



DSA-Zertifizierung (M 4.2.3)

Version 26.02.2019

Cluster 4

Verantwortlicher Partner SUB Göttingen

DARIAH-DE Überführung der digitalen Forschungsinfrastrukturen für die e-Humanities in die Operational Phase (Betriebsphase)

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1610A bis J, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



**Bundesministerium
für Bildung
und Forschung**

Projekt: DARIAH-DE: Überführung der digitalen Forschungsinfrastrukturen für die e-Humanities in die Operational Phase (Betriebsphase)

BMBF Förderkennzeichen: 01UG1610A bis J

Laufzeit: März 2016 bis Februar 2019

Dokumentstatus: Final

Verfügbarkeit: Öffentlich

Autoren:

Claudio Leone, SUB Göttingen



Dieses Werk ist unter einer Creative Commons Lizenz vom Typ Namensnennung 3.0 Deutschland zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie <http://creativecommons.org/licenses/by/3.0/de/> oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.

Inhaltsverzeichnis:

1. Einleitung	4
2. Rahmen und Stand der Vorbereitungsarbeiten	5
2.1. Selbstevaluation und Dokumentation	5
2.2. Einreichung und erste Evaluierung durch die Gutachter	8
2.3. Finale Schritte.....	10
3. Literaturverzeichnis	10

1. Einleitung

Im Rahmen der Arbeitsschwerpunkte von Cluster 4 wurde die Weiterentwicklung des TextGrid und DARIAH-DE Forschungsdaten-Repositorys weiterverfolgt und diente dem Aufbau einer nachhaltigen Forschungsdaten-Föderationsarchitektur, die aus einer Vielzahl von Tools und Services besteht, welche die Nutzung von verteilten Forschungsdaten und wissenschaftlichen Sammlungen ermöglichen. Die Zusammenarbeit mit TextGrid spielte dabei eine tragende Rolle. Die bereits etablierte Kooperation, insbesondere bei der Entwicklung des Repositorys, wurde in der dritten Förderphase von DARIAH-DE aktiv fortgeführt. Die technologische Zusammenführung der beiden Repositorys wurde vorangetrieben und modular nachnutzbare Technologiekomponenten entwickelt. In diesen Kontext gehört auch die angestrebte Zertifizierung des TextGrid und DARIAH-DE Repositorys mit dem Data Seal of Approval¹ (DSA), die Gegenstand dieses Berichtes ist, in dem der Stand der dazugehörigen Arbeiten dokumentiert wird.

Die Zertifizierung dient der Schaffung von Vertrauenswürdigkeit, sowohl seitens der Wissenschaftler und Wissenschaftlerinnen, die Ihre Daten und Ergebnisse publizieren, als auch seitens der Nutzer und Nutzerinnen, die Daten/Publicationen suchen, beurteilen können und nachnutzen wollen. Vertrauenswürdigkeit ist wichtig für die Verbreitung von wissenschaftlichen Ergebnissen, weitergehende Forschung, wissenschaftlichen Austausch und Kommunikation. Hierbei spielen Qualitätskriterien, die stabile Referenzierbarkeit, Nachprüfbarkeit und Nachvollziehbarkeit von wissenschaftlichen Arbeiten/Daten eine wichtige Rolle, aber auch ethische und rechtliche Aspekte.²

Vertrauenswürdigkeit lässt sich diesbezüglich definieren als „die Eigenschaft eines Systems [...], gemäß seinen Zielen und Spezifikationen zu operieren (d.h. es tut genau das, was es zu tun vorgibt) und dies auch in geeigneter Weise glaubhaft zu machen [...]“³

Der Kriterienkatalog von Zertifikaten übersetzt in der Regel Erwartungen von Wissenschaftlern und Wissenschaftlerinnen und Anforderungen sowie Standards an das wissenschaftliche Arbeiten an sich in Mindestanforderungen, die durch Dienste im Bereich Forschungsdatenmanagement und Publikation zu erfüllen sind. Die Kriterien, Anforderungen und Maßstäbe, die zur Erlangung des Zertifikats erfüllt werden müssen, können als Grundlage zum Aufbau einer entsprechenden Infrastruktur oder zur Selbstevaluierung bereits bestehender Dienste dienen, da z.B. die Langzeitspeicherung von Daten mit großen Herausforderungen auf der Ebene der Organisation, der Technologie, der Finanzierung sowie auf der rechtlichen Ebene verbunden ist. Genau auf diese

