



Assessment Information

[CoreTrustSeal Requirements 2017–2019](#)

Repository:

TextGrid Repository

Website:

<https://textgridrep.org/en>

Certification Date:

15 May 2020

This repository is owned by:

Humanities Data Centre

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org



TextGrid Repository

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository, Publication repository, Research project repository

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

Being a part of the DARIAH-DE research infrastructure and a component of the TextGrid virtual research environment (VRE), the TextGrid Repository is especially designed for researchers in the arts and humanities and disciplines dealing with text-based scientific research. It is furthermore optimised for TEI and XML data formats as the TextGrid VRE offers a variety of tools and services for scholars in the humanities in digital scholarly editing, storing and publishing. Several cooperating research projects use the TextGrid virtual research environment for research endeavours and to manage their data internally in the own storage (private area) of the repository and finally publish the results in the public storage of the repository. For more information and an overview see the following description of the repository and his designated community.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of the Repository's Designated Community.

The TextGridRep, which is object of this certification, is the result of the community driven project TextGrid (<https://textgrid.de/en/>) running from 2006 to 2015 to develop a virtual research environment (TextGridVRE) for digital scholar editing, for collaborative creation, analysis and publication of text and images (mainly digitised manuscripts). Organisational structure, practices, policies and content are related to its history and community and are only understandable against this background.

TextGrids community consists of scientists from the humanities, libraries, computing centres and of information scientists integrating established standards and best practices into the virtual research environment which have constantly been further developed and adapted in the project lifetime. Underlying processes have been coordinated by the Research and Development Department of the Gottingen State and University library who has been leading the project and carried out the development of the repository from the funding proposal in 2005 onwards. (See the list of partners for all three funding

phases on the following webpages: TextGrid

<https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/textgrid/> ; TextGrid II <https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/textgrid-vernetzte-forschungsumgebung-in-den-ehumanities/> ; TextGrid III <https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/textgrid-iii-virtuelle-forschungsumgebung-fuer-die-geisteswissenschaften-1/>)

In 2016 the TextGrid Repository (TextGridRep) and its virtual research environment (TextGrid VRE) became part of the DARIAH-DE research infrastructure, which offers several tools and services for researchers in the humanities in a broader context (<https://de.dariah.eu/en/>).

The TextGridRep is a discipline specific repository and commits itself to ensure availability and long-term preservation of data in the humanities. Because of its designated community and its history the TextGrid Repository defines itself more specific as a searchable and citable long-term archive especially suitable and designed for digital editions and text based scientific research. Therefore, the repository is optimised for XML/TEI formats and offers here additional services. As part of the virtual research environment TextGrid VRE it is linked to the TextGrid Laboratory (TextGridLab) where most of its publications are collaboratively edited, prepared and finally published out of the laboratory in the repository. Further access is provided by a TextGridRep API such as TG-import allowing permanent and referenceable storage of data in different formats, whereby the presentation of XML/TEI is highly supported. Several cooperating research projects (https://wiki.de.dariah.eu/display/TextGrid/Data+Policies?preview=/83230842/105710815/190725_Koop-TextGrid-DARIAH_english.png#DataPolicies-CollectionDevelopmentPolicyandDataQuality/Reuse) and researchers from the arts and humanities use the TextGrid Repository as part of the TextGrid VRE for research endeavours and to manage their data internally within the protected own storage area and finally publish the results in the public storage of the repository. When it comes to depositing data, the designated community are the researchers using the TextGrid Laboratory. All published data receives a persistent identifier and a citation recommendation. Represented disciplines of data depositors and users are the following:

- Editorial Philology
- German Philology
- Slavic Studies
- Jewish Studies
- Ancient American Studies
- Theology
- Philosophy
- Ethnology
- Historical Science
- Legal History
- Cultural History
- Art History
- Musicology

The TextGrid Repository is an open access repository, which means that access to deposited data is always free and the data is publicly available to everyone – means to everyone who has internet access. This is in line with the open access commitment of the repository as declared in the TextGrid Repository mission statement (<https://textgridrep.org/en/>) and also suitable to the type of data the designated community do publish and reuse. The repository is not suitable for publication of sensitive data or personal data with disclosure risk. A code of conduct and measures as well as legal implications in case of non-compliance or abuse are described in detailed in the TextGrid Repository Terms of Use (<https://textgrid.de/terms-of-use>).

Only for the TextGrid OwnStorage of the repository and in the context of research projects using the TextGrid Laboratory for collaboratively editing of data a secure environment with access control and a differentiated role management ensures access and exchange of data within a research group that are not publicly available.

Besides the Terms of Use, consultation and recommendations (<https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-RecommendationsandListofPreferredFormats>) as well as workshops, user meetings (<https://textgrid.de/en/veranstaltungen> , <https://de.dariah.eu/veranstaltungen>), publically available tutorials (<https://textgrid.de/en/tutorials>) and further information (https://textgrid.de/en_US/dokumentation-und-manuals) ensures the skills and information researchers need to have to use the repository and related services.

For a more detailed description of the TextGrid Repository, its mission and its designated community, see <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-MissionandLong-TermPreservation> , <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-TheTextGridRepositoryanditsDesignatedCommunity>

For a description of data publication and reuse related to community practice see: <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-CollectionDevelopmentPolicyandDataQuality/Reuse> , <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-DataReuse>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

A. Content distributed as deposited, B. Basic curation – e.g. brief checking; addition of basic metadata or documentation

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept.

Comments

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Outsource Partners. If applicable, please list them.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Other Relevant Information.

Current number of editions in the TextGridRep (03/09/2020): 93978

(https://textgridlab.org/1.0/tgoaipmh/oai?verb=ListIdentifiers&metadataPrefix=oai_dc)

Current number of registered users (03/09/2020): 2445

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The TextGrid Repository (TextGridRep) is a digital preservation archive for human sciences research data providing a variety of data for teaching and research purposes. It promotes open access to research data including open standards allowing an efficient reuse for research. The TextGridRep also provides researchers with a comprehensive and reliable service to store their data permanently, well described and with a stable reference for citation and reuse. The TextGridRep is part of the TextGrid Virtual Research Environment (VRE) and linked to the TextGrid Laboratory (TextGridLab), where most of its publications are collaboratively edited, prepared and finally published out of the laboratory in the repository. In this context the repository as archive for depositing and reusing data is linked to an environment which offers besides digital preservation also open-source software for collaborative creation, analysis and publication of text and images. It is a searchable and citable long-term archive especially suitable for digital editions and text based scientific research.

Therefore, the repository is optimised for XML/TEI formats and offers here additional services. An independent publication from the TextGridLab including other types of data and formats is equally possible by tools using the TextGridRep API such as TG-import, whereby the presentation of XML/TEI is highly supported. For a more detailed description see information in the public wiki on the “TextGrid Repository and his Designated Community”: <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-TheTextGridRepositoryanditsDesignatedCommunity>.

The TextGrid Repository is a community orientated result of a national program to establish a Digital Humanities infrastructure in Germany and operates in terms of long term operation together with the DARIAH-DE Repository as part of the Humanities Data Center (HDC) founded by the Göttingen State and University Library (SUB – <https://www.sub.uni-goettingen.de/en/about-us/portrait/>) and the Gesellschaft für wissenschaftliche Datenverarbeitung

Göttingen mbH (GWDG – <https://www.gwdg.de/about-us>), which operates as computing centre and IT-competence centre for the University of Göttingen and the Max Planck Society. Both institutions provide their resources and services in terms of technical infrastructure, administration, human resources and expertise through the DARIAH-DE operating cooperation agreement (specified in an internal document “TextGrid_FTE”) as well as through involvement in CLARIAH-DE for strategic developments and potential disciplinary innovations concerning the repository (see [https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructureandLong-TermSustainability](https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-OrganisationalInfrastructureandLong-TermSustainability)). SUB and GWDG control all parts of the data lifecycle in-house and responsibilities and tasks related to each institution are documented publically: <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-ResponsibleInstitutions:RulesandObligations>).

The mission of the TextGridRep is to serve national and international research, teaching and learning by providing long term preservation, continued access, reuse, openly sharing and dissemination of digital research data according to ethical and scientific standards of the international research community. The TextGridRep’s mission is publicly stated and approved by the DARIAH-DE Coordination Office (<https://de.dariah.eu/en/kontakt>).

The full description of the mission statement is available at the TextGrid website (<https://textgridrep.org/en/>) as well as in the public wiki (<https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-MissionandLong-TermPreservation>).

