Identifier umgeschrieben und technische Metadaten generiert. Schließlich werden, nach dem Erzeugen der Kollektions-Datei, alle referenzierten Daten samt Metadaten aus dem OwnStorage an den DARIAH-crud weitergegeben.

Wird der Aufruf des Publish Services erfolgreich beendet, ist die Kollektion der Nutzer erfolgreich publiziert worden. Dies bedeutet zunächst, dass

- alle Dateien in den PublicStorage geschrieben wurden, wo sie öffentlich zugänglich sind,
- alle Dateien einen PID[30] erhalten haben und nachhaltig referenzierbar sind,
- die Kollektion und ihre Inhalte über den DARIAH-OAI-PMH-Service abfragbar sind und
- für den Forschenden ein Entwurf einer Sammlungsbeschreibung in der Collection Registry angelegt wurde. Dieser kann nun noch um weitere Metadaten ergänzt und schließlich dort veröffentlicht werden. Nach diesem Schritt gilt die Kollektionsbeschreibung als publiziert und kann von der Generischen Suche per OAI-PMH-Schnittstelle indiziert werden. Erst dann sind die Daten auch über die Generische Suche recherchierbar.

30 Als PIDs werden hier die Handles des EPIC-Konsortiums genutzt (EPIC API v2), vgl. http://www.pidconsortium.eu/ und http://epic.gwdg.de/wiki/index.php/EPIC:API.

6 DARIAH-crud-Service

Der DARIAH-crud-Service ist der Speicher-Service des DARIAH-DE-Repositoriums und stellt, genau wie der TG-crud-Service, grundlegende Speicher-Operationen zur Verfügung: Create, Retrieve, Update und Delete. Es sind zwei Instanzen des DH-crud-Services in Betrieb. Die eine ist nur intern zu erreichen (z. B. vom DARIAH-publish-Service), diese ist vornehmlich für die Erzeugung und Verwaltung von Daten zuständig (Create und evtl. Delete für administrative Zwecke). Hier werden die Metadaten und Daten aller Objekte

- im DARIAH-DE PublicStorage gespeichert,
- die Metadaten in die Indexdatenbank ElasticSearch für einen späteren Abruf per OAI-PMH Service eingetragen und
- ein PID erzeugt, der jedes Objekt eindeutig und dauerhaft identifiziert und referenziert.

Die zweite Instanz, die nur lesenden Zugriff auf die Daten erlaubt, ist von extern zu erreichen und gibt Daten- sowie Metadaten der gespeicherten Objekte heraus (Read und ReadMetadata).

7 Collection Registry

Die Publish GUI sendet bei erfolgreichem Aufruf des Publish-Services den Metadatensatz der Kollektion als Entwurf einer Sammlungsbeschreibung an die Collection Registry. Dieser Schritt ist zum einen nötig, um den Wissenschaftlern die vollständige Kontrolle über die Registrierung der Kollektion zu geben – und damit über die Entscheidung, ihre Kollektion über die Generische Suche verfügbar zu machen –, zum anderen sind verschiedene Angaben zur Kollektion nötig, die nicht schon in der Publish GUI abgefragt bzw. nicht automatisiert an die Collection Registry weitergegeben werden können.

8 Generische Suche

Sobald die Kollektion in der Collection Registry fertig beschrieben und veröffentlicht wurde (dort wird u. a. die URL zur OAI-Schnittstelle festgelegt), kann die Generische Suche die Daten indexieren und über die Webseite recherchierbar machen. Der OAI-PMH-Data-Provider kann öffentlich nach neuen Datensätzen des DARIAH-Repositoriums – nach dem OAI-PMH-Protokoll – angefragt werden. Dieser nutzt für seine Antworten den ElasticSearch-Index, der vom DARIAH-crud-Service gefüllt wird. So kann die Gene-

rische Suche alle Daten des Repositoriums indexieren und allen Nutzern zur Verfügung stellen. Es werden nur die Daten indiziert, die in öffentlichen Kollektionen der Collection Registry publiziert sind.

9 Schluss

Das DARIAH-DE-Repositorium basiert auf den Anforderungen von Fachwissenschaftlern, die eine technische Möglichkeit einforderten, um Forschungsdaten aus Forschungsprojekten dauerhaft und referenzierbar speichern zu können. Der Fokus lag hierbei insbesondere darauf, eine technologische Infrastruktur zu entwickeln, die einerseits modular aufgebaut ist und zugleich persistente Speichermöglichkeiten bietet. Auf diese Weise wurde sichergestellt, dass die umgesetzten Publikationsprozesse generischen Charakter aufweisen und u. a. hinsichtlich Usability und der graphischen Nutzerführung in unterschiedlichen disziplinären Kontexten genutzt werden können. Das DARIAH-DE-Repositorium wird im Frühjahr 2016 den Produktivbetrieb in der Version 1.0 aufnehmen und zugleich wird, wie beschrieben, an der Implementierung weiterer Funktionalitäten gearbeitet. Gerade die Kombination von gleichen technologischen Komponenten für die Entwicklungen den Betrieb zweier Repositorien, die sich auf unterschiedliche Daten- und Dateiformate fokussieren, zeigt den Bedarf und zugleich die grundlegende Stärke einer digitalen Forschungsinfrastruktur, wie sie beispielsweise von DARIAH-DE betrieben wird.

Dr. Stefan Schmunk
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Abteilung Forschung und Entwicklung
Platz der Göttinger Sieben 1
D-37073 Göttingen
schmunk@sub.uni-goettingen.de

Stefan E. Funk
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Abteilung Forschung und Entwicklung
Platz der Göttinger Sieben 1
D-37073 Göttingen
funk@sub.uni-goettingen.de