¹ www.datasealofapproval.org/

² Vgl. Dobratz, Susanne; Schoger, Astrid: 5.2. Grundkonzepte der Vertrauenswürdigkeit und Sicherheit. In: nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3. Hrsg. von Neuroth, Heike et al. Göttingen: nestor c/o Niedersächsische Staats- und Universitätsbibliothek 2010, S. Kap.5:2, URN: urn:nbn:de:0008-2010030597

³ Ebenda

Aspekte gehen die Kriterien ein und sorgen für eine transparente Dokumentation nach innen wie nach außen.

2. Rahmen und Stand der Vorbereitungsarbeiten

Im Rahmen der Arbeiten von Cluster 4 wurde eine interne Roadmap erstellt und mit allen beteiligten Akteuren aus der Niedersächsischen Staats- und Universitätsbibliothek Göttingen sowie der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen als Provider der technischen Infrastruktur besprochen. Die für die Zertifizierung notwendigen Arbeiten wurde dabei mit Rücksicht auf gegebene Ressourcen an Verantwortliche und inhaltlich zuständige Mitarbeiter verteilt. Die Koordination und Gesamterstellung der Dokumentation sowie die umfangreiche Ausformulierung der Einreichung oblag dabei Cluster 4 und war auf entsprechende Zuarbeit angewiesen.

Im Laufe der Vorbereitung der Zertifizierung kam es zu einer Änderung der Anforderungen aufgrund eines Zusammengehens von Data Seal of Approval (DSA) und World Data System (WDS) in eine neue Zertifizierungsorganisation mit neuem Siegel - dem Core Trust Seal (CTS). In Cluster 4 wurde in Rücksprache mit allen Akteuren beschlossen, die Arbeiten im Hinblick auf die neuen, überarbeiteten Anforderungen neu zu organisieren und aufzurollen, wodurch eine spätere Einreichung anvisiert werden musste. Weiterhin wurde aufgrund nicht geringer technischer sowie organisatorischer Gemeinsamkeiten und Überschneidungen zwischen dem TextGrid Repositorium und dem DARIAH-DE Repositorium beschlossen, zuerst die Einreichung des TextGrid Repositoriums vorzubereiten und nach entsprechender Rückmeldung weite Teile für das DARIAH-DE Repositorium nachzunutzen.

2.1. Selbstevaluation und Dokumentation

Das Data Seal of Approval (wie auch der Nachfolger Core Trust Seal) stellt in Bezug auf das European Framework for Audit and Certification⁴ die erste von drei Stufen dar, die sich neben unterschiedlichen Anforderungen auch im Hinblick auf den Evaluierungsprozess unterscheiden:

- Stufe1: „basic certification“ – Selbstevaluierung mit Peer-Review (Data Seal of Approval bzw. aktuell Core Trust Seal)
- Stufe2: „extended certification“ – Selbstevaluierung, die von externen Gutachtern geprüft wird (ISO 16363 oder DIN 31644/nestorSeal)
- Stufe3: „formal certification“ – vollständiges externes Audit (ISO 16363 oder DIN 31644/nestor Seal)

⁴ <http://www.trusteddigitalrepository.eu>

Mit der Vorbereitung der Zertifizierung für das Core Trust Seal (CTS) gingen eine entsprechende Selbstevaluation des TextGrid Repositoriums und notwendige Anpassungen und Erweiterungen einher. Elementar für die Erlangung der Zertifizierung ist dabei, dass die im Rahmen der Selbstevaluierung getroffenen Selbstaussagen in Form einer öffentlichen und transparenten Dokumentation belegt sind. Diese betrifft eine Vielzahl umfangreicher Kriterien, die sich auf die drei übergeordnete Themen Organisationsstruktur, Datenmanagement und Technologie beziehen, wie nachfolgende aus dem offiziellen Anforderungskatalog exzerpierte Kurzübersicht in Form von Tabellen illustriert⁵:

Table 1: Organisational Infrastructure

Anforderungskriterium	Kurzfassung der Anforderung
Mission/Scope	The repository has an explicit mission to provide access to and preserve data in its domain.
Licences	The repository maintains all applicable licenses covering data access and use and monitors compliance.
Continuity of access	The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.
Confidentiality/Ethics	The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.
Organisational infrastructure	The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.
Expert guidance	The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).

