

Annotieren und Publizieren mit DARIAH-DE und TextGrid

Kollatz, Thomas

kol@steinheim-institut.org
Steinheim-Institut für deutsch-jüdische Geschichte Essen,
Deutschland

Hegel, Philipp

hegel@linglit.tu-darmstadt.de
Technische Universität Darmstadt, Deutschland

Veentjer, Ubbo

veentjer@sub.uni-goettingen.de
Niedersächsische Staat- und Universitätsbibliothek
Göttingen, Deutschland

Söring, Sibylle

sibylle.soering@sub.uni-goettingen.de
Niedersächsische Staat- und Universitätsbibliothek
Göttingen, Deutschland

Funk, Stefan E.

funk@sub.uni-goettingen.de
Niedersächsische Staat- und Universitätsbibliothek
Göttingen, Deutschland

Annotieren und Publizieren mit DARIAH-DE und TextGrid

Im Rahmen des halbtägigen Workshops werden den Teilnehmerinnen und Teilnehmern Werkzeuge zum Publizieren und Annotieren von Forschungsdaten demonstriert, die im Rahmen von Hands-On-Einheiten anhand eigener und / oder bereitgestellter Daten erprobt werden können.

Vorgestellt und angewendet werden das TextGrid- und DARIAH-DE Repository, der DARIAH-DE Publikator und die DARIAH-DE Annotation Sandbox. Zudem wird in die Arbeit mit dem Text-Bild-Link-Editor des TextGrid Laboratoriums eingeführt und exemplarisch gezeigt, diese Text-Bild Relationen mit Hilfe des Web-Publikationstools „SADE – Scalable Architecture for Digital Editions“ in eine digitale Präsentation bzw. ein Web-Portal zu übernehmen.

Der Workshop richtet sich an Geisteswissenschaftlerinnen und -wissenschaftler aus text- und bildbasierten Disziplinen aller Phasen des akademischen Werdegangs ebenso wie an Vertreterinnen und Vertreter von Institutionen – etwa Bibliotheken,

Forschungsverbünde oder Archive –, die im Rahmen ihrer Vorhaben digitale Forschungsinfrastruktur nutzen bzw. nutzen wollen, um ihre Forschungsdaten nachhaltig digital zu publizieren und zu annotieren.

Der Workshop liefert durch Kurzvorträge und Hands-On-Einheiten Einblicke in verschiedene Verfahren, Anwendungen und Workflows liefern, um Geisteswissenschaftlerinnen und Geisteswissenschaftlern die maschinenlesbare Annotation von Text- und Bilddaten sowie die Publikation solcher Forschungsdaten in einem Repository zu ermöglichen. Nach einer kursorischen Einführung in die Angebote von TextGrid und DARIAH-DE liefert ein Überblick über das Annotieren in den digitalen Geisteswissenschaften verschiedene Anwendungsszenarien, -anforderungen, -modelle und -technologien. Dabei werden neben bereits bestehenden Angeboten wie dem TextGrid Text-Bild-Link-Editor auch neuere Entwicklungen wie die Annotation Sandbox und das DARIAH-DE Repository und seine Publish GUI (Publikator) demonstriert und in interaktiven Übungen durch die Teilnehmenden anhand eigener bzw. zur Verfügung gestellter Daten erprobt.

Der Workshop ist Teil zweier konzeptionell eigenständiger Einreichungen zu den Angeboten der digitalen Forschungsinfrastrukturen TextGrid und DARIAH-DE. Der Besuch beider Workshops ermöglicht eine grundlegende und umfassende Einführung in und Anwendung von Architektur, Tools, Diensten und Workflows zum Annotieren, Sammeln, Modellieren, Recherchieren und Publizieren geisteswissenschaftlicher Forschungsdaten.

Annotationen in den digitalen Geisteswissenschaften

Digitales Annotieren ist zentrale Praxis bei der Wissensgenerierung und variiert je nach spezifischer wissenschaftlicher Zielsetzung und Forschungsgegenstand. Verfahren des fachwissenschaftlichen digitalen Annotierens bilden heute eine der Kernanwendungen der Digital Humanities. Im Zentrum steht dabei ein weites Spektrum von Daten und / oder Objekten, z.B. Texte, Bilder und Musik (Töne, Noten). Digitale Annotationen unterscheiden sich daher in Form, Funktion und Tragweite. Einführend werden die technischen Ebenen und theoretischen Dimensionen der digitalen Annotation in den Geisteswissenschaften exemplarisch erörtert. Die vermittelten Grundlagen können danach im Workshop praktisch angewandt werden.