The TextGrid Repository fulfills its mission and ensures long-term preservation and efficient reuse by an appropriate organisational and technical infrastructure in terms of data policies and digital object management.

A description of the collection development of the TextGrid Repository also including data enrichment or format changes is documented here: <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-CollectionDevelopmentPolicyandDataQuality/Reuse> .

A description of the preservation policy of the TextGrid Repository is also available on the public wiki page on Data Policies: <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-PreservationPolicy> .

An overview about the main technologies and procedures used for an appropriate data management including import and publish workflows is available at: <https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management> . See in this context also the description for requirement “VIII Data Appraisal” below.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The licence is determined by the data producer/ data owner. The researcher, who uploads data choose the licence. The indication of licences for the use of the deposited data by others are mandatory at all levels (aggregations of data like collections, editions and for single objects). Due to the design of the upload process of the data in the repository, uploading data without indicating the type of licence is not possible, because mandatory licence and right holder metadata fields are implemented for the publishing process (for an overview of the TextGrid metadata, see <http://textgrid.info/namespaces/metadata/core/2010> .

The type of data deposited is not restricted as long as it is in compliance with ethical and legal criteria as stated in the Terms of Use of the TextGrid Repository and recommendations are given for long term preservation suited and open formats that are in line with the nestor criteria Catalogue of Criteria for Trusted Digital Repositories (see: <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-RecommendationsandListofPreferredFormats> , http://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf).

Access to deposited data is always free and the data is publicly available to everyone - means to everyone who has internet access. This is in line with the open access commitment of the repository as declared in the mission statement (<https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-MissionandLong-TermPreservation>). A code of conduct and measures as well as legal implications in case of non-compliance or abuse are described in detailed in the Terms of Use (<https://textgrid.de/terms-of-use>).

In order to ensure that the relevant legal frameworks for the protection of personal data, especially the EU data protection directive, DARIAH-DE as service provider of the TextGrid Repository deals with private data as furthermore described in the privacy notice (see <https://de.dariah.eu/datenschutz>).

To ensure that adherence to the copyright and other laws, before publishing data every user has to accept the already mentioned in the TextGrid Repository Terms of Use. Every user has to declare that he/she own[s] all necessary rights to publish this collection including all data and metadata and to allow re-use by third parties (see <https://textgrid.de/en/terms-of-use> , “§ 8 Terms of Use of the TextGrid Repository”).

The Terms of Use also include regulations with regard to the privacy policy: “Data, collections, or metadata that allow conclusions to be drawn about individual persons may not be imported unless the author obtains explicit confirmation from the persons concerned or their legal representatives that they are in agreement with publication in the DARIAH-DE Repository. This confirmation must be presented to DARIAH-DE in writing.” (see <https://textgrid.de/en/terms-of-use> , “§ 8 Terms of Use of the TextGrid Repository”). The depositors of data are informed in a legally binding manner that the repository is not suitable to publish personal and sensitive data which has to be protected according to applicable law.

As data can only be uploaded after the authentication of the user (DARIAH-DE account or eduGAIN), the misuse can be traced back to the perpetrator. After the publication the data is stored securely in the TextGrid Repository and is publicly accessible.

Following the open access policy of the Textgrid repository creative commons licences (see: <http://creativecommons.org/licenses/>) are recommended to the community.

Monitoring the compliance of the data owner’s / data producer’s licences is not done actively by the repository operators. In case of a notification of the repository operators, illegal deposited or not-compliant data will be removed and the persistent identifier then leads to an information page.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The TextGrid Repository as part of the DARIAH-DE research infrastructure ensures long-term access and preservation of its holdings. A commitment to long-term availability and accessibility of research data stored in the Textgrid Repository is publicly stated in the Terms of Use and reflects the mission statement as pointed out at the repository website:

“DARIAH-DE commits itself, within the framework of technical and organizational possibilities, to permanently store the data / collections and their descriptive metadata, to provide them with persistent IDs, and to make them accessible for research via international data networks” (see <https://textgrid.de/terms-of-use> (§8, 8.) , <https://textgridrep.org/>).

Furthermore, the sustainable operation of the repository is ensured by his incorporation into the Humanities Data Centre (HDC – <http://humanities-data-centre.de/>), which is supported by central and persistent institutions in the field of research infrastructures and humanities: The Göttingen State and University Library (SUB), which is one of the biggest research libraries in Germany engaged in several research infrastructure projects, and the “Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG), which operates with long-term preservation expertise as computing centre and IT-competence centre for the University of Göttingen and the Max Planck Society.

The SUB and the GWDG committed themselves in the founding manifesto of the HDC to ensure a sustainable and persistent data centre for the humanities, its business continuity and to take all necessary responsibilities, technical and organisational steps to realise and maintain services for long-term access and archiving of research data over time and changing circumstances (https://humanities-data-centre.de/wp-content/uploads/2016/08/HDC_Erklaerung-Aufbau-Forschungsdatenzentrum_2016-07-27_gez.pdf).

No data will ever be removed from the repository without ensuring a relocation, direct advice from the data provider/owner or transmitting data to the provider/owner. Last but not least several technical measures are undertaken in the digital object management to ensure long-term preservation as described in the public wiki (<https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management#DigitalObjectManagement-LongtermPreservationandDataCuration>).

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The repository points out relevant issues related to sensitive data and to applicable disciplinary norms in the official Terms of Use. Every user/author has to accept the Terms of Use before publishing data. Therefore, the process of uploading data into the TextGrid Repository is designed in a way that makes every user confirm that he/she is aware of data with disclosures risk. As stated in the terms of use authors/depositors of data have to declare that he/she own[s] all necessary rights to publish this collection including all data and metadata and to allow re-use by third parties (see <https://textgrid.de/en/terms-of-use>, “§ 8 Terms of Use of the TextGrid Repository”). By accepting the Terms of Use they declare to be responsible for compliance with legal and ethical criteria related to personal privacy, copyright issues, computer fraud, abuse and dissemination of unlawful offending material, etc. Rights and obligations of users (authors/depositors) as well as measures in case of non-compliance are clearly described. This includes removing of already published data, exclusion from use and legal implications.

The TextGrid Repository declares to be an open access repository and tries to promote open access in the scientific community as an important part of the repository’s policy (see <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-PreservationPolicy>).

Open Access is also an important part of DARIAH-DE's policy, the Göttingen University's Open Access Policy (see <https://www.sub.uni-goettingen.de/en/electronic-publishing/open-access/>) and Göttingen State and University Library's strategic goals (see <https://www.sub.uni-goettingen.de/en/about-us/portrait/strategy/#c13120>). This means that all data, once published, is clearly available open access without restrictions and can be used freely according to the licence provided. Users of deposited data have not to register to access data, as the repository don't allow publishing of data not suitable for open access according to applicable law. Nevertheless, German Copyright Law applies in this case.

TextGrid therefore advises authors/depositors to be aware of data with disclosures risk, which are not allowed to be published without permission and an explicit confirmation from the persons concerned or their legal representatives that they are in agreement with publication. This has to be presented in writing to DARIAH-DE.

If permissions are not given for open access publication, non sensitive data are to be seperated from sensitive data with disclosure risk for publishing. Sensitive data will have to be published in a secure environment – not in an open access repository as the TextGridRep. Besides regulations by the Terms of Use, information and recommendations are given in the public wiki (see <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-Personalsensitivedatawithdisclosurerisk>).

All authors/depositors declare to follow the principles of good scientific practice as formulated and highly recommended by the German Research Foundation (DFG) (Deutsche Forschungsgemeinschaft (DFG) (ed) (2013) Safeguarding Good Scientific Practice: Memorandum – Recommendations of the Commission on Professional Self Regulation in Science", Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. DOI: <https://dx.doi.org/10.1002/9783527679188.oth1>). See also the public wiki page on "Important Ethical and Disciplinary Norms", where some highlighted principles are presented (<https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-ImportantEthicalandDisciplinaryNorms>).

As part of the TextGrid research environment and the DARIAH-DE research infrastructure the Terms of Use of the repository are formulated as part of parent regulations. Additionally §8 of the Terms of Use is explicitly dedicated to the repository. The Terms of Use are accepted when accounting for the use of the DARIAH-DE research infrastructure, when publishing in the repository and are publically available in German and English (see <https://textgrid.de/en/terms-of-use>).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

In terms of long term operation, stability and sustainability the TextGrid Repository is integrated in a complex organisational infrastructure as illustrated by the following figure and explanations in the dedicated public wiki page (see <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-OrganisationalInfrastructureandLong-TermSustainability>). This complex organisational infrastructure is due to the history of the repository and to the necessity to balance community involvement, already existing infrastructures and agreements as well as a long term guarantee for its operation.