Table 2: Digital Object Management

Anforderungskriterium	Kurzfassung der Anforderung
Data integrity and authenticity	The repository guarantees the integrity and authenticity of the data.

⁵<https://www.coretrustseal.org/wp-content/uploads/2017/01/20180629-CTS-Extended-Guidance-v1.1.pdf>

Anforderungskriterium	Kurzfassung der Anforderung
Appraisal	The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.
Documented storage procedures	The repository applies documented processes and procedures in managing archival storage of the data.
Preservation plan	The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.
Data quality	The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.
Workflows	Archiving takes place according to defined workflows from ingest to dissemination.
Data discovery and identification	The repository enables users to discover the data and refer to them in a persistent way through proper citation.
Data reuse	The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Table 3: Technology

Anforderungskriterium	Kurzfassung der Anforderung
Technical infrastructure	The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
Security	The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users

Hinter jeder Kurzbeschreibung der Anforderung steht eine Vielzahl von Fragen, die es zu berücksichtigen und zu beantworten gilt. Für das TextGrid Repository wurden diese

en Detail durchgegangen und eine entsprechende Dokumentation hierfür erstellt, der wiederum eine Vielzahl von zusammenhängenden Arbeiten als Resultat der Selbstevaluierung vorangegangen ist, wie z.B.:

- die Beschreibung und Klärung von Governance-Aspekten und der zugrundeliegenden basalen technischen Infrastruktur mit entsprechenden Partnern
- eine visuell unterstützte Erarbeitung von Workflows zum Datenmanagement
- eine Erneuerung und Aktualisierung von Informationen auf den Webseiten des Repositoriums
- eine Überarbeitung der Nutzungsbedingungen, um z.B. ethischen und rechtlichen Aspekten gerecht zu werden
- Die Erarbeitung und schriftliche Fixierung von Notfallplänen mit entsprechenden zu klärenden Verantwortlichkeiten intern und extern

Neben der Ausformulierung der Einreichungstexte und Anpassungen an bereits vorhandenen Dokumentationen wurde insbesondere für die Themenbereiche Datenmanagement und Technologie eine neue Dokumentationsseite im Wiki erstellt: <https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management>, die folgende Aspekte inhaltlich abdeckt:

- Import and Publish Procedures
- TextGrid Object API Integrity
 - TextGrid Resolver and PIDs
- Longterm Preservation and Data Curation
 - Persistent Identifiers
 - Technical Metadata Extraction
 - Longterm Preservation Archiving
 - Storage System
 - Backup Strategy and Data Recovery
 - Data Integrity and Authenticity
- Technical Infrastructure
 - TextGrid Repository Components
- Security Issues

2.2. Einreichung und erste Evaluierung durch die Gutachter

Im Rahmen der Erarbeitung der Dokumentation wurde klar, dass es einige Bereiche gibt, die vorerst ausreichend für den Betrieb aufgestellt sind, aber langfristig weiterentwickelt werden können. Dies gilt insbesondere für den Bereich Data Appraisal. Hier

wurde entschieden, die Evaluation durch die Gutachter für eine Einschätzung abzuwarten und den bis dato erarbeiteten Stand einzureichen, da der Zertifizierungsprozess einem iterativen Verfahren folgt und somit die Expertise des Peer-Reviewing zielführend einfließen kann.

Eine erste Einreichung für das Textgrid Repository erfolgte im Sommer 2018. Die Rückmeldung der Ergebnisse durch zwei Gutachter und die Core Trust Seal Zertifizierungsorganisation nahm allerdings einige Monate in Anspruch.

Die Rückmeldungen der Gutachter wurden zuletzt mit allen beteiligten Akteuren besprochen und einer ersten inhaltlichen Strukturierung und Gruppierung unterzogen, an deren Ende zwei weitere neue Dokumentationsseiten im Wiki stehen, die für mehr Transparenz sorgen sollen. Diese werden zurzeit sukzessive mit Inhalt gefüllt und stehen, sobald finalisiert, öffentlich zur Verfügung: <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure>, <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies>. Zwei große Themenbereiche haben sich dabei herauskristallisiert, an denen bereits gearbeitet wird: Data Policies und Organisational Infrastructure.