Annotieren im Rahmen einer digitalen Infrastruktur

Forschungsinfrastrukturen wie TextGrid und DARIAH-DE haben zum Ziel, methodologische Fähigkeiten auf diesem Gebiet zu vermitteln, entsprechende Verfahren

zu evaluieren bzw. bereitzustellen und die nachhaltige Anwendung dieser Verfahren in den Fachwissenschaften zu ermöglichen.

Die DARIAH-DE Annotation Sandbox (Beta) ermöglicht heute die Text- und Bildannotation der Bestände des TextGrid Repository. Darüber hinaus können beliebige Webseiten über den DARIAH-DE Annotationsdienst annotiert werden. Zudem lässt sich der DARIAH-DE Annotationsdienst in eigene Webseiten einbetten; hierzu wurden die digitalen Werkzeuge Annotator.js, Via und ein Annotation Manager über die DARIAH AAI (Authorization and Authentication Service) verfügbar gemacht.

Die DARIAH-DE Annotation Sandbox gestattet die direkte Verbindung der in den Repositorien publizierten Forschungsdaten mit ihrer digitalen Annotation. Diese schließt sowohl die disziplinübergreifenden Nachnutzung als auch die Datenanreicherung oder die Analyse ein. Mittelfristig können Annotationen somit als Zwischenschritt des Forschungsprozesses, aber auch als genuines Forschungsergebnis - etwa im Sinne einer Mikropublikation - verstanden bzw. generiert, verfügbar gemacht und als solches nachgenutzt werden. Im Rahmen einer digitalen Infrastruktur fließen sie wie die Forschungsdaten, auf die sie Bezug nehmen, ebenfalls in die Archivierung ein, um weiterverarbeitet und nachgenutzt zu werden.

Bilder in TextGrid annotieren

Ein weiteres Anwendungsszenario digitaler Annotation stellt die Annotation von Bildern bzw. Bilddaten dar. Eine Vielzahl von Werkzeugen im TextGrid Laboratory erlaubt das Arbeiten mit Texten und Bildern, aber auch beispielsweise mit Noten und Digitalisaten. Eine dieser Komponenten, die auch für die Annotation von Bildbereichen dienen kann, ist der Text-Bild-Link-Editor. Er unterstützt den in TextGrid integrierten XML-Editor bei der Alignierung von Text- und Bildelementen. Ziel ist die Erstellung einer Ausgabedatei, die die Textelemente und die topographische Position von rechteckigen und polygonen Bildbereichen in SVG miteinander verknüpft, wie dies zum Beispiel bei der Verbindung von Faksimiles und Transkriptionen in kritischen Editionen der Fall ist. Auch können Bilder auf diese Weise im Rahmen kunsthistorischer Untersuchungen annotiert werden.

Text-Bild-Relationen publizieren

Die Software SADE der Berlin-Brandenburgischen Akademie der Wissenschaften ist als „Skalierbare Architektur für digitale Editionen“ in TextGrid eingebunden, um eigene Webportale für die Publikation gestalten zu können. Sie enthält ein Modul, mit dem die Verknüpfungen, die mit dem Text-Bild-Link-Editor erstellt wurden, in ein Web-Portal übernommen werden können.

Dieses Modul basiert auf dem in DARIAH-DE integrierten Werkzeug „Semantic Topological Notes“ (SemToNotes). Es erlaubt unter anderem, Zeilen auf einem Digitalisat auszuwählen und Transkriptionen anzuzeigen.

Publizieren via Infrastruktur:

Das DARIAH-DE Repository und der DARIAH-DE Publikator

Das DARIAH-DE Repository bildet eine zentrale Komponente der Infrastruktur, auf die mittels verschiedener Dienste und Anwendungen zugegriffen werden kann. Das Repository erlaubt es, Forschungsdaten zu speichern, diese mit Metadaten zu versehen und die Forschungsdaten durch die Generische Suche aufzufinden. Die Daten werden im DARIAH-DE Storage sicher gespeichert. Darüber hinaus ermöglicht das Repository die nachhaltige und sichere Archivierung von Datensammlungen bzw. Kollektionen.