First, the TextGrid Repository is operated by the Humanities Data Centre, which has been explicitly founded for this purpose. The data centre is funded by the Göttingen State and University Library (SUB – <https://www.sub.uni-goettingen.de/en/about-us/portrait/>) and the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen mbH (GWDG – <https://www.gwdg.de/about-us>; <https://www.gwdg.de/home>), which operates as computing centre and IT-competence centre for the University of Göttingen and the Max Planck Society.

These two well known institutions in the field of digital research infrastructure take over the necessary responsibilities for the long-term-operation, and sustainability of the HDC – as publicly declared in their founding manifesto: https://humanities-data-centre.de/wp-content/uploads/2016/08/HDC_Erklärung-Aufbau-Forschungsdatenzentrum_2016-07-27_gez.pdf , https://humanities-data-centre.de/wp-content/uploads/2016/08/HDC_Erkl%C3%A4rung-Aufbau-Forschungsdatenzentrum_2016-07-27_gez.pdf. The founding of the HDC was a very important step for stability as it ensures a sustainable operation of the repository with all necessary resources by two important institutions – in addition and if necessary independently of public project funding.

Second, the TextGrid Repository is part of the DARIAH-DE research infrastructure since 2016. The operation of the

repository is therefore additionally ensured until the end of 2021 by the DARIAH-DE operating cooperation – a legally binding cooperation agreement between all consortium members to keep services and tools of DARIAH-DE running for the transition phase into the planned long-term funded national research infrastructure for 2021 onwards.

The SUB and GWDG are important consortium members of DARIAH-DE and the main pillars for the sustainable operation. The SUB is responsible for the administrative coordination of the operating cooperation and the GWDG takes the responsibility for the technical coordination. SUB and GWDG provide their resources and services in terms of technical infrastructure, administration, human resources and expertise for a sustainable operation of the TextGrid Repository through the DARIAH-DE operating cooperation agreement (specified in an internal document “TextgridRep_FTE”).

Third, both institutions are also involved in associated projects as CLARIAH-DE, which provides additional resources in means of financial and human resources for proposing and developing specific extensions for the TextGrid repository and its virtual research environment TextGrid. CLARIAH-DE consists of partners of CLARIN-D and DARIAH-DE and is in constant exchange with the DARIAH-DE steering committee. Strategic developments and potential disciplinary innovations concerning the repositories are implemented in two ways:

- Minor or regular adjustments are provided by the employees of the DARIAH-DE-Coordination Office responsible for the TextGrid Repository and VRE.
- Project-specific extensions are proposed and developed within the framework of projects as CLARIAH-DE, whose ‘Work Package 2 – Tools and Virtual Research Environments’, among other things, is concerned with the TextGrid VRE. Like every work package within CLARIAH-DE, Work Package 2 conducts regular telephone conferences as well as face-to-face meetings.

SUB and GWDG take over a variety of tasks within the framework of this complex organisational infrastructure, especially within the research infrastructure DARIAH-DE operating cooperation (service provider) and of the associated project CLARIAH-DE. All tasks fulfilled by SUB and GWDG as running members of the HDC and consortium members of DARIAH-DE are in-house and listed in the following overview.: <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-ResponsibleInstitutions:RulesandObligations> . This overview shows which responsibility each organization has in sustaining the repository and that both institutions cover several necessary competences. Some are highlighted in the following.

The SUB offers the assignment of DOIs (Digital Object Identifier) at present only for the DARIAH-DE Repository, but this is planned for the TextGrid Repository as well, an implementation plan has already been set up. As one of the most advanced libraries in the field of research data management the SUB offers a high quality of consulting competence in this field. The SUB furthermore is responsible for the TextGrid Repository's presentation layer and APIs – the frontend of the repository and for the design of the tools for the ingest and administration of research data.

The GWDG has high competences in the field of Bitstream-Preservation, it ensures the preservation and accessibility of data and offers the assignment of EPIC-PIDs for the TextGrid Repository. The GWDG provides the basic technical infrastructure for its operation. It is also responsible for the DARIAH-DE Authorisation and Authentication Infrastructure

(AAI), which is also used for TextGrid authentication and rights management. Publication access to the repository is provided through the authentication infrastructure DFN-AAI-Basic (<https://www.aai.dfn.de/de/der-dienst/>) of DFN, the communications network for Science and research in Germany (Deutsches Forschungsnetz – <https://www.dfn.de/en/>) as well as through the DARIAH Authentication and Authorization Infrastructure (DARIAH AAI) for institutions or scientists (<https://wiki.de.dariah.eu/display/publicde/DARIAH+AAI+Documentation>), not registered or being part of DFN-AAI-Basic. The ISO 9001 certification of GWDG (https://www.gwdg.de/documents/20182/62973/Zertifikat+ISO+901-2008_2016-05-20+bis+2018-09-14_deutsch.pdf; <https://www.iso.org/standard/46486.html>) also shows that there are several defined procedures to ensure an efficient operation of the data centre).

All mentioned responsibilities and tasks are related to dedicated employees of SUB and GWDG, which regularly participate on conferences, workshops and project related working groups relevant for their work. This ensures to stay in touch with the state-of-the-art in their fields of expertise. Additional qualification and training offers are discussed individually between employee and respective supervisor. An overview in terms of full time equivalents related to relevant tasks in sustaining the repository is stated in an internal document (“TextgridRep_FTE”). SUB and GWDG ensures all necessary resources through commitment to the DARIAH-DE operating cooperation agreement, involvement in CLARIAH-DE as well as through the Humanities Data Center as a long term commitment.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Expert guidance is an important aspect for the adoption and further development of the workflows and technologies in use related to the TextGrid Repository and the TextGrid Virtual Research Environment (VRE). Due to the complexity of the organisational infrastructure of the repository guidance is provided from a variety of involved organisation and projects (see: <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-OrganisationalInfrastructureandLong-TermSustainability>).

TextGrid workshops on specific topics are regularly organised through the DARIAH-DE activities in this domain. Common topics of workshops are:

- DARIAH-DE and TextGrid tools for teaching purposes
- annotation of texts
- digital editions
- archiving and publishing
- images in TextGrid
- digital editing with TextGrid
- norm data and metadata
- adaptations and extension within TextGrid
- digital editions and geo-referential data
- “Around the World in a TextGrid Day: edit, manage, publish and explore your TEI data with TextGrid and DARIAH-DE”

Participants of the workshops are students, researchers and cooperating research projects that are using TextGrid as well as involved developers and IT experts adapting the TextGrid VRE for their needs (for an overview of the designated community see also: <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-TheTextGridRepositoryanditsDesignatedCommunity>).

A further format of presenting TextGrid as virtual research environment and the repository are our “Grand Tours”, where experts for the TextGrid VRE from DARIAH-DE cover a variety of topics and offer small working areas for people showing their projects in order to discuss their issues.

Furthermore, TextGrid is presented together with the DARIAH-DE research infrastructure on conferences, poster sessions and scientific community events (<https://textgrid.de/en/veranstaltungen>, <https://de.dariah.eu/veranstaltungen>). This ensures additional feedback in a broader context from the scientific community concerning their needs, methods and standards.

Feedback from the community is taken into consideration by all consortium members of DARIAH-DE dealing with the TextGridVRE. All partners of the DARIAH-DE operating cooperation (<https://de.dariah.eu/en/der-forschungsverbund>) discuss the strategic development of all of the DARIAH-DE services including the TextGrid Repository in monthly telephone conferences. This ensures scientific advisory from a network of organisations providing expertise in different disciplinary fields including the user point of view (by research members mainly from literary studies and editorial philology using the TextGrid VRE) as well as the technical perspective (by members of the development team).

The Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen mbH (GWDG) is hosting the technical infrastructure of TextGrid, which furthermore ensures technical expertise and an infrastructure adapted to the designated community.

