1 Organisational Infrastructure

2 Content

- Mission and Long-Term Preservation
- The TextGrid Repository and its Designated Community
- Organisational Infrastructure and Long-Term Sustainability
- Responsible Institutions: Rules and Obligations
  - DARIAH-DE Services and Tasks
  - CLARIAH-DE Services and Tasks
  - Expertise of the SUB and GWDG

3 Mission and Long-Term Preservation

The mission of the TextGrid Repository (TextGridRep) is to serve nationally and internationally 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 research community. The publicly stated mission (https://textgridrep.org/) is approved by the DARIAH-DE Coordination Office (https://de.dariah.eu/en/kontakt).

The repository sees his mission in line with the Open Access strategy of the University of Göttingen and its research data policy. It provides all necessary resources to promote and support making the research results of its researchers as widely accessible and usable as possible. This commitment to open access is reflected in the organisational and technical infrastructure as well as in its archiving procedures of the repository to allow the use of publications and data without any access restriction in order "to support research and innovation in science […] and society in a direct and lasting way". In terms of data management, publication and preservation workflows are based on the Open Archiving Information System, see TextGrid Repository – Digital Object Management.

The commitment is strongly supported by the two relevant institutions ensuring also the long-term sustainability of the repository and its data: The Göttingen State and University Library (SUB) and the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen mbH (GWDG).

Both institutions share a commitment to the sustainability of services and to FAIR principles in research and its infrastructures. For the SUB research data management is an important aspect of the strategic aims of Göttingen State and University Library. Not only for research data, but for all digital resources, Göttingen State and University Library follows a policy, which contains guiding principles in order to ensure the quality for access, metadata and IT architecture.

In the context of open access, the Göttingen State and University Library also participates in national and international projects, such as the Confederation of Open Access Repositories (COAR) and OpenAIRE. In this perspective the TextGrid Repository is also in line with open access requirements of important funders of the German research system as the German Research Foundation (DFG) (see https://www.dfg.de/formulare/2_00/v/dfg_2_00_de_v1215.pdf, p. 44, section 12.2.1) and the European Union. Mandates of the European Commission and the European Research Council require as stated e.g. in the European Open Access Pilot on Open Data all funded projects to publish their results in Open Access (see the Horizon 2020 Online Manual). The Research Department at Göttingen University offers detailed information about the European Union Open Access Pilot also on its web pages.

The TextGridRep is a disciplin specific repository and commits itself to ensure availability and long-term preservation of data in the Humanties. Because of its designated community and its history the TextGrid Repository defines furthermore itself as an 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. It is part of a virtual research environment and linked to the TextGrid Laboratory 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.

4 The TextGrid Repository and its Designated Community

The TextGrid Repository is the result of the community driven project TextGrid running from 2006 to 2015. Organisational structure, practices, policies and content are related to its history and community and are only understandable against this background.

From its beginning, the projects aim has been to develop a virtual research environment (TextGridVRE) for digital scholar editing, for collaborative creation, analysis and publication of text and images. In continuous exchange with the research community TextGrid developed within its three funding phases several tools and services in order to answer an increasing demand for digital and collective research features in the humanities.

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 TextGrid project pages of first and second funding period, and TextGrid III).

At the beginning mainly scientist from different philologies, art history and musicology have been involved. In the project lifetime more and more scientists from further disciplines in the humanities dealing with digital editions and text based research became part of the TextGrid community. The most represented discplines 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

Furthermore, many cooperations with a variety of research projects using TextGridVRE have been established (see Collection Development Policy and Data Quality/Re-Use, Fig. 1). In terms of a more abstract categorizations, three different target groups can be stated for the repository and its virtual research environment:

- Scholars of the Arts and Humanities using TextGrid for their research projects
- Developers adapting TextGrid tools and services for specific scholarly needs
- Academic institutions (such as archives and libraries) storing data in TextGrid or linking data to own data bases

Since 2016 – when the funding of TextGrid by the Federal Ministry of Education and Research (BMBF) ended – technological core components such as the user administration and repository technologies have been migrated into the digital research infrastructure DARIAH-DE - Digital Research Infrastructure for the Arts and Humanities (also funded by the BMBF) as to guarantee a sustainable and long-term usage of TextGrid's state-of-the-art services.

The TextGrid Virtual Research Environment consists of the TextGrid Laboratory (TextGridLab) as client application and the TextGrid Repository (TextGridRep) which are linked to each other. Via the TextGridLab, researchers can access as a single point of entry specialized tools, services and content to create, manage and edit their XML-based research data and publications. Core elements are the project and user administration of a non publicly available area (OwnStorage), the text-image-link editor, the integrated xml editor, dictionaries, search functionalities and access to aggregations of data. Further open source tools and services optimised for use with TextGrid are available for integration via the "Market Place" of the TextGridLab. In the perspective of digital scholarly editing, storing and publishing the TextGridRep is the publication archive and at the same time one data source of the virtual research environment.