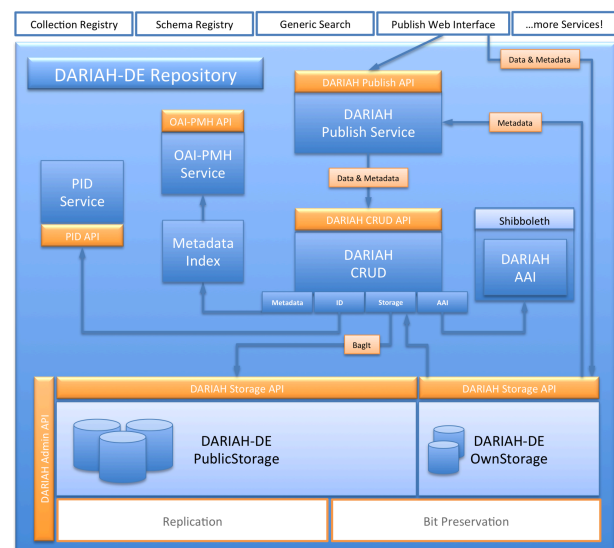


Abb.1: DARIAH-DE-Repository: Architektur

Dies ist komfortabel und intuitiv über ein Web-Interface des DARIAH-DE Portals im Browser möglich, dem DARIAH-DE Publikator. Daten im Repository sind in Kollektionen organisiert, die zunächst vom Nutzer über den Publikator angelegt und mit Metadaten ausgezeichnet werden. Einer Kollektion können beliebig viele Dateien zugeordnet werden, die ebenfalls über den Publikator hochgeladen und mit Metadaten ausgezeichnet werden. Eine publizierte Kollektion sowie alle darin enthaltene Objekte können unmittelbar nach dem Publizieren per Persistent Identifier (PID) referenziert werden und sind damit öffentlich zugänglich und nachhaltig referenzier- und zitierbar. Im nächsten Schritt kann die Kollektion in der Collection Registry nachgewiesen und veröffentlicht werden. Sobald die Kollektion selbst ebenfalls in der

Collection Registry publiziert wurde, sind die Daten auch mit der Generischen Suche recherchierbar.

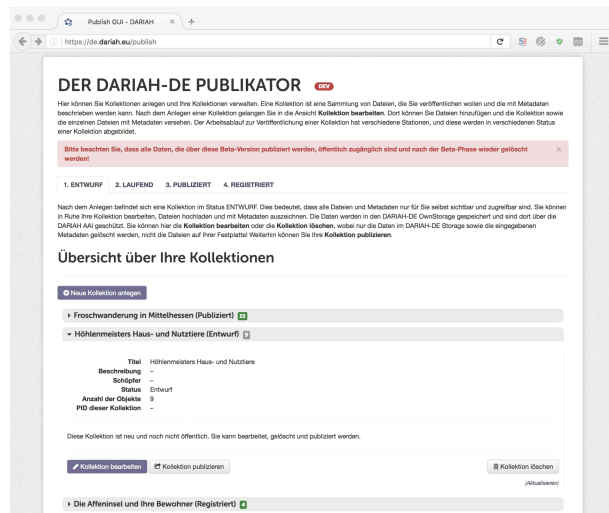


Abb. 2: DARIAH-DE Publikator: Übersicht über die Kollektionen

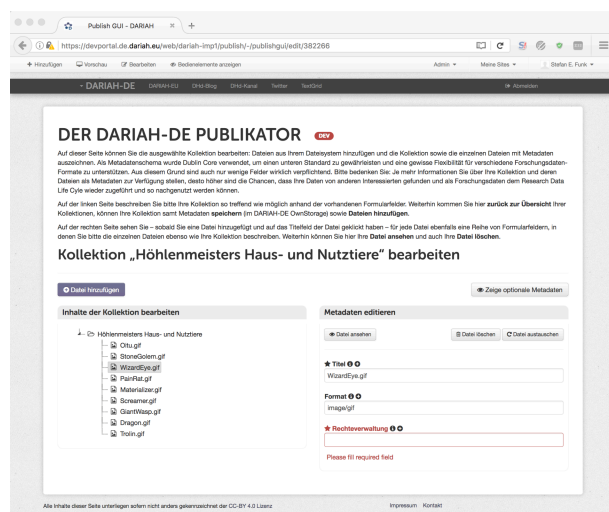


Abb. 3: DARIAH-DE Publikator: Kollektion bearbeiten

Anforderungen

Im Workshop werden exemplarisch Annotationen an einem Digitalisat in TextGrid vorgenommen. Zu diesem Zweck ist ein eigener Rechner mitzubringen, auf dem im Idealfall TextGrid bereits installiert ist – <https://textgrid.de/download>.

Eine Registrierung für TextGrid und DARIAH kann online beantragt werden unter <http://auth.dariah.eu/>

Bitte teilen Sie uns im Vorfeld des Workshops (möglichst bis zum 5. Februar 2017) mit, ob und welche eigenen Materialien Sie verwenden wollen.