The GWDG is certified with ISO 9001 which shows the ability of ensuring an efficient operation of the data centre (https://www.gwdg.de/documents/20182/62973/Zertifikat+ISO+901-2008_2016-05-20+bis+2018-09-14_deutsch.pdf). As part of the DARIAH-DE infrastructure, TextGrid is also advised by the DARIAH-DE scientific advisory board (<https://wiki.de.dariah.eu/display/publicde/Wissenschaftlicher+Beirat>), which consists of scientists with data science expertise, disciplinary experts and IT experts. Furthermore there is advice by the CLARIN DARIAH Technical Advisory Board (TAB) which consists of six international IT specialists (<https://wiki.de.dariah.eu/display/publicde/Technical+Advisory+Board>). CLARIN-D and DARIAH-DE are cooperating in the context of CLARIAH-DE as associated project for proposing and developing specific extensions to the TextGrid Repository.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The TextGrid Repository guarantees the integrity and authenticity of the data through ingest workflows technical background processes and evaluation and control mechanisms.

The following aspects and measures are implemented to ensure data integrity:

- At publication, each object is assigned a EPIC2 Handle PID. In addition to the URL, the PID record also contains the internal object ID (TextGrid URI), the file size and the checksum of an object. A background process regularly compares the file size and the checksum stored in the PID record with the attribute of the actual object stored in the repository and its metadata. This ensures that the object still exists and hasn't been modified. The source code of the background process can be found at: <http://hdl.handle.net/11022/0000-0007-C2EA-6> and is included in the TextGrid server's configuration.
- This process is part of the digital object management of the repository and documented in the TextGrid public wiki (section "Data Integrity and Authenticity": <https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management>)
- At the process of publication mandatory metadata are to be indicated. They are complemented by automatically generated metadata and may be further extended by metadata indicated in the optional fields. The metadata schema and concept of the TextGrid Repository is documented and accessible: <http://textgrid.info/namespaces/metadata/core/2010> , <https://wiki.de.dariah.eu/download/attachments/12189756/Metadata-Cheatsheet.pdf?api=v2>

Fields which are required for publishing an edition:

- `tg:object/tg:edition/tg:isEditionOf`
- `tg:object/tg:edition/tg:license`

Fields which are required for publishing an item:

- `tg:object/tg:item/tg:rightsHolder`

Fields which are required for publishing a collection:

- tg:object/tg:collection/tg:collector

Fields which are required for publishing a work:

- tg:object/tg:work/tg:agent
- tg:object/tg:work/tg:dateOfCreation
- tg:object/tg:work/tg:genre

After data is published, it can't be modified. Only the publication of a new version (revision) is possible by using an update operation. Versioning is realised by linking previous versions into the metadata of the new published object. Only data owners have the necessary rights to create and publish new revisions, the date of creation as well as the data contributor are listed automatically in the respective metadata (see

<https://textgridlab.org/1.0/tgcrud-public/rest/textgrid:19044.2/metadata>):

```
<textgridUri extRef="">textgrid:19044.2</textgridUri>  
<revision>1</revision>
```

- The previous version would then be the predecessor: textgrid:164pv.1 (see: <https://textgridlab.org/1.0/tgcrud-public/rest/textgrid:19044.1/metadata>). It is possible, that predecessors of published objects have not been published.

- Therefore, revisions of the data can be controlled allowing to follow a chain back to the first version. Furthermore, it is possible to show all revisions through the TG-search functionality "show revisions". (see TG-search documentation <https://textgridlab.org/doc/services/submodules/tg-search/docs/api/info.html#get-info-textgriduri-revisions> , <https://textgridlab.org/1.0/tgsearch-public/info/textgrid:164pv/revisions> , <https://wiki.de.dariah.eu/display/TextGrid/Using+Revisions>).

- TextGrid makes use of international standards and conventions, such as the metadata schema and concept is based on DC terms, Marc, PREMIS, LoC Terms, RDVocab, Peral.org, creative commons, etc. The TextGrid Metadata Schema XSD file shows a variety of used standards: <http://textgrid.info/namespaces/metadata/core/2010>

- Furthermore, the TextGrid Repository supports the visualisation of TEI and XML formats (for files stored in the repository) as HTML. (The repository supports TEI-Visualisation by user-generated XSLT-transformation, for an example of a visualisation in the TextGridRep see: https://textgridrep.org/en/browse/-/browse/vz4c_0)

The following aspects are implemented to ensure data authenticity:

- Data that is published can't be changed or deleted anymore, only new revisions of an object can be published. Information on changes realised by creating new revisions are registered in the respective metadata. Data owner have full

control of the rights for the publication of new revisions (checksums are generated at the ingest).

- Provenance information is provided in the TextGrid metadata record.
- The metadata of the deposited objects contains the respective PID. For an example of TextGrid metadata see: <https://textgridrep.org/textgrid:3qxtk.0/metadata> ; for an example of handle metadata see: <https://hdl.handle.net/21.11113/0000-000B-D5C3-8?noredirect>.
- Whereas TextGrid doesn't actually compare essential properties of different versions, versions can be controlled manually by users comparing the respective metadata or data.
- For publication on the TextGrid Repository a DARIAH-DE account or access to the DFN-AAI is necessary. The identity of depositors is ensured by the DARIAH-DE Authentication and Authorization Infrastructure (DARIAH AAI : <https://wiki.de.dariah.eu/display/publicde/DARIAH+AAI+Documentation>; DFN-AAI-basic: <https://www.aai.dfn.de/en/>)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

To understand how data appraisal is organised for the TextGrid Repository, it is crucial to understand that data are usually published out of the TextGrid Laboratory into the repository or deposited directly to the TextGrid Repository by research projects that are editing data in groups of research experts from specific disciplines (see <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-TheTextGridRepositoryanditsDesignatedCommunity>) and preparing publications. When it comes to depositing data, the designated community are the researchers using the TextGrid Laboratory. These research projects are consulted and supported by staff members from the SUB and DARIAH-DE and are using with the TextGrid VRE a domain specific virtual research environment. This ensures up to now that data published in the repository is in line with the self-understanding of the designated community and the mission and self-definition of the repository as especially designed for the publication of digital editions and related data suitable for text based research in the humanities. The long-term research data archive TextGridRep offers in this perspective safe storing, publishing and researching for versatile digital material such as XML/TEI formatted text and images.

That's why the collection development policy focuses mainly in direction of digital editions and related data contributing at the same time to teaching and the broader context of the development and set up of a service for digital editions at Göttingen State and University Library (<https://www.sub.uni-goettingen.de/en/about-us/portrait/strategy/>).

The TextGrid Repository is following a successful self-regulatory approach in terms of collection development supported by its designated community and the collection is also in addition actively enriched by further content – if necessary or considered as useful. The content of the repository has been enriched actively e.g. in 2011 by the so called TextGrid Digital Library to open the TextGridRep and his data to a broader public and other students and researchers outside the research projects using the TextGrid VRE. Format conversions have been undertaken for this purpose from XML to a valid TEI format in order to make data available for further processing for example in editions and text corpora. At present, the collection development is expanded to increase the interoperability with other external data sets – for instance the German Text Archive (<http://www.deutschestextarchiv.de/>) to enhance reuse and establish connectivity to further text analysing tools. Collections of TextGrid, the German Text Archive and CLARIN can be merged in the future e.g. for text analysis. The works are done within the framework of CLARIAH-DE, such as approaches for converting the collections of the TextGrid Repository stored in the TEI format to the DTA basic format (DTABf – <http://www.deutschestextarchiv.de/doku/basisformat/>).

All this is stated and explained also on the public wiki page of the TextGrid Repository (<https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-OrganisationalInfrastructureandLong-TermSustainability>). Apart from active enrichment and format conversions of content, the use of the VRE as editing and publication platform by the designated community and due to the self-regulatory approach, the repository has no further content-specific preselection of data for publishing from repository side, though automated technical and mandatory metadata checks are implemented for every import workflow. The responsibility is here at the depositor's side with the limitation to follow the Terms of Use of the TextGrid Repository (<https://textgrid.de/en/terms-of-use>). The terms of use exclude abuse of the repository to publish or disseminate unlawful and offending content as well as content that allow conclusions to be drawn about individual persons without explicit permission. All authors declare to follow the principles of good scientific practice as formulated in the recommendations of the German Research Foundation (See Deutsche Forschungsgemeinschaft

(DFG) (ed) (2013) Safeguarding Good Scientific Practice : Memorandum - Recommendations of the Commission on Professional Self Regulation in Science", Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi: <https://dx.doi.org/10.1002/9783527679188.oth1>).

The description to make data accessible and reusable is provided by all necessary metadata that is available for every single object of the TextGridRep.