Research projects working with the TextGridLab finally publish their usually peer reviewed research results and related data in the PublicStorage of the TextGridRep to make it publically available – open access and for free. This includes also new versions of digital editions or restructured XML schemas to adapt data for special research questions. Content of the repository may be considered – depending on the perspective of its users – simultaneously as publications, research results, research data for new endeavours, teaching material or part of the ongoing digital transcription of our cultural heritage. Hence, the repository is to be considered as a publication repository and a data repository. Primary text material and xml/TEI encoded and marked entities may become relevant (meta)data for research. The Terms of Use as well as mandatory licence indications (usually recommended Creative Common Licences) specify legal frameworks for re-use.

In this context the TextGrid Repository offers with TextGrids Digital Library an additional and important canon of XML/TEI encoded texts for research and teaching whose copyright has expired – ranging from the beginning of the printing press up to the first decades of the 20th century and related to literary and cultural history. All data of the repository is intended to be not only used for reading purposes but mainly for further processing, such as analysing and visualising tools, editions and text corpora.

According to the TextGrid Repository Mission Statement and it's designated community, all content of the repository is open access and publicly available except as for the content administrated by research projects in their OwnStorage area. This area is publically not available and provided with a rights and role management system. The PublicStorage contains all published and therefore publically availalable data. This data are provided with a persistent identifier (ePIC PID) and are accessible via the website of the TextGrid Repository, via the TG-search API as well as via the DARIAH-DE Generic Search. All data can be visualised using different external tools or DARIAH-DE tools such as Digivoy.

Even if the repository is mainly used in a national context it is also open for international scientists and users. As part of the VRE it is already in use within international research projects, such as Maps of God, which is elaborating and editing a digital encyclopedia of diagrams of the Jewish Kabbalah Doctrine involving a research group in Israel. In addition and in line with its open and free access policy (Data Policies) the technical infrastructure allows also the use of the repository independently from the TextGridLab.

5 Organisational Infrastructure and Long-Term Sustainability

In terms of long term operation, stability and sustainability the TextGrid Repository is integrated in a complex organisational infrastructure as illustrated by Fig. 1.
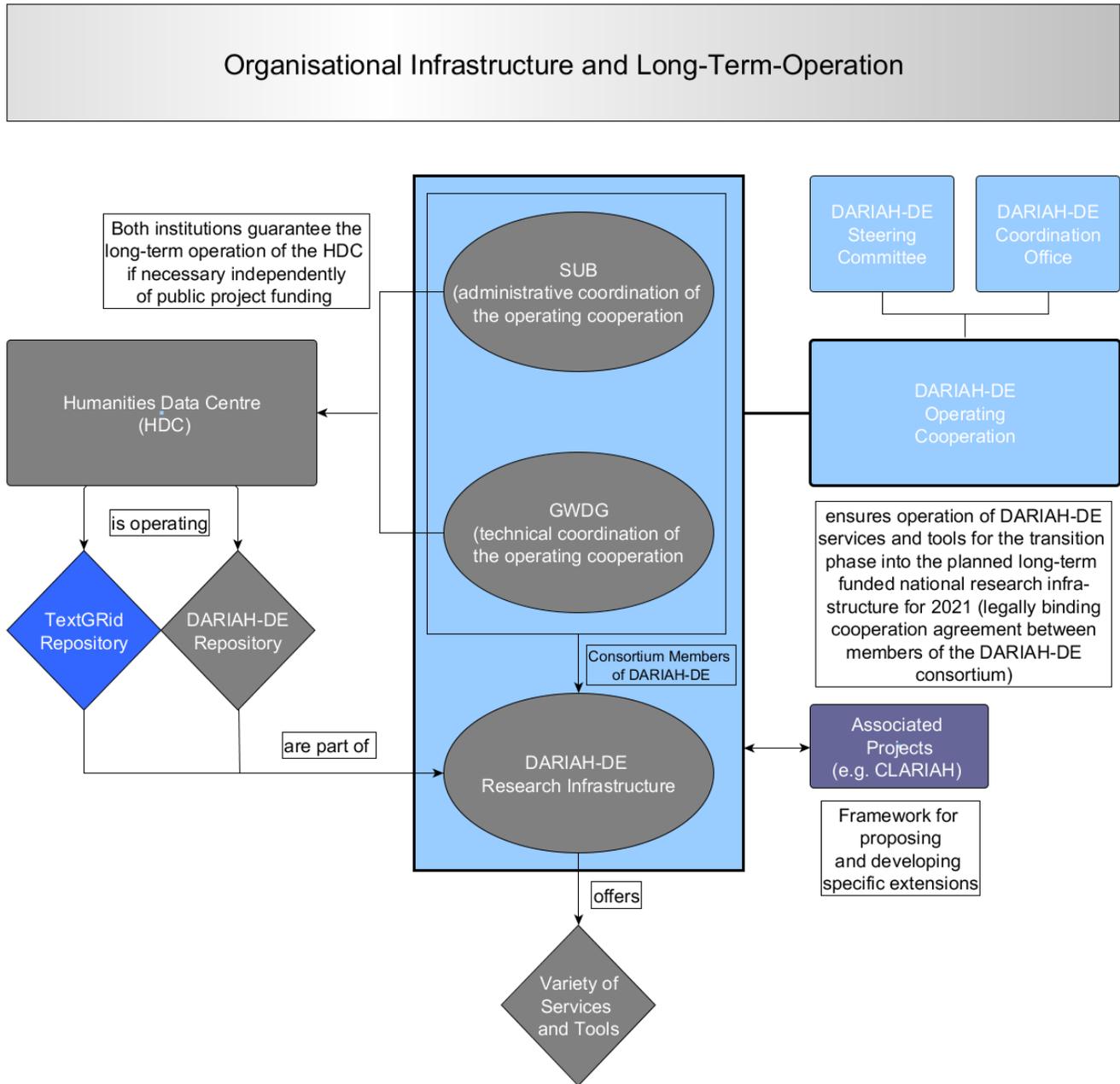