Für den ersten Themenbereich gilt es Data Policies für das Textgrid Repository zu formulieren und folgende Unterbereiche (Unterkapitel) abzudecken:

- Collection Development Policy and Data Quality/Re-Use
- Preservation Policy
 - Aims and Requirements of the Policy
 - Recommendations and List of Preferred Formats
 - Legal and Regulatory Framework
- Important Ethical and Disciplinary Norms

Für den Bereich Organisational Infrastructure gilt es noch genauer und klarer die Organisationsstruktur hinter dem nachhaltigen Betrieb des Repositorys vor allem im Hinblick auf eine garantierte Langzeitarchivierung zu illustrieren, da durch die Vielzahl beteiligter Projekte und Institutionen (Textgrid, DARIAH-DE und Nachfolgeprojekt CLARIAH-DE, Humanities Data Centre mit den dahinterstehenden Institutionen SUB und GWDG) Unklarheiten bezüglich des Zertifizierungsgegenstandes aber auch bezüglich Verantwortlichkeiten und Rollen aufkommen. Hierfür werden gerade Grafiken und Texte zu folgenden Unterkapiteln erstellt:

- Mission and Long Term Sustainability
- The TextGrid Repository and his Designated Community
- Organisational Infrastructure and Long-Term Sustainability
- Responsible Institutions: Rules and Obligations

2.3. Finale Schritte

Die finale Einreichung für das TextGrid Repository erfolgt im März 2019, die Vorbereitung für die Einreichung für das DARIAH-DE Repository läuft ab März parallel, da viele Inhalte aus den Vorbereitungen für das TextGrid Repository in Anlehnung nachgenutzt werden können. Die Einreichung für das DARIAH-DE Repository erfolgt nach Erhalt des finalen Gutachtens zum TextGrid Repository. Eine Zertifizierung mit dem weitaus seltener vergebenen Nestor-Siegel^{6 7} wurde geprüft und wird erst zukünftig angestrebt, da es hierfür einer weiteren Optimierung der Repositorien auf relevanten Ebenen bedarf. Mit der Erlangung des Core Trust Seals als dem am weitesten verbreiteten Standard ist hierfür eine erste Grundlage zur Weiterarbeit gelegt.

3. Literaturverzeichnis

Core Trust Seal. <https://www.coretrustseal.org/> [Stand 26.02.19]

Core Trust Seal: CoreTrustSeal Data Repositories Requirements - Extended Guidance. <https://www.coretrustseal.org/wp-content/uploads/2017/01/20180629-CTS-Extended-Guidance-v1.1.pdf> [Stand 26.02.19]

Data Seal of Approval. <https://www.datasealofapproval.org/en/> [Stand 26.02.19]

Dobratz, Susanne; Schoger, Astrid: 5.2. Grundkonzepte der Vertrauenswürdigkeit und Sicherheit. In: nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3. Hrsg. von Neuroth, Heike et al. Göttingen: nestor c/o Niedersächsische Staats- und Universitätsbibliothek 2010, S. Kap.5:2. URN: urn:nbn:de:0008-2010030597

European Framework for Audit and Certification of Digital Repositories. <http://www.trusteddigitalrepository.eu> [Stand 26.02.19]

Nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive Version 2. Frankfurt am Main: nestorc/o Deutsche Nationalbibliothek 2008. (nestor-materialien 8). URN: urn:nbn:de:0008-2008021802

Nestor-Siegel für vertrauenswürdige digitale Langzeitarchive. http://www.langzeitarchivierung.de/Subsites/nestor/DE/Siegel/siegel_node.html [Stand 26.02.19]

⁶ Vgl. Nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive Version 2. Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek 2008. (nestor-materialien 8). URN: urn:nbn:de:0008-2008021802

⁷ Nestor-Siegel für vertrauenswürdige digitale Langzeitarchive. http://www.langzeitarchivierung.de/Subsites/nestor/DE/Siegel/siegel_node.html