A minimal mandatory set of metadata ensures the completeness of a description to allow understandability of data and meets the needs of the designated community. The TextGrid metadata schema was developed as a chained schema that serves editions (<http://textgrid.info/namespaces/metadata/core/2010#editionType>), collections (<http://textgrid.info/namespaces/metadata/core/2010#collectionType>), works (<http://textgrid.info/namespaces/metadata/core/2010#workType>), and items (<http://textgrid.info/namespaces/metadata/core/2010#itemType>) with different mandatory metadata for every level (see Bibliographic Metadata sheet:

<https://wiki.de.dariah.eu/download/attachments/12189756/Metadata-Cheatsheet.pdf?api=v2>). So any researcher can decide which complexity shall be used and fits best to their requirements (see minimal and mandatory metadata). In the development of the metadata schema the designated community as well as metadata experts from SUB (<https://www.sub.uni-goettingen.de/en/contact/departments-a-z/departamental-and-unit-details/abteilunggruppe/metadaten-und-datenkonve>) were involved from the beginning through the first two funding phases of the TextGrid project (2006-2012) (see <https://textgrid.de/en/projekt>). The complexity of the chained metadata schema is fully implemented in the TG-search service to sustain the findability of data.

Besides machine wise checking of necessary metadata, the designated community evaluates the quality of data by the metadata provided – included automatically generated metadata. Technical metadata for long-term preservation are added to data during the publish process, and are orientated towards needs of the designated community (JHOVE (<http://jhove.openpreservation.org/>) and FITS (<https://projects.iq.harvard.edu/fits/home>) are the tools used, which are widely spread). Data can't be published, if mandatory metadata is not provided by depositors. For reuse of the data (mainly XML/TEI transcribed texts and images of digitised manuscripts) the mandatory metadata fields are sufficient as the community evaluates the quality of data usually by accessing the encoded texts and digital editions – a procedure inherent to text-based scientific disciplines such as Editorial Philology or other disciplines of the Humanities when dealing with textual sources. A full text search allows that enhanced and (project) specific metadata are additionally accessible, such as TEI headers and other marked up elements in TEI XML files (see TG-search documentation – <https://textgridlab.org/doc/services/submodules/tg-search/docs/index.html>). This allows the community not only to access those metadata but also to elaborate, annotate or enrich a source with an own XML schema or to add additional metadata and republish it as a new citable source suitable for their research needs. The designated community usually evaluates this data when it is disseminated and reused. In terms of the DARIAH-DE Research Data Lifecycle this process is part of 'source definition', e.g. when researchers select data for specific analysis or editing of new digital editions and when they publish data. This corresponds to the part 'sources' at the ingest level and to the part 'access and curation' at publication level of the publication workflow for the TextGrid VRE as described on the public wiki page on data reuse (<https://wiki.de.dariah.eu/pages/viewpage.action?pageId=83230842#DataPolicies-DataReuse>).

Evaluation of data quality is an integrated part of the scientific discourse in the humanities and in text-based research. Additionally, DARIAH-DE and the TextGrid Repository follow the recommendations of the nestor guide (http://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf).

No formats are excluded from publishing, but TEI and XML as well as TIFF are communicated within the dedicated community as preferred formats as they are more suitable to long-term preservation. A list of preferred formats is provided on the TextGrid public wiki (<https://wiki.de.dariah.eu/pages/viewpage.action?pageId=83230842#DataPolicies-RecommendationsandListofPreferredFormats>). TEI and XML are especially supported by the TextGrid Repository due to the fact that the TextGrid Repository is part of a broader virtual research environment that offers services and tools for collaboratively creating and publishing of digital editions for scholars in the digital humanities. Non-preferred formats are accepted, but the repository doesn't offer supplementary services as for TEI or XML such as HTML visualisation (<https://wiki.de.dariah.eu/display/publicde/Das+DARIAH-DE+Repository+und+das+TextGrid+Repository>). For non-preferred formats the repository offers bitstream preservation and dissemination ensuring data integrity and authenticity. For preferred formats further services are offered or further changes in addition to bitstream preservation may be undertaken in the future – if necessary. Technical innovations and changes as well the scientific practice are observed by members of the staff dealing with the repository and its community. Necessary changes will be discussed and implemented within the framework of its organisational infrastructure and resources (see TextGrid Preservation Policy – <https://wiki.de.dariah.eu/pages/viewpage.action?pageId=83230842#DataPolicies-PreservationPolicy>).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

All relevant processes concerning archival storage of data – covering ingest, storage, metadata and access to data - are documented in the public TextGrid wiki. Import and publish procedures are described in detail:

<https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management> .

The TextGrid Repository is part of a virtual research environment. Therefore, there are two distinct storage areas: An OwnStorage (dynamic) and a PublicStorage (static). The dynamic storage area contains all data which are not meant for public consumption. Moreover, these data is also secured by the authentication and authorization component (TG-auth* – <https://textgridlab.org/doc/services/submodules/tg-auth/docs/index.html>). By means of this component, data owners can restrict the access on specific datasets. Usually, data always belongs to a project. Each project in turn, has different members, with individual access or modification permissions.

The PublicStorage in turn contains only published datasets (published by the user), that are available for public consumption. Furthermore, as public data only contains the outcome of the research endeavour, any other data from the dynamic storage area is only transferred to the PublicStorage if published. TextGrid users are informed by the Terms of Use they agree to about the fact that all data stored in the public area of the repository are publicly accessible. In this context also the issue of sensitive data is pointed out. The public wiki contains also a short section on personal sensitive data with disclosure risk with some links to further information and recommendations.

(<https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-Personalsensitivedatawithdisclosurerisk>) In case of violation of applicable law and of the TextGrid Terms of Use, DARIAH-DE will remove such data when ever discovered or pointed too by third parties and take measures and legal actions in the framework of the terms of use.

User guides are provided besides the technical documentation for a step by step instruction for the data provider in terms of data collection, enriching with mandatory and optional metadata suitable for the designated community and so forth (<https://wiki.de.dariah.eu/display/TextGrid/User+Manual+2.0> , <https://wiki.de.dariah.eu/display/TextGrid/Main+Page>).

The Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen mbH (GWDG), who is the storage provider for the repository, guarantees long-term bit-stream preservation of data for at least ten years. This means all stored data is retrievable as it was initially stored. As the storage system of the repository is provided and maintained by GWDG, the adequate techniques used by the computing centre to mitigate storage media deterioration are applied. The most important mechanism is RAID6. Also, consistency checks are in place. Details about the technical infrastructure and import and publish procedures including back-up strategies, mitigation of data loose and risk management are provided in the public Wiki. The particular procedures related to storage, backup and data recovery are publicly documented under

the Longterm Preservation and Data Curation section: <https://wiki.de.dariah.eu/display/DARIAH3/TextGrid+Repository%3A+Digital+Object+Management#TextGridRepository:DigitalObjectManagement-LongtermPreservationandDataCuration> .

The GWDG follows standards that includes risk management techniques for a variety of scenarios which are fully described in an internal risk management plan including a plan for a recovery in case that data is lost.

In case of data loss due to services breakdown or server hacks, servers and services can be restored by setting up new servers and services from scratch (VM and Puppet: <https://github.com/DARIAH-DE/puppetmodule-dhrep>), and data from backup storage. For each scenario a workflow with specified actions indicating the responsible staff member is documented (internal TextGrid risk management plan).

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The TextGrid Repository has a Preservation Policy

(<https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-PreservationPolicy>) which is in line with the open

access strategy of the University of Göttingen and its research data policy as well as reflected in the mission statement of the repository. It represents a clear commitment to open access of research data in promoting and making data of the designated community of the repository as widely accessible and usable as possible. Within its organisational and technical infrastructure, the repository ensures continuous access to its resources by following an active preservation policy in managing this in a planned and documented way. Within the policy description aims and requirements the repository strives to ensure, recommendations and list of preferred formats as well as the legal and regulatory framework are stated and explained. Related technical aspects and data management issues including backup and data recovery strategies as well as security issues are described in a separate public wiki page about “Digital Object Management” (see: <https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management>).

For TEI/XML and TIFF as most suitable long-term preservation formats according to the state of the art of the designated community and respective technical standards the repository offers additional services (see section “VIII Data Appraisal” and the public wiki page <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-DataReuse>) and will also undertake format changes in the future if necessary. For non-preferred formats the repository offers at present long-term access and only a minimum preservation level with bitstream preservation including authenticity and integrity of data. However, data format changes and technological developments as well as community best practice are observed for all formats represented in the repository by staff members of SUB and GWDG and changes are discussed and decided within its designated community, organisational infrastructure and will then be implemented if considered necessary (see <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-OrganisationalInfrastructureandLong-TermSustainability>).