Fig. 1: Organisational Infrastructure and Long-Term-Operation of the TextGrid Repository

First, the TextGridRep is operated by the Humanities Data Centre to ensure its long-term sustainability. The HDC has been founded explicitly for this purpose by the Göttingen State and University Library and the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen mbH, which operates as computing centre and IT-competence centre for the University of Göttingen and the Max Planck Society. Both institutions guarantee the long-term stability and sustainability of the repository in addition and if necessary independently of public project funding – as publicly declared in their founding manifesto.

Second, the TextGrid Repository is since 2016 part of the DARIAH-DE Research Infrastructure which offers a variety of services and tools – the TextGridRep is one of them. Additionally, the operation of the repository is ensured until the end of 2021 by a DARIAH-DE operating cooperation between all consortium members. It's a legally binding cooperation agreement to keep services and tools of DARIAH-DE running for the transition phase into the planned long-term funded national research infrastructure for 2021. The SUB and GWDG as founders of the HDC are also important consortium members of DARIAH-DE. While the SUB is responsible for the administrative coordination of the operating cooperation, 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 Reprository through the DARIAH operating cooperation agreement.

The specific governance structure of the DARIAH-DE operating cooperation ensures that disciplinary and technical innovations are taken into account and implemented to an appropriate extent, The DARIAH-DE steering committee is the joint central decision-making body at cooperation level. In monthly telephone conferences, it discusses and decides about strategic developments and potential disciplinary innovations of the DARIAH-DE Infrastructure, including the TextGridRep. The DARIAH-DE Coordination Office coordinates the DARIAH-DE operational cooperation and is located at the SUB. It also documents the DARIAH steering committee's decisions.

Third, associated projects provide additional resources in means of financial and human resources for proposing and developing specific extensions. The most prominent one for the transition phase into the national research infrastructure is CLARIAH-DE, where the DARIAH-DE and the CLARIN-D consortia are involved and preparing the next steps in terms of integration and developing their service portfolio. 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 the related virtual research environment.
- Project specific extensions are proposed and developed within the framework of projects, for example CLARIAH-DE, whose Work Package 2 (Tools and Virtual Research Environments), among other things, is concerned with TextGrid. Like every work package within CLARIAH-DE, Work Package 2 conducts regular telephone conferences as well as face-to-face meetings.

6 Responsible Institutions: Rules and Obligations

As illustrated above for the organisational infratstructure of the TextGrid Repository, SUB and GWDG as founders of Humanities Data Centre take over a variety of tasks within the framework of the research infrastructure DARIAH-DE operating cooperation (service provider) and of the assocoiated 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.

6.1 DARIAH-DE Services and Tasks

- Göttingen State and Unviersity Library (SUB), adminstrative coordinator of the operating cooperation:
  - TextGridLab Application Management
  - TextGridRep Service Management
  - Consulting for digital text data: standards, file formats etc.
  - Consulting for Digital Editions
  - Consulting for use of the repository
  - Consulting for ingesting large amounts of data
  - Dissemination
  - Maintenance
  - Updates
  - User request
  - Documentation
  - TextGrid workshops and training
- Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen mbH (GWDG), technical coordination of the operating cooperation:
  - DARIAH Authentication and Authorization Infrastructure (DARIAH AAI)
  - Persistent Identifiers
  - Virtual Machines
  - Storage and backup
  - Monitoring

6.2 CLARIAH-DE Services and Tasks

- Project Partners (SUB, GWDG, and more)
  - Evaluation of new features for TextGrid
  - Implementation of new features for TextGrid
  - Evaluation of possibilities for format- and collection development

6.3 Expertise of the SUB and GWDG

- Software engineers
- Information scientists
- Metadata experts
- Experts for digital editions
- Digital Humanists
- Text and data mining experts