Für Rückfragen erreichen Sie uns unter workshop@de.dariah.eu

Kontaktdaten

Mirjam Blümm, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Abt. Forschung und Entwicklung, Papendiek 14, 37073 Göttingen, Forschungsinteressen: Virtuelle Forschungsumgebungen, Digitale Forschungsinfrastrukturen, Digitale Editionen

Stefan E. Funk, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Abt. Forschung und Entwicklung, Papendiek 14, 37073 Göttingen, Forschungsinteressen: Forschungsdatenmanagement, Digitale Langzeitarchivierung, Repositoriums-Technologien.

Canan Hastik, Technische Universität Darmstadt, Dolivostraße 15, Institut für Sprach- und Literaturwissenschaft, 64293 Darmstadt, Forschungsinteressen: Digital Humanities, Semantisches Wissensmanagement, Digitale Kultur und Kunst

Philipp Hegel, Technische Universität Darmstadt, Institut für Sprach- und Literaturwissenschaft, Dolivostraße 15, 64293 Darmstadt, Forschungsinteressen: Digitale Editionen, virtuelle Forschungsumgebungen

Thomas Kollatz, Salomon Ludwig Steinheim-Institut für deutsch-jüdische Geschichte, Essen, Edmund-Körner-Platz 2, 42157 Essen, Forschungsinteressen: Digitale Epigraphik, Jüdische Studien

Sibylle Söring, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Abt. Forschung und Entwicklung, Papendiek 14, 37073 Göttingen, Forschungsinteressen: Virtuelle Forschungsumgebungen, Digitale Forschungsinfrastrukturen, Digitale Editionen

Ubbo Veenster, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Abt. Forschung und Entwicklung, Papendiek 14, 37073 Göttingen, Forschungsinteressen: Digitale Forschungsinfrastrukturen, Text- und Bild-Annotation, Visualisierungstechnologien.

Zahl der möglichen Teilnehmerinnen und Teilnehmer.

Aufgrund des hohen Praxisanteils soll die Zahl der Teilnehmerinnen und Teilnehmer auf möglichst 25 beschränkt bleiben.

Angaben zu einer etwa benötigten technischen Ausstattung.

WLAN / Beamer / Stellwände / Verlängerungskabel

Fußnoten

1. Siehe auch Workshop "Daten sammeln, modellieren und durchsuchen mit DARIAH-DE"

Bibliographie

Becker, Rainer / Bender, Michael / Borek, Luise / Hastik, Canan / Kollatz, Thomas / Lordick, Harald / Mache, Beata / Rapp, Andrea / Reiche, Ruth / Walkowski, Niels-Oliver (2016): „Digitale Annotationen in der geisteswissenschaftlichen Praxis“, in: *Bibliothek – Forschung und Praxis* 40 (2): 186–199 <https://www.degruyter.com/view/j/bfup.2016.40.issue-2/bfp-2016-0042/bfp-2016-0042.xml?format=INT>.

Bender, Michael / Borek, Luise / Kollatz, Thomas / Reiche, Ruth (2015): "Wissenschaftliche Annotationen: Formen – Funktionen – Anforderungen", in: *DHd-Blog* <http://dhd-blog.org/?p=5388>.

Borek, Luise / Reiche, Ruth (2014): „Round Table „Annotation von digitalen Medien“ (Veranstaltungsbericht), in: *DHd-Blog* <http://dhd-blog.org/?p=3831>.

Blümm, Mirjam / Funk, Stefan E. / Söring, Sibylle (2015): „Die Infrastruktur-Angebote von DARIAH-DE und TextGrid“, in: *Information. Wissenschaft & Praxis* 66 (5–6): 304–312.

Neuroth, Heike / Rapp, Andrea / Söring, Sibylle (2015): *TextGrid: Von der Community für die Community – Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften*. Göttingen http://www.univerlag.uni-goettingen.de/handle/3/Neuroth_TextGrid.

Schmunk, Stefan / Funk, Stefan (2015): „Das DARIAH-DE- und das TextGrid-Repositorium: Geistes- und kulturwissenschaftliche Forschungsdaten persistent und referenzierbar langzeitspeichern“, in: *Bibliothek Forschung und Praxis* 40 (2): 213–221 10.1515/bfp-2016-0020

Söring, Sibylle (2016): „Technische und infrastrukturelle Lösungen für digitale Editionen: DARIAH-DE und TextGrid“, in: *Bibliothek Forschung und Praxis* 40 (2): 207–212 10.1515/bfp-2016-0040.