During the publication process, each object is assigned a EPIC2 Handle, and technical metadata is extracted and stored using the File Information Toolset (FITS - <https://projects.iq.harvard.edu/fits>). So, a format migration of certain objects will be possible, if necessary. As ensured by the Terms of Use (<https://textgrid.de/terms-of-use>, especially by §8), the repository has all necessary rights to copy, transform and migrate data, if necessary. Consistency of data and metadata is also checked during publication. Preservation planning is understood and handled as described in the OAIS recommendation. (see [https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management#DigitalObjectManagement-TextGridandtheOpenArchivalInformationSystem\(OAIS\)](https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management#DigitalObjectManagement-TextGridandtheOpenArchivalInformationSystem(OAIS))). In addition to the EPIC Handle PIDs, DataCite DOIs will be provided for every object in the near future.

Furthermore, correction of eventually false checksums and other incorrect metadata is done if necessary. The outcome of the procedures described in <https://wiki.de.dariah.eu/display/DARIAH3/TextGrid+Repository%3A+Digital+Object+Management#TextGridRepository:DigitalObjectManagement-DataIntegrityandAuthenticity> indicates incorrect metadata.

There are plans to additionally put every published TextGrid object into a LTP archive (a so called dark archive) to enhance the level of preservation (see: <https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management#DigitalObjectManagement-LongtermPreservationandDataCuration>).

The TextGrid Repository which belongs to the DARIAH-DE research infrastructure commits itself in the TextGrid terms of use to "permanently store the data / collections and their descriptive metadata, to provide them with persistent IDs, and to

make them accessible for research via international data networks" (<https://textgrid.de/terms-of-use>). Preservation at the governance level are described in section "III Continuity of access".

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XI. Data quality

R11. 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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The TextGrid Repository ensures that there is enough available and understandable metadata for its designated community to allow an estimation of the quality of the data.

Each deposited dataset is automatically checked for mandatory metadata needed by the TextGrid metadata schema.

TextGrid metadata validity is checked during the publish process. Additionally, more metadata can be imported as data object. The quality of the data itself is up to the user, the metadata is documented in the TextGrid wiki (see <https://wiki.de.dariah.eu/display/TextGrid/Metadata>).

Fields which are required for publishing an edition:

- `tg:object/tg:edition/tg:isEditionOf`
- `tg:object/tg:edition/tg:license`

Fields which are required for publishing an item:

- `tg:object/tg:item/tg:rightsHolder`

Fields which are required for publishing a collection:

- `tg:object/tg:collection/tg:collector`

Fields which are required for publishing a work:

- `tg:object/tg:work/tg:agent`
- `tg:object/tg:work/tg:dateOfCreation`
- `tg:object/tg:work/tg:genre`

(see for mandatory fields:

<http://www.textgridlab.org/doc/services/submodules/kolibri/kolibri-tgpublish-service/docs/configuration.html#tgpublish>)

Links to related works or citations can be given in the custom metadata field (`object/custom`) if necessary, see <https://textgridrep.org/textgrid:39dj6.0/metadata> .

The following provenance metadata are generated automatically (see <https://textgridrep.org/textgrid:165hp.0/metadata>):

- `object/generic/provided/generated/created` – Date of creation of the resource
- `object/generic/provided/generated/lastModified` – Date on which the resource was changed
- `object/generic/provided/generated/dataContributor` – Used for a person that submits data in textgrid

Object specific provenance metadata would then be:

- at generic level (see <https://textgridrep.org/textgrid:165hp.0/metadata>)
 - `object/generic/relations/isDerivedFrom` – Link to an object the described textgrid object is derived from
- at collection Level (see <https://textgridrep.org/textgrid:165hp.0/metadata>)
 - `object/generic/collection/collector` – Person or corporate body creating a collection

- at edition level (see <https://textgridrep.org/textgrid:3740r.0/metadata>)
- object/generic/edition/isEditionOf – Manifestation of which work? Value must be the TextGrid URI of a TextGrid Work object. Field is mandatory on publication
- object/generic/edition/source/bibliographicCitation and object/generic/edition/source/objectCitation – Relation between the textgrid object and the description of the source of this object

Technical metadata are also created automatically and can be downloaded for every object (including text and image objects): <https://textgridlab.org/1.0/tgcrud/rest/textgrid:3qxtk.0/tech>

The TextGrid Repository monitors the development of community standards and technical change and can undertake required preservation actions if necessary, based on the technical metadata extracted.

At import process, administrative metadata is created, that contains information such as file size, creation and lastModified date, and checksum (see <http://textgrid.info/namespaces/metadata/core/2010>, genericType).

For an example see: <https://textgridrep.org/textgrid:1659b.0/metadata>

Workflows and technical measures to ensure data authenticity and integrity as part of data quality are already described above in section "VII. Data integrity and authenticity" and in the TextGrid public wiki: <https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management#DigitalObjectManagement-DataIntegrityandAuthenticity> .

As already described for VIII Data Appraisal (see above), the repository is strongly related to its designated community and follows a self-regulatory approach which is accompanied by automatized procedures at ingest and publication of data and supported by consultation. That's why the repository operates with recommendations, consultation and support (<https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-RecommendationsandListofPreferredFormats>) and does not have quality restrictions besides the mandatory metadata fields. When the designated community of the repository evaluates data quality, the mandatory metadata sets and automatically generated metadata during ingest and storing are a first orientation. The designated community evaluates quality of data mainly by accessing the encoded texts and digital editions – a procedure inherent to text-based scientific disciplines such as Editorial Philology or other disciplines of the Humanities when dealing with textual sources (mainly XML/TEI transcribed texts). Images are mainly digitised manuscripts and a suitable file format for long term preservation is recommended (TIFF 6.0) as listed in the TextGrid public wiki and according to the Nestor criteria Catalogue of Criteria for Trusted Digital Repositories (see http://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf and <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-RecommendationsandListofPreferredFormats>). The TextGrid Repository monitors the development of community standards and technical change and can undertake required preservation actions if necessary. For the text data of the repository – mainly XML/TEI encoded texts – a full text search (see TG-search documentation - <https://textgridlab.org/doc/services/submodules/tg-search/docs/index.html>) allows to access additionally metadata depending on their relevance for specific research question or projects. This may be TEI headers and other marked up elements in TEI XML files (see TG-search documentation – <https://textgridlab.org/doc/services/submodules/tg-search/docs/index.html>). That's why the minimal and mandatory

metadata are sufficient for reuse of data. This chained metadata system was developed from the beginning in the first two funding phases (2006-2012) in close cooperation between the designated community (<https://textgrid.de/en/projekt> , https://textgrid.de/en_US/textgrid-forschungsverbund) and metadata experts from SUB (<https://www.sub.uni-goettingen.de/en/contact/departments-a-z/departamental-and-unit-details/abteilunggruppe/metadaten-und-datenkonversion/>) to ensure understandably and reuse of data for the community. For the information already given and some additional details see also the section on data quality and data reuse at the public wiki page about Data Policies (<https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-CollectionDevelopmentPolicyandDataQuality/Reuse>)

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

All workflows concerning archiving are documented at different levels.

The user documentation describes all necessary procedures to manage data through the repository tools, see: <https://wiki.de.dariah.eu/display/TextGrid/User+Manual+2.0> and <https://textgrid.de/tutorials> .

At the technical level the public repository API is presented, see: <https://textgridlab.org/doc/services/index.html> .

At a process level a general Digital Object Management documentation describes all underlying aspects of the architecture and the storage system, illustrates the ingest, publication and access workflows in OAIS terms as well as all relevant aspects concerning Long Term Preservation and Data Curation, see:

<https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management> .

All necessary rights to handle data and the necessary workflows as well as legal and privacy policy issues as well as ethical aspects are described in the TextGrid Terms of Use (<https://textgrid.de/en/terms-of-use>).

Security issues are documented by internal risk management plans, namely the internal TextGrid risk management plan and the GWDG internal risk management plan.

The TextGrid Repository has no direct appraisal and selection of data, despite the mandatory metadata fields. Authors and Users are committed to follow the recommendations for a good scientific practice as formulated by the German Research Foundation (DFG) (see: (DFG) (ed) (2013) Safeguarding Good Scientific Practice : Memorandum - Recommendations of the Commission on Professional Self Regulation in Science", Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi: <https://dx.doi.org/10.1002/9783527679188.oth1>).

TEI and XML are preferred formats to publish in TextGrid and are supported with additional services such as a HTML visualisation within the TextGridRep (<https://textgridrep.org>), a full text search via TG-search for text and valid XML files and better possibilities for editing data in the TextGridLab (see:

<https://wiki.de.dariah.eu/display/publicde/Das+DARIAH-DE+Repository+und+das+TextGrid+Repository>).

An overview of all the workflows is available at <https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management#DigitalObjectManagement-ImportandPublishProcedures> and will be updated if changes occur.

As the repository is especially designed for data of digital editions and related research, data ingest usually are published out of the TextGrid Laboratory, where research projects and researchers edit their data before publication (<https://wiki.de.dariah.eu/display/TextGrid/Data+Policies?preview=/83230842/108135272/tg-publication-cycle.png#DataPolicies-CollectionDevelopmentPolicyandDataQuality/Reuse>). That's why all data published is in line with the TextGridRep being a discipline specific archive. However due to the repository's commitment and mission to enable open access to its data (<https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-MissionandLong-TermPreservation>) without restrictions, the repository does not exclude different data to be published as long as it is compliant to all ethical and legal guidelines as stated in the terms of use.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:
Accept

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Open access to research data and reusing research data are important parts of the mission of the DARIAH-DE research infrastructure and of the TextGrid Repository.

The TextGrid Repository offers various ways for data discovery and identification. It provides a search engine for metadata and full text data (TG-search). It can be accessed via the TextGrid Repository website (<https://textgridrep.de>) and TG-search API (documentation accessible at <http://www.textgridlab.org/doc/services/submodules/tg-search/docs/index.html>).

Furthermore, the TextGrid Repository provides access via the Textgrid OAI-PMH data provider (see <https://textgridlab.org/1.0/tgoaipmh/oai?verb=Identify>) and collections can additionally be registered in the DARIAH-DE Collection registry (<https://colreg.de.dariah.eu>) for being indexed in the DARIAH-DE Generic Search (<https://search.de.dariah.eu/search>), such as the “TextGrid Digitale Bibliothek” (<https://textgridrep.org/en/search?filter=project.value%3aDigitale+Bibliothek>).

A suggested citation style is documented in the TextGrid Repository Documentation (see <https://textgridlab.org/doc/services/citation.html>), and is also available for every TextGrid object as a text snippet on <https://textgridrep.org> (see Citation: https://textgridrep.org/browse/-/browse/vqn2_0#citation).

Persistent Identifiers are provided with each object's metadata set (EPIC2 Handle PIDs) for stable reference. As a part of future extensions each object of the TextGrid Repository will also be assigned DataCite DOIs, as already implemented for the DARIAH-DE Repository.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

TextGrid data objects are described by the TextGrid metadata schema (<http://textgrid.info/namespaces/metadata/core/2010>), that uses different XML standards for data description. TextGrid makes use of international standards and conventions, such as the metadata schema and concept is based on DC terms, Marc, PREMIS, LoC Terms, RDvocab, Peral.org, creative commons, etc.

The TextGrid metadata schema is constantly observed for possible enhancements by the TextGrid community. Changes

will be taken into account for future versions of the TextGrid services. Important changes that have little or no impact on running services will be deployed instantly. Furthermore, the TextGrid schema is well-documented in the TextGrid User Guide as well as in the metadata schema itself.

TextGrid proposes to use the TEI/XML format for storing text files according to the Nestor criteria Catalogue of Criteria for Trusted Digital Repositories (http://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf). XML for long term usage and retrieval and TEI as humanities based format as it is well-documented and widely used (see also a list of recommended formats:

<https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-RecommendationsandListofPreferredFormats>). It is orientated on long term preservation formats and community standards to ensure understandability of the data (see also public wiki pages on the designated community and data reuse: <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-TheTextGridRepositoryanditsDesignatedCommunity> , <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-DataReuse>).

For images and digitised manuscripts TextGrid proposes TIFF as recommended format according to the Nestor criteria Catalogue of Criteria for Trusted Digital Repositories (see http://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf and <https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-RecommendationsandListofPreferredFormats>). The TextGrid Repository monitors the development of community standards and technical change and can undertake required preservation actions if necessary.

As already described in the section "VIII Data Appraisal", for non-preferred formats the repository offers at present long term access and only a minimum preservation level with bitstream preservation including authenticity and integrity of data. However, data format changes and technological developments as well as community best practice are observed for all formats represented in the repository by staff members of SUB and GWDG and changes are discussed and decided within its designated community, organisational infrastructure and will then be implemented if considered necessary and relevant to the designated community (see <https://wiki.de.dariah.eu/display/TextGrid/Organisational+Infrastructure#OrganisationalInfrastructure-OrganisationalInfrastructureandLong-TermSustainability>).

The technical metadata is extracted for every TextGrid object during the publishing process and is then available for every TextGrid object. It is stored as an extra file beneath the data and metadata files. You can get the object's technical metadata from textgridrep.de on the object site (see Tech. Metadata (XML):

https://textgridrep.org/browse/-/browse/3qxtk_0) or directly from the TG.crud via <https://textgridlab.org/1.0/tgcrud/rest/textgrid:3qxtk.0/tech> .

The repository has all necessary rights to undertake data migration or transformation if technically required in the future (<https://textgrid.de/terms-of-use>).

In terms of data re-use the TextGrid Repository takes into account the DARIAH-DE Research Data Lifecycle as described in the following publication: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-4-4> . This lifecycle and its processes can be further specified for the TextGrid repository in terms of a publication process. The publication process is a central part

of scientific research in the humanities and for the TextGrid Repository and VRE. Publication is the basis for the dissemination of research results and data to allow all researchers to refer (cite) to each other's work. New research can follow up on the basis of already existing scientific results and sources. The TextGrid Repository covers the fundamental processes of the DARIAH-DE research data lifecycle in its publication lifecycle within its framework of the virtual research environment. For an illustration and short description in English see (especially the final part of it):

<https://wiki.de.dariah.eu/display/TextGrid/Data+Policies#DataPolicies-DataReuse> .

Trainings and guidance for data users are offered through the respective support and communication channels:

- consultation of research projects (textgrid-support@gwdg.de)
- online-tutorials (<https://textgrid.de/en/tutorials>)
- email for technical support (textgrid-support@gwdg.de)
- user's mailing list (<https://textgrid.de/en/maillinglisten>)
- user administration (register@dariah.eu)
- self-service for registration

(<https://auth.de.dariah.eu/cgi-bin/selfservice/ldapportal.pl?mode=selfreg&active-menuitem=selfreg>)

Trainings are regularly offered in TextGrid user meetings and within the framework of DARIAH-DE events related or including the TextGrid VRE and Repository as topic. The meetings and workshops are as well for the research projects and researchers that publish data as well as for data users, see:

- <https://textgrid.de/en/veranstaltungen>
- <https://de.dariah.eu/veranstaltungen>

All responsibilities and tasks for the repositories are related to dedicated persons and institutions and stated in an internal document in terms of full time equivalents related to the DARIAH-DE operating cooperation agreement ("TextGridRep_FTE").

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

TECHNOLOGY

XV. Technical infrastructure

R15. 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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The technical infrastructure of the TextGrid Repository runs on a well-supported operating system. The hardware, software and used technologies are appropriate to serve nationally and internationally research, teaching and learning by providing long term preservation, further processing, openly sharing and dissemination of digital research data according to ethical and scientific standards of the international research community. The documentation of the main components of the technical infrastructure is accessible publicly:

<https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management#DigitalObjectManagement-TechnicalInfrastructure>
, <http://www.textgridlab.org/doc/services/submodules/tg-search/docs/index.html> .

The repository is hosted by the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen mbH (GWDG), which operates as computing centre and IT competence centre for the University of Göttingen and the Max Planck Society. Moreover, the GWDG is also certified with the ISO 9001 certification (https://www.gwdg.de/documents/20182/62973/Zertifikat+ISO+901-2008_2016-05-20+bis+2018-09-14_deutsch.pdf – recertification is in progress). Most of the common services needed for the operation of the TextGrid Repository are provided by the data centre. More specifically, the infrastructure of the TextGrid repository at GWDG is hosted on a virtualised environment. The GWDG is employing multiple virtualisation platforms with different service level agreements.

The TextGrid repository is hosted on the VMware-Cluster (Elastic Sky X – ESX), which provides a high service level agreement at GWDG. This includes:

- higher availability of the service

- higher bandwidth to allow faster use of the infrastructure
- best effort to provide around the clock connectivity

In addition, to ensure a consolidated operation of the various Virtual Machines (VMs) of the TextGrid Repository, the Puppet framework (<https://puppet.com>) is employed, which is also provided by GWDG. The function of Puppet is basically to ensure that all the common services are distributed to the various TextGrid VMs.

The technical infrastructure of the TextGrid Repository supports all the functions the repository fulfils, namely Ingest, Archival Storage, Data Management, Administration, Access, Preservation Planning, common Services – in order to meet the needs of the scientific community. In principle, the TextGrid Repository composition complies with the OAIS reference model, as it is composed of three main components, which cover the above listed functions.

We refer to these components of the TextGrid repository as the so called "Three Pillars of TextGrid":

1. TG-auth* (<https://textgridlab.org/doc/services/submodules/tg-auth/docs/index.html>): The function of this component is to provide user authentication and authorisation. Whereas the authentication functionality is realised via Shibboleth (see DARIAH AAI), and the authorisation by the OpenRBAC model, using LDAP databases.
2. TG-search (<https://textgridlab.org/doc/services/submodules/tg-search/docs/index.html>): The search components enables to query about datasets which contain a specific requested pattern. The search functionality is based on the underlying Elasticsearch and Sesame databases.
3. TG-crud (<https://textgridlab.org/doc/services/submodules/tg-crud/service/tgcrud-webapp/docs/index.html>): As the core component of the TextGrid repository, the function of the CRUD component is to provide an abstraction layer on top of the underlying archival storage system. Hence, TG-crud is used for creation, retrieving, updating and deletion of TextGrid objects stored in the archival storage system. (see for further description of the archival storage system section "IX. Documented storage procedures"). This component incorporates all the functions of the remaining components. Ingest, Access and Data Management are realised by an interaction between TG-crud and TG-auth*.

During Ingest, which is accomplished by the CREATE operation, TG-crud basically checks the permissions, creates a Textgrid URI, processes the metadata and stores the actual data and the metadata into the attached storage system. In addition, TG-crud also provisions the search index with the incoming processed metadata, which then constitutes the overall Data Management.

Access is realized by the READ operation, which first checks the permissions and is followed by the object retrieval.

The updating of existing objects then again is in principle performed in the same manner as for Ingest. However, TG-crud in addition computes new revision number for the new version of the object.

The administration of the repository is accomplished by a 24h monitoring of components. In case of failures, the responsible persons are notified to initiate the troubleshooting during business hours. For failures, which cause the repository to be unavailable, a notification is sent to the users. The preservation is then ensured by the interaction between the PID system, TG-crud and the archival storage system, which was already described in section IX.

Over time the components of the TextGrid Repository have been adapted to newer technological developments to improve specific aspects, such as the performance or stability. The TextGrid Repository service infrastructure is now fully integrated in the DARIAH-DE research infrastructure. Further connections with DARIAH-DE services are discussed and planned in the near future.

The source code of the above described services is open source and available at TextGrid's GIT repository (see <https://projects.gwdg.de/projects/tg>). Building and deployment of the services are supported by continuous integration and continuous deployment using Jenkins (<https://ci.de.dariah.eu/jenkins/>), Nexus (<https://nexus.gwdg.de>), Sonarqube (<https://ci.de.dariah.eu/sonarqube/>), and Puppet. Apache Tomcat is used as a Java servlet container to provide the TextGrid Repository web services.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The Textgrid Repository monitors relevant issues for long term preservation of data and therefore has elaborated an internal risk management plan including the handling of possible scenarios threatening the technical infrastructure and operation of the repository. This document is an internal document between the Göttingen State and University Library (SUB) and the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) as the two relevant institutions running the repository. The plan relates risk scenarios to measures to be undertaken which themselves are related to automatic processes and responsible persons to undertake certain actions. Each of these responsible persons has one back-up person to realise the same responsibilities in case of unavailability,

To detect outages and failures, the GWDG is monitoring all its technical infrastructure. To detect more specific parameters, the developers/maintainers of the TextGrid Repository – which are located at both above mentioned institutions – are also monitoring individual components. This components are also documented in the public wiki (<https://wiki.de.dariah.eu/display/TextGrid/Digital+Object+Management#DigitalObjectManagement-SecurityIssues>). The list of monitored components and services is quite long, but the most relevant components for the administration of the daily operations and maintenance of the repository are:

- Nginx, the proxy web-server on top of the underlying repository components which allows to keep the internal architecture of the TextGrid Repository isolated from world wide access, which serves as a protection component and access point to all TextGrid components
- Apache is dedicatedly employed as a proxy for TG-auth and TG-noid
- TG-auth is the authentication and authorisation system of the repository which manages access rights and permissions
- TextGrid LDAP (internal) is the database for authentication and authorisation, where all relevant data or access rights and permissions are stored.
- DARIAH IdP regulates access via the federated identity provider Shibboleth (TextGrid is accessible via Shibboleth)#
- TG-crud represents the data management-system for creating, updating, retrieving and deleting of data
- TG-search is the search engine on top of the TextGrid Repository
- Elastic-Search is one of the underlying technologies of TG-search used for search functionalities of the repository
- Sesame, the RDF database for rdf queries, that are related among others to the versioning system of the repository
- TG-noid, the component which generates internal identifiers (naming component) for the objects and datasets

- StorNext, represents the archival system (storage system) of the repository
- Tivoli Backup, the backup-system of the Repository
- ePIC PID, the reference system providing persistent identifiers to deposited data and objects

In case of an outage, there are clear procedures to ensure re-operation of all the services provided by the GWDG and SUB. Furthermore, core services provided by the GWDG, such as storage systems and virtual machines are hosted on different locations in the city of Göttingen.

Furthermore, the ISO 9001 certification of GWDG

(https://www.gwdg.de/documents/20182/62973/Zertifikat+ISO+901-2008_2016-05-20+bis+2018-09-14_deutsch.pdf, <https://www.iso.org/standard/46486.html>) also shows that there are several defined procedures to ensure an efficient operation of the data centre, e.g.:

- the process “06 – Administration and System Maintenance”, defines the general measures, which have to be complied with to administrate the various systems in place.
- the process “07 – Backup and Archive”, defines steps, which have to be taken into account for the backup and archival systems.

In addition, the document “Disaster Plan” describes for all the important services and hardware machines the responsible persons and their reachability. The document “Actionmap” describes the general steps, which have to be performed in case of an emergency situation. By another document, the “IT-Security Guidelines”, the GWDG describes the measures to mitigate threats related to IT security. This document also specifies the persons at the GWDG who are in charge with the IT-Security topic.

For a smooth run of the TextGrid Repository IT security system, operating system updates are done weekly by the system administrators, and rebooted if necessary (on updates such as Linux security kernel updates). The system administration and maintenance access to the TextGrid Repository machines are managed using the DARIAH-DE Puppet configuration (scope: textgrid) and can only be accessed and updated by authorized personnel. Remote access to TextGrid Repository machines is only possible by the use of SSH and RSA public keys security infrastructure, remote access using username/password and remote root access is disabled on all machines.

Bug Fixing of security issues is done by the service developers and maintainers using the Gitflow workflow (<https://datasift.github.io/gitflow/IntroducingGitFlow.html>), fixed services are deployed and released regularly and security issues are released immediately after implementation, testing, and deployment.

For the secure operation of the TextGrid Repository there is no single dedicated security officer, as a whole team of developers and maintainers from SUB and GWDG is monitoring the whole system and undertakes necessary actions whenever necessary. This is organised and coordinated for the core infrastructure as already described by the disaster plan.

In case of a breach of the implemented security measures, the whole system and (virtual) machines can be set up from scratch using the Puppet configuration and the data can be imported again from daily backups.

The TextGrid Repository services are monitored not only for correct operation using Icinga/Nagios (<https://icinga.de.dariah.eu/icinga/>), they are also constantly monitored using Grafana (<https://metrics.gwdg.de>) for system analytics and possible service misuse (security scanning), such as unusual high server load or low performance of certain services. Developers and maintainers are notified by email in such cases.

The access to the machine room is strictly regulated by the document "Machine Room Access Rules". Finally, to prevent threats from social media, GWDG also has a Social Media Guideline.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: