Assessment Information

CoreTrustSeal Requirements 2017–2019

Repository:	Tübingen CLARIN-D Repository
Website:	http://www.sfs.uni-tuebingen.de/ascl/clarin-center/repository.html
Certification Date:	27 March 2019
This repository is owned by:	University of Tübingen

Tübingen CLARIN-D 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

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of the Repository's Designated Community.

The Tübingen CLARIN-D Repository (https://uni-tuebingen.de/en/134314) is part of a network of research infrastructure centres located throughout Europe. The aim of these centres is to provide language data, tools and services in an integrated, interoperable and scalable infrastructure to researchers in the humanities and social sciences.

CLARIN (https://www.clarin.eu/) is an acronym for "Common Language Resources and Technology Infrastructure". It is a research infrastructure that was initiated from the vision that all digital language resources and tools from all over Europe and beyond are accessible through a single sign-on online environment for the support of researchers in the humanities and social sciences. The CLARIN infrastructure is fully operational in many countries, and a large number of participating centres are offering access services to data, tools and expertise.

In 2012, nine CLARIN member countries created CLARIN-ERIC (European Research Infrastructure Consortium), which is an international legal entity that governs and coordinates CLARIN activities. CLARIN-ERIC members are governments or intergovernmental organisations which pay an annual fee to support the development and maintenance of the CLARIN research infrastructure.

Germany is one of the founding members of CLARIN-ERIC and contributes to CLARIN-ERIC via CLARIN-D (https://www.clarin-d.net/en/). CLARIN-D is an acronym for "Common Language Resources and Technology Infrastructure Deutschland".

The CLARIN-D Resource Center Tübingen is one of currently eight German CLARIN-D Resource and Service Centers which form a web and centers-based research infrastructure for the humanities and social sciences. The aim of CLARIN-D and its service centres is to provide language data, tools and services in an integrated, interoperable and scalable infrastructure for researchers in the humanities and social sciences and related disciplines. The research infrastructure is rolled out in close collaboration with expert scholars in the humanities and social sciences, to ensure that it meets the needs of users in a systematic and easily accessible way. The CLARIN-D Resource Centre Tübingen is part of the CLARIN-D consortium funded by the German Federal Ministry for Education and Research.

CLARIN-D is building on the achievements of the preparatory phase of the European CLARIN initiative as well as CLARIN-D's Germany-specific predecessor project D-SPIN. These previous projects have developed research standards to be met by the CLARIN service centres, technical standards and solutions for key functions, a set of requirements which participants have to provide, as well as plans for the sustainable provision of tools and data and their long-term archiving.

Within CLARIN, this resource centre is a certified centre of type B. CLARIN distinguishes a number of different centre types that have different impact for the language resources and tools infrastructure. Type B centres offer services that include the access to the resources stored by them and tools deployed at the centre via specified and CLARIN compliant interfaces in a stable and persistent way.

The following requirements hold for CLARIN centres of type B, and are fulfilled by this resource center:

- Centres need to offer useful services to the CLARIN community.

- Each centre needs to refer to CLARIN in a visible way on its website.

- Each centre needs to make explicit statements about its funding support state and its perspectives in this respect.

- Each centre needs to make explicit statements about CLARIN compliant resources and services available at the centre.

- Each centre needs to make clear statements about their policy of offering data and services and their treatment of IPR issues.

- The centre has to implement the GÉANT Data Protection Code of Conduct (DP-CoC) for each of its federated Service Providers.

- Centres need to have a proper and clearly specified repository system and participate in a quality assessment procedure as proposed by the CoreTrustSeal.

- Centres need to adhere to the security guidelines, i.e. the servers need to have accepted certificates.

- Centres need to join the national identity federation where available and join the CLARIN service provider federation to support single identity and single sign-on operation based on SAML2.0 and trust declarations.

- Centres need to offer component based metadata (CMDI) that make use of elements from accepted registries such as the CCR in accordance with the CLARIN agreements, i.e. metadata needs to be harvestable via OAI-PMH.

- Centres need to associate (handle) PIDs with their metadata records. These PIDs should be suitable for both human and machine interpretation, taking into account the HTTP-accept header. Individual files (e.g. a text, zip or sound file) can be referred to with either the PID of the describing metadata record in combination with a part identifier or with another PID.

- Centres can choose to participate in the Federated Content Search with their collections by providing an SRU/CQL Endpoint.

A short overview of all requirements for centres of type B is also given in the form of a checklist (https://office.clarin.eu/v/CE-2013-0095-B-centre-checklist-v6.pdf).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

B. Basic curation – e.g. brief checking; addition of basic metadata or documentation

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

This repository distributes data as deposited. However, prior to ingestion into the repository, data is checked for suitability for inclusion, and extensive metadata are created. Among the resources currently available in this repository, researchers can find widely used treebanks of German (e.g. TüBa-D/Z), the German wordnet (GermaNet), the first manually

annotated digital treebank (Index Thomisticus), as well as descriptions of tools used by the WebLicht execution engine for natural language processing.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Outsource Partners. If applicable, please list them.

1) Gesellschaft für Wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
The repository makes use of a common CLARIN PID service (https://www.clarin.eu/files/pid-CLARIN-ShortGuide.pdf) based on the Handle System (http://www.handle.net/) and in cooperation with the European Persistent Identifier Consortium (EPIC). The usage of PIDs is mandatory for resources in CLARIN thus all resources added to the repository may be referenced using PIDs. CLARIN-D has a contractual relationship with GWDG concerning the provision of PID-services via EPIC API v2. The following document lists the services which are stipulated: http://www.clarin-d.de/mwiki/images/0/0b/GWDG_PID.pdf

2) CLARIN-D
The repository in one of currently eight resource and service centres of CLARIN-D. As part of the CLARIN-D consortium, the repository has signed the "Kooperationsvereinbarung" - Cooperation Agreement - which states the rights and obligations of all CLARIN-D centres. A condensed version of this contract (in German only) is available at: https://www.clarin-d.net/de/ueber/zentren/zusammenarbeit

CLARIN-D offers several services to it's member institutions, among them the following:

- CLARIN-D HelpDesk (https://support.clarin-d.de/mail/): A central system for user support, which allows for the distribution of user questions and feedback to qualified personnel at the centres.
- CLARIN-D website (https://clarin-d.de/en/): A starting point for researchers to find information on CLARIN-D and to access CLARIN-D services.
- CLARIN-D wiki (https://www.clarin-d.de/mwiki/index.php/Hauptseite): A central platform for CLARIN-D-related staff.
- CLARIN central monitoring (https://monitoring.clarin.eu/): A monitoring service offered to all CLARIN-ERIC members and maintained by the resource centre Leipzig.

3) CLARIN-ERIC
CLARIN-D is a member of CLARIN'S European Research Infrastructure Consortium (ERIC). CLARIN-ERIC offers central services to it's members and users, as stated here: https://www.clarin.eu/value-proposition

The services are available to all centres in the member countries of the CLARIN-ERIC (https://www.clarin.eu/content/overview-clarin-centres).

The most important services of the ERIC cover the search functionality for the German CLARIN centres:

- Virtual Language Obervatory - VLO (https://vlo.clarin.eu): CLARIN's central metadata-based search engine, which contains metadata of all German CLARIN-centres.
- Metadata harvester: The VLO is kept up to date using the metadata harvester run by the CLARIN- ERIC.
- Federated Content Search - FCS (https://www.clarin.eu/contentsearch): Optionally, centres can provide the actual data of their resources for this central content search.
- CMDI Component Registry (https://catalog.clarin.eu/ds/ComponentRegistry): CLARIN's registry for components and profiles according to ISO-24622-1.

In addition, CLARIN-ERIC offers several further services such as central registries, user statistics management and, as an official EUDAT community, access to advanced EUDAT services.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Other Relevant Information.

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 mission of the Tübingen CLARIN-D Repository is to ensure the availability and long-term preservation of resources in the field of Humanities and Social Sciences, to preserve the knowledge gained in research, to aid the transfer of knowledge into new contexts, and to integrate new methods and resources into university curricula.

This mission is supported by the infrastructure of the University of Tübingen and by the integration of the repository into the national and international CLARIN infrastructures. As part of the CLARIN-D infrastructure, it shares the CLARIN-D mission to provide linguistic data, tools and services in an integrated, interoperable and scalable infrastructure for the Humanities and Social Sciences (https://www.clarin-d.net/en/about), and is committed to play an active role in the development of CLARIN's repository infrastructure.

The CLARIN-D center in Tübingen supports data from the Humanities and Social Sciences with a clear emphasis on language related material, both for disciplines working with language analysis as the objective of research and as a research method. This covers data especially from Linguistics, Psycholinguistics, Corpus Linguistics, Syntax, Semantics, Lexicography, etc. but also includes other areas such as literary studies, political sciences, history, etc.

For an overview of the mission and goals of the CLARIN research infrastructure, see the following publication by Erhard Hinrichs (national coordinator of CLARIN-D) and Stephen Krauwer (former executive director of CLARIN-ERIC):

Hinrichs, E.; Krauwer, S. (2014a): The CLARIN Research Infrastructure: Resources and Tools for E-Humanities Scholars. In: N. Calzolari et al. (Eds.), Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). 1525–1531. ELRA, Reykjavík, Island. http://www.lrec-conf.org/proceedings/lrec2014/pdf/415_Paper.pdf

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

Upon renewal, the repository is asked to provide a formal mission statement online as opposed to in an academic paper.

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:

Before data can be deposited into the data repository, the depositor must provide an appropriate End User License Agreement (EULA) and sign a Depositor's Agreement.

The EULA is an agreement between the depositor and the user, is provided by the depositor, and must be accepted by the user before obtaining the data. The repository encourages depositors to place data under open licenses such as Creative Commons whenever possible. Information about a dataset's licensing is stored in the CMDI metadata. In case of misuse of data, legal action may be taken by the depositor or property rights owner.

The Depositor's Agreement is an agreement between the depositor and the repository, where the repository is represented by the University of Tübingen. The Depositor's Agreement includes granting distribution rights to the repository, specifying access rights to the data (public, academic, individual), and assures that IPR and privacy rights are respected in the deposited data. The data provider retains all intellectual property rights to their data.

Access rights are enforced technically and can restrict download of a dataset to certain individuals or to the academic community. For some resources (e.g. those with individual access), the user may need to sign a license agreement with the depositor before the repository can give access to the resource. This is the case for datasets which are based on copyrighted material, such as newspaper text, and a license is required to protect the rights of the copyright holder. In this case, credentials are provided to the individual, which allows them to download the dataset. Access to resources that are limited to academic use is protected via Shibboleth and is only available to those that are able to login through IDPs operated at institutions taking part in the DFN-AAI or similar AAI federations that are part of CLARIN. This currently includes over 2,000 academic institutions in Europe.

If a problem is discovered in a dataset (such as personal data disclosure), the depositor will be contacted immediately, and steps will be taken to ensure that the data is not distributed until the issue can be resolved. These steps include blocking download access to the dataset and preventing the metadata from being harvested. Additionally, records are kept by the repository for those resources which require the user to sign a license before gaining access credentials. These records of licensed users enables the repository to contact users in case of a legal issue with a specific version of a dataset, for example in case of legal action against a depositor from a third party. Users can also be informed if a new version of the dataset, in which legal conflicts have been removed, becomes available. The records of users are kept private according to German privacy legislation.

Please see the repository agreements and guidelines for more information(https://uni-tuebingen.de/en/134320).

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:

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:

As stated in the Depositor's Agreement (https://uni-tuebingen.de/en/134320), the repository ensures that the deposited data will remain archived in a legible, accessible, and sustainable manner to the best of its ability and resources.

All CLARIN centers commit to ensuring long-term availability, access to and preservation of datasets submitted to their repositories, as set out in their mission statements. CLARIN centers are set up as a distributed network, where each center institution is a hub of the digital humanities and brings its own financial resources into CLARIN-D, which ensures continued availability. Thus, in case of a withdrawal of CLARIN-D funding, University of Tübingen funding would be used to ensure intermediate-term maintenance of the Center infrastructure, during which time the repository's content would be moved to another CLARIN center if possible. The legal aspects of the process of relocating data to another institution is addressed by templates of license agreements provided in CLARIN.

The depositing agreement makes provisions to allow such a transfer between institutions maintaining the same access restrictions – if any – in the case of a transfer of data to another CLARIN center. This is especially important as there may not be any other contractual relationship between a depositor and a data center stepping in for another center.

All CLARIN-D centers have agreed to a Memorandum of Understanding for taking over research data by other CLARIN-D centres (https://www.clarin-d.net/en/about/centres/mou-taking-other-centre-s-data), which states the terms under which the centers agree to transfer data from one center to another.

The repository is currently in the process of developing its future strategy, allowing it to guarantee preservation periods to data depositors. Discussions with the BMBF, the state of Baden-Württemberg, and the University of Tübingen are ongoing.

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 process of depositing data into the repository at the CLARIN-D center in Tübingen includes steps to ensure that the legal and ethical requirements of the archive are met. This is achieved organizationally through a contractual agreement (Depositor's Agreement) between the depositor and the archive, and also through close collaboration with the depositor prior to ingesting the data into the repository.

The Depositor's Agreement explicitly states that the depositor has fulfilled ethical and legal obligations, has resolved any disclosure risks related to privacy issues, and has respected Intellectual Property Rights (IPR) with regard to the data. In particular, data must be anonymized when applicable. Data depositors are responsible for compliance with all relevant national or international legal regulations. The depositor can choose to make the data publicly available, restrict access to academics via AAI (Authentication and Authorization Infrastructure), or to restrict access to individual users (which may be necessary if the dataset is based on copyrighted material such as newspaper text).

Only data that is in compliance with the Depositor's Agreement and with the University of Tübingen's guidelines for safeguarding good scientific practice (https://uni-tuebingen.de/en/119123) will be considered for depositing. In addition, we ask the depositor whether the data to be deposited contains any parts that have been contributed by third parties, and thus constitute a potential disclosure risk. If so, written documentation that the third party has consented to redistribution of the data must be provided by the depositor.

Neither the CLARIN-D resource center, nor the repository run by it, are legal entities on their own. This also holds for the General and Computational Linguistics Department ("Seminar für Sprachwissenschaft", SfS) where the CLARIN-D center and its repository are located. All are part of the University of Tübingen, which is a legal entity - specifically, like all public German universities, a Körperschaft des öffentlichen Rechts, an institution governed under public law. Hence, the

university as an institution is the contractual party in the depositor's agreement with appropriate authorities signing the contract.

The Depositor Agreement is governed by German law, hence the authoritative version is in German. An informative version in the English translation has also been made. Both are available here: https://uni-tuebingen.de/en/134320. For legal reasons, these agreements are templates only, to be adjusted on a case-by-case basis. A Depositor's agreement must be signed prior to depositing data.

The repository encourages users, depositors, and researchers to report violations. The repository has created an email address for this purpose, and a means for reporting violations will be added to the repository website by the end of 2018. In the case where the repository discovers that the Depositor's agreement has been violoated, distribution of the data and metadata will be halted, with the possibility of ingesting a new version in which the problematic parts have been removed. In the case of violation of the EULA by a user, further access by that user may be blocked. In all cases, the original data provider will be contacted.

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:

This repository is hosted by the CLARIN-D center Tübingen within the General and Computational Linguistics Department of the University of Tübingen. The center currently has seven staff members, four of whom are responsible for CLARIN-D technical activities, including operation of the repository. The repository staff have both the technical expertise and knowledge of language data required to ensure the safety of the repository and the quality of its data.

CLARIN centers are hosted by scientific institutions - their repository staff members have access to training on data management, metadata, long-term preservation and professional development (offered by CLARIN-D and CLARIN-ERIC). This includes regular developer meetings, mobility grants for sharing of expertise, conferences, meetings with their respective scientific communities (called discipline-specific working groups) as well as a centralized knowledge base (user guide, wiki, bugtracker and mailing lists). CLARIN has a wide field of expertise in its collaborative network of centers, which come from within their respective fields of digital humanities.

As part of CLARIN-D, staff members also have access to information on a wide range of topics that CLARIN-D offers, including teaching material, help on data management plans and other, discipline-specific support.

By being part of the CLARIN-D consortium, the repository gains access to funding for running and further developing a sustainable repository and resource center to support its goals. Besides staff resources, this includes a budget for attending national and international meetings such as conferences, workshops or internal developer meetings and meetings with the subject-specific working groups.

Currently CLARIN-D is funded by the Federal Ministry of Science and Education of Germany (Bundesministerium für Bildung und Forschung, BMBF). The current project phase has a runtime of 4 years and is funded until 30.09.2020. The Tübingen CLARIN-D center is also supported by the University of Tübingen with 2 FTE (full-time equivalent) positions.

The tasks of the staff members are stated in the "Division of Labour" of CLARIN-D. An overview is available at: https://www.clarin-d.net/en/about/centres/division-of-labour-of-clarin-d-centres

As an alternative to project based funding, CLARIN-D currently pursues a permanent continuation of funding. Discussions and negotiations on this topic are ongoing.

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:

The repository, through its membership in CLARIN-D, is supported by external advisory committees.

The International Advisory Board (IAB), CLARIN-D's scientific advisory board, is a group of CLARIN-D external experts who are consulted on new developments and discuss strategic and content related developments, also with a bird's-eye view of other developments in the communities. With experienced experts from various backgrounds, a high-profile international committee was formed for this purpose. Members of the IAB are currently: Reinhard Altenhöner, Christiane Fellbaum, Björn Granström, John Nerbonne, Heike Renner-Westermann and Achim Streit.

The joint Technical Advisory Board (TAB) of CLARIN-D and DARIAH-DE, is a committee that supports collaboration on the fundamental technical level between the two large research infrastructures for the humanities and social sciences. The issues of the Collaboration are: questions of technical protocols, infrastructural requirements on the level of archiving, interconnection, search, etc. Based on requirements, small working groups (for example on persistent identifiers, authorization and identification) are being formed in areas with an overlap of requirements. This avoids duplication of developments and allows an increased efficiency in implementation, but also interoperability where overlaps exist. This includes for example an option to grant access to one infrastructure for users of the other. Members of the Technical Advisory Board are currently: Jan Haji■ (Prague Institute of Formal and Applied Linguistics, CLARIN Center), Margareta Hellström (ICOS Carbon Portal staff member), Peter Leinen (German national library, head of information technology), Karlheinz Mörth (Austrian Acacemy of Sciences), Wolfgang Nagel (Technical University of Dresden, Head of the centre of information services and high performance computing), and Sabine Roller (University of Siegen, head of the centre of

information and media technologies).

CLARIN is committed to boosting humanities research in a multicultural and multilingual Europe, by facilitating access to language resources and technology for researchers and scholars across a wide spectrum of domains in the humanities and social sciences (HSS). To reach this goal and to contribute to overcoming the traditional gap between the Humanities and the Language Technology communities, we established an active interaction with the research communities in HSS in so-called discipline-specific working groups.

These groups act as a link between the CLARIN-D resource centres and the research communities which represent the users of the CLARIN-D infrastructure. Currently eight working groups act as consultants for the needs of the humanities, social sciences and particular disciplines. All together they consist of more than 200 academic professionals. Their main role is to advise CLARIN-D during the development and implementation of the infrastructure so that these efforts can best meet the needs of all research communities involved. The working group chairs further coordinate dissemination and best practice using CLARIN-D services in their member communities.

CLARIN-D organizes joint activities of the working groups. This includes the organization of working group meetings, organization of specialized and interdisciplinary workshops and the creation of joint reports. Further, communications between CLARIN-D centres and the working groups as well as groups among themselves are coordinated. Virtual meetings are held on a bi-monthly basis. Activities of the WG are described on the CLARIN-D Website (https://www.clarin-d.net/en/disciplines/). For communication, mailing lists and wiki contents are maintained.

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 integrity and quality of digital objects are ensured by two processes: A manual process at the time of ingestion, and an automatic process for continual monitoring of integrity. The identity of data providers is known and a Depositor's Agreement must be signed before data is archived.

Prior to ingest, an initial quality check and assessment of the provided data and metadata is performed by the archive manager. This involves checking the consistency and validity of the metadata and a check of the data provided. The metadata is validated for syntactic correctness and manually evaluated for completeness and soundness. Software tools (both standard and custom-built) support the archive manager in this process, for example in checking the validity of an XML document against its schema, or checking metadata for consistency. Any issues that arise can be resolved between the depositor and the data manager prior to ingest.

The integrity of the data is ensured by the version control in the Fedora-Commons backend. Metadata is a data stream within the digital object, and as such is version-controlled like object data. The system performs integrity checks of the individual data streams based on MD5 checksums. Problems and changes of files are reported to the archive manager for immediate action and restoring from backup if necessary.

The repository stores data but does not process or alter it in any way. Alterations of primary data is not allowed, but new versions of the data may be made available. New versions are assigned a version number and are stored in a separate data stream, which has an associated checksum which is automatically computed by the repository. In the rare case where a datastream needs to be updated, the previous version is automatically maintained by a version control system built into the repository back end. Metadata may be updated if need arises by the data depositor or the archive manager (e.g. to update contact information, addresses, or to add descriptions in other languages, etc). The repository includes an RDF store of system metadata and also an audit trail which can be used to inspect past activities.

Metadata according to ISO 24622-1 (CMDI; http://www.clarin.eu/cmdi/) is uploaded or created during the archiving process. This step is required in the uploading process, since data without metadata is technically not accepted in the system. The front-end of the archiving system includes software to assist the depositor in creating valid CMDI metadata using components and profiles stored in the Component Registry (http://catalog.clarin.eu/ds/ComponentRegistry/).

Transformation into MARC21, DublinCore and HTML are currently being tested to support interoperability with other standards. A roll-out of this additional service is planned by the end of 2018.

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:

This repository supports data from the humanities and social sciences with a clear emphasis on language related material, both for disciplines working with language analysis as the objective of research and as a research method. This covers data especially from Linguistics, Psycholinguistics, Corpus Linguistics, Syntax, Semantics, Lexicography, etc., but also includes other areas such as literary studies, political sciences, history, etc.

Prior to ingesting a dataset into the repository, quality checks of data are performed by repository staff, members of the General and Computational Linguistics Department, or other local experts in the field. External data is only accepted in the repository if the project seeking deposition of the data has been externally reviewed, for example in a grant application process. In the case where a dataset cannot be reviewed locally, the repository may recommend another CLARIN center with a collection profile which is better matched to the dataset (https://www.clarin-d.net/en/disciplines). Currently, many of the datasets in the repository are widely-used language resources created locally by the Seminar für Sprachwissenschaft (e.g. GermaNet and TüBa-D/Z), or externally created datasets that are well known within the community (e.g. Index Thomisticus Treebank).

Depositors are encouraged to use formats listed in the CLARIN standard recommendations (http://www.clarin.eu/content/standards-and-formats/) when possible. Use of these formats will ensure that the data is interoperable within the CLARIN infrastructure. If possible, data stored in other formats will be converted to an acceptable format before it is archived.

The depositor, with assistance from a data manager if necessary, creates CMDI (http://www.clarin.eu/cmdi/) metadata using components and profiles stored in the Component Registry (http://catalog.clarin.eu/ds/ComponentRegistry/). Metadata is manually checked for completeness and correctness, and automatically validated for syntax and consistency, prior to ingestion. Metadata is required, since data without metadata is technically not accepted in the system.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

Please provide a public appraisal and selection document when renewing your certification.

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:

The repository's preservation policy (https://uni-tuebingen.de/en/137029) includes local and distributed backups, reinstalling the repository from backup, and integrity tests of stored data. The processes and procedures of the repository (including ingest, metadata checks, recovery from backup, etc.) are documented in an internal wiki. Locally developed software is stored and documented in an internal git repository. Both the wiki and the git repository are backed up regularly.

The Information, Communication, and Media Center (IKM) of the University of Tübingen is the central information center of the university. It is formed through cooperation between the university library (UB) and the university computing center (ZDV), and reports directly to the rectorate of the University of Tübingen. The computing center of the university provides all central IT-services, including data storage. Storage service is provided in cooperation with the Universities of Stuttgart and Hohenheim under the umbrella of a state internal cooperation plan. The repository makes use of this central infrastructure for backup and operating services.

The repository is physically located at the Central Data Center (ZDV) at the University of Tübingen. They provide the server and its maintenance (including system updates, firewalls, etc), storage, and remote backup services. The repository server is configured to perform daily backups to the University of Ulm data center, appoximately 90 km away. Physical access to the servers is allowed only by authorized ZDV staff and is strictly controlled.

Additionally, the repository uses the B2SAFE (http://www.eudat.eu/b2safe) service provided by the European Data Infrastructure project EUDAT (http://eudat.eu/) to replicate the data in a data center located in Garching, Germany, which is approximately 250 km away. B2SAFE currently uses iRODS, a data management middleware, to enable implementation of data management policies. A B2SAFE backup is performed automatically on a weekly basis.

In order to maintain the integrity of archived data, checksums based on the MD5 algorithm are calculated and the stored objects are assessed regularly. In addition, checksums are automatically computed each time a data stream is downloaded. Deviations are visible to the archive managers for taking immediate action.

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 repository's preservation policy (https://uni-tuebingen.de/en/137029) includes local and distributed backups, reinstalling the repository from backup, and integrity tests of stored data. The repository and backups are located in dedicated computing centers with strict access control, and administrator access to the repository is limited to a small group of trained experts.

Depositors are encouraged to use formats listed in the CLARIN standard recommendations (http://www.clarin.eu/content/standards-and-formats/) when possible. The list of accepted data formats may be extended to include new, widely-used formats in the field. In this case, the repository staff will determine which datasets it would be feasible and possible to convert. For example, the repository has converted many of its treebank resources to a new format that has gained popularity in recent years. In the case that a data format is removed from the list of acceptable formats in the future, every effort will be made to convert datasets into an acceptable format.

The Information, Communication, and Media Center (IKM) of the University of Tübingen is the central information center of the university. It is formed through cooperation between the university library (UB) and the university computing center (ZDV), and reports directly to the rectorate of the University of Tübingen. The computing center of the university provides all central IT-services, including data storage. Storage service is provided in cooperation with the Universities of Stuttgart and Hohenheim under the umbrella of a statewide concept for data. The repository makes use of this central infrastructure for backup and operating services.

The repository backend was selected for ease of long-term maintenance and compliance to best practice. It has low technical requirements for extracting the resources from the system without additional and proprietary software, making the transfer of the data to new hardware straight-forward. Long-term access is ensured by the hardware, open protocols, and organizational embedding in sustainable departmental structures of the university.

Neither the CLARIN-D resource center nor the repository run by it, are legal entities on their own. This also holds for the General and Computational Linguistics Department ("Seminar für Sprachwissenschaft", SfS) where they are located. All are part of the University of Tübingen which is a legal entity - specifically, like all public German universities, a

Körperschaft des öffentlichen Rechts, an institution governed under public law.

Depositors must sign a Depositor's agreement (https://uni-tuebingen.de/en/134320) with the University of Tübingen, which ensures that they own all necessary rights required to deposit the data, that they are in compliance with all relevant national and international legal regulations, and that they grant the repository permission to distribute the data in accordance with the access model chosen (public, academic, or individual). Data providers retain all intellectual property rights to their data. In case a violation of conditions is observed, steps will be taken to ensure that the data is not distributed until the issue can be resolved.

The repository is currently in the process of developing its future strategy, allowing it to guarantee preservation periods to data depositors. Discussions with the BMBF, the state of Baden-Württemberg, and the University of Tübingen are ongoing.

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 General and Computational Linguistics Department of the University of Tübingen develops and maintains language resources and tools for community use (https://uni-tuebingen.de/en/134257). Many of these widely used resources have been made available via the Tübingen CLARIN-D repository. These resources, as well as the external resources currently hosted, each have a dedicated webpage containing a detailed description of the resource and contact information for users to ask questions, request a license if required, or give feedback about the resource. Links to the webpage are recorded in the CMDI metadata so that the webpage can also be found using CLARIN search applications.

As part of CLARIN, the repository also participates in additional channels through which members of the designated communities can give feedback on data and metadata hosted by certified centers. These channes include the metadata portal CLARIN Virtual Language Observatory (https://vlo.clarin.eu/), the CLARIN-D Help Desk, and discipline-specific working groups.

The CLARIN Virtual Language Observatory (VLO, https://vlo.clarin.eu/) harvests ISO 24622-1 conformant metadata (CMDI) and displays the large number of available resources through faceted browsing and search facilities. Both in the overview, i.e. when browsing or searching for relevant resources, and on the individual resource pages displaying further information on a specific resource, the user can report an issue or give feedback on metadata records or resources using a designated button connected via a form to the CLARIN-D Help Desk.

The CLARIN-D Help Desk, maintained by the CLARIN centre at the University of Hamburg, manages support and feedback workflows for national centres and various international services, such as the CLARIN VLO. Depending on the type of feedback, help desk agents can thus both forward issues directly to the responsible CLARIN centre and, for issues with a wider impact, contact relevant institutions and bodies at the European level, such as the CLARIN Metadata Curation Taskforce, which is responsible for improving and harmonizing metadata within the infrastructure.

Furthermore, the discipline-specific working groups (https://www.clarin-d.net/en/disciplines) within the CLARIN-D project are yet another communication channel, through which the various designated communities can provide more general input and feedback on data and metadata to ensure CLARIN-D centres provide relevant resources and resource descriptions.

The metadata profiles used by the CLARIN-D centre in Tübingen have been selected for descriptive appropriateness for the data types deposited in the repository. ISO 24622-1 provides the framework for selecting these metadata profiles.

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:

The repository implements explicitly defined workflows described in the repository's preservation policy (https://uni-tuebingen.de/en/137029). Workflows for both depositing data and accessing data are summarized on the repository website (https://uni-tuebingen.de/en/134314). The depositing workflow consists of packaging the resource, creating metadata, and a quality check of data and metadata including PID (Persistent Identifier) assignment. The access workflow includes options for restricting data access. Where possible, processes in the workflow are automated and enforced by the Fedora Commons back-end or through a custom front-end to Fedora Commons developed by the Tübingen CLARIN-D repository staff.

Detailed workflow documentation is maintained in an internal wiki. A general description was also published by Dima, E. et al. (2012): A Repository for the Sustainable Management of Research Data (http://www.lrec-conf.org/proceedings/lrec2012/pdf/470_Paper.pdf).

The repository and its backups are located in dedicated computing centers with strict access control, and administrator access to the repository is limited to a small group of trained experts. This ensures that the data storage and backup is always managed by professionals.

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:

All CLARIN centres (https://www.clarin.eu/content/overview-clarin-centres) provide their metadata in the CMDI format. The Component MetaData Infrastructure (CMDI) (https://www.clarin.eu/content/component-metadata was initiated by CLARIN to provide a flexible framework for describing metadata based on components and concepts. Each metadata record is based on a profile that is registered in the CLARIN CMDI Component Registry (https://catalog.clarin.eu/ds/ComponentRegistry). Profiles can make use of components. Those building blocks are also registered in the CMDI Component Registry and describe specific aspects or properties of a resource. Elements of CMDI records link to concept definitions that are stored in external registries (like the CLARIN Concept Registry, https://openskos.meertens.knaw.nl/ccr/browser/). Since different communities use different names for the same concepts, linking CMDI elements to concepts enables communities to stick to their terminology while enabling users to find concepts independent of the naming.

A strict requirement for CLARIN centres, and therefore for the Resource Centre Leipzig as well, is to make metadata for all resources available through the established and well documented Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (https://www.openarchives.org/pmh/). This standard enables harvesting of the metadata from the repository via http(s).

Main search facilities are currently not provided by the repository itself. Instead, services of the CLARIN-ERIC are utilized. The provision of harvesting services for metadata and the provision of central metadata and

data search facilities is stated in the value proposition:
https://www.clarin.eu/value-proposition

Metadata:
The CLARIN Virtual Language Observatory (VLO) (https://vlo.clarin.eu) of the CLARIN-ERIC harvests the metadata in CMDI format from all CLARIN centres via OAI-PMH. Metadata from CLARIN centres (and other relevant archives and repositories) are browsable and searchable via the VLO website. CLARIN has defined a set of facets to narrow down the selection of resources in the VLO. These facets are again based on concept sets and allow access to potential heterogeneous metadata stocks. The search in the VLO combines a full text query with a selection of (multiple) values in facets.

Data:
For a subset of resources of the CLARIN-infrastructure a “deep search” within the actual data is supported by the means of the CLARIN Federated Content Search (http://www.clarin.eu/fcs) interface. The Tübingen CLARIN-D Repository also offers this kind of access for some of its resources.

PIDs:
The repository uses the common CLARIN PID service (https://www.clarin.eu/files/pid-CLARIN-ShortGuide.pdf) based on the Handle System (http://www.handle.net/) and in cooperation with the European Persistent Identifier Consortium (EPIC). The usage of PIDs is mandatory for resources in CLARIN thus all resources added to the repository may be referenced using PIDs.

CLARIN has a contractual relationship with GWDG concerning the provision of PID-services via EPIC API v2 as mentioned in R0 on repository context.

Citation:
Necessary information for citing resources of the repository - such as PID, resource name, and responsible organization - can be found on the respective page of the VLO, as well as in the metadata. Discussions with CLARIN-ERIC have been initiated on integrating recommended data citations into the VLO.

The Tübingen CLARIN-D repository provides preliminary recommendations on how to cite data in the data's metadata. This includes the landing page of the resource, the name of the resource, its creator, release date, and an actionable persistent identifier (in the form of a URL). The landing pages for the General and Computational Linguistics Department resources are also being reviewed and updated to include standard citation recommendations where they are not already listed.

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:

All CLARIN centres (https://www.clarin.eu/content/overview-clarin-centres) provide their metadata according to ISO 24622-1 (CMDI) via OAI-PMH. The Component MetaData Infrastructure (CMDI; https://www.clarin.eu/content/component-metadata) was initiated by CLARIN to provide a flexible framework for describing metadata. With this metadata framework it is possible to create metadata schemas tailored to each specific type of resource. It also allows inclusion of all pieces of information deemed useful for potential future users to understand the data in a reuse scenario. This also includes basic information utilized by research data search engines. To avoid proliferation and to allow for transparent structures this is technically realized by bundling descriptive categories into components that can be reused for other data types. Components themselves are then bundled into datatype-specific schemas called profiles. Each metadata record is based on a profile that is registered in the Component Registry (https://catalog.clarin.eu/ds/ComponentRegistry). The data category of CMDI records link to concept definitions that are stored in external registries (like the CLARIN Concept Registry (https://openskos.meertens.knaw.nl/ccr/browser/)). Since different communities use different names for the same concepts, linking CMDI data categories to concepts enables communities to retain their terminology while enabling users to find concepts independent of the naming.

The Tübingen CLARIN-D repository uses the following CMDI profiles, which were designed with concept registry links:

First generation profiles used:

ExperimentProfile: clarin.eu:cr1:p_1302702320451
LexicalResourceProfile: clarin.eu:cr1:p_1290431694579
TextCorpusProfile: clarin.eu:cr1:p_1290431694580
ToolProfile: clarin.eu:cr1:p_1290431694581
WebLichtWebService: clarin.eu:cr1:p_1320657629644
Resource Bundle: clarin.eu:cr1:p_1320657629649
OLAC-DcmiTerms: clarin.eu:cr1:p_1288172614026
DcmiTerms: clarin.eu:cr1:p_1288172614023

Second Generation Profiles:

ExperimentProfile:clarin.eu:cr1:p_1447674760337
TextCorpusProfile: clarin.eu:cr1:p_1442920133046
LexicalResourceProfile: clarin.eu:cr1:p_1445542587893
SpeechCorpusProfile: clarin.eu:cr1:p_1485173990943
CourseProfile: clarin.eu:cr1:p_1505397653792
ToolProfile: clarin.eu:cr1:p_1447674760338

The second generation profiles extend the first generation profiles, for example by allowing metadata files to include authoritative IDs for individuals and institutions, such as VIAF links and ORCID IDs. The repository is currently updating metadata files from first generation profiles to second generations profiles.

Prior to ingest, the depositor and the data manager check that all relevant metadata fields have been filled in correctly, and as completely as possible. Special attention is given to these components:

- General Information, including name of the resource, type of resource, version of the data, life cycle status,legal owner, start year, field of research, modality
- Project, including relevant information about the project in which the resource was created, if it was created within a specific research project
- Publications, which describe the resource or are based on the resource
- Creation, including the information on each individual involved in the resource creation, software tools used in the creation of the dataset, and third party data contained in the dataset
- Documentation of the resource, i.e. external descriptions of the resource
- Access Information, i.e. under which licence a data user may receive the resource, associated software that can be used to work with the data
- Technical Information for each file part of the data set
- Resource Type Specific Information, such as the size of a text collection in terms of number of words or the number of

recording hours for speech corpora

The components for General Information, Project, Creation, Access and Technical Information are required in most profiles, and the data manager ensures that those are completed. Data Type Specific Information is strongly encouraged by the data manager. Since Documentation, Project, and Publications may not be available for each resource, they are optional.

Depositors are encouraged to use formats listed in the CLARIN standard recommendations (http://www.clarin.eu/content/standards-and-formats). Use of these formats will ensure that the data is interoperable within the CLARIN infrastructure. If possible, data stored in other formats will be converted to an acceptable format before archiving. In the case that a particular format is replaced by a more widely-used format, data will be converted and archived under a new PID. The landing pages for resources developed at the General and Computational Linguistics Department of the University of Tübingen contain a detailed description of the resource, including which data format(s) are used and any associated software that can be used with the data (e.g. TüBa-D/Z treebank: https://uni-tuebingen.de/en/134290; GermaNet wordnet: http://www.sfs.uni-tuebingen.de/GermaNet/index.shtml). Resources have been, and will continue to be, converted to new formats as the need arises.

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:

As part of CLARIN-D we are committed to play an active role in the development of CLARIN's repository infrastructure. General plans for maintaining and further developing the infrastructure have been formulated as part of the project proposal.

The central goal is to improve the usability of the research infrastructure for typical research tasks such as the retrieval of resources, the evaluation of data or the publication of results. To achieve this, modifications and extensions to a variety of infrastructure components in the repository and in the central infrastructure are necessary. Meetings of all centres to monitor advances in infrastructure development take place quarterly.

Further important goals of infrastructure development (https://www.clarin.eu/content/clarin-technology-introduction) are:
- To ensure resilience, integrity, and availability of the sustainable repositories and the central infrastructure
- To integrate new resources and tools based on the needs of the user communities
- To allow for better interoperability of tools and resources in the infrastructure
- To enhance the central content search to be more useful in actual research tasks
- To optimize metadata of the resources provided and to enhance user experience in central metadata search

Additional strategic infrastructure planning takes place on the European level in the coordinating committee of the technical centres of the CLARIN ERIC where CLARIN-D also participates.

The repository adheres to all standards and best practice recommendations set forth by CLARIN, as well as meeting the requirements of OAIS, as described in the preservation policy (https://uni-tuebingen.de/en/137029).

The repository is built on a reliable and stable technical infrastructure, which has been tested and evaluated and has been determined to be fully functional for the needs of the repository. The main technical components used by the repository include:

- The core repository infrastructural software is Fedora Commons 3 (currently upgrading to 4), running on a server at the central data center (ZDV) of the University of Tübingen. Fedora Commons software is compliant with the OAIS reference model, and its use promotes adherence to OAIS in all functions of the repository.

- Server maintanence and geographically distributed daily backups (https://uni-tuebingen.de/en/2944) are performed by

the ZDV.

- An additional backup mechanism is also in place, making weekly backups to a different remote location. These backups are performed with the B2SAFE (http://www.eudat.eu/b2safe) service, provided by the European Data Infrastructure project EUDAT (http://eudat.eu/). B2SAFE is built upon iRODS (https://irods.org/) data management middleware.

- The repository has developed and maintains a customized administrator interface to the Fedora Commons backend that aids in carrying out the workflows and functions described in the preservation policy.

- All locally developed software is housed in a local git repository, and all workflow documentation in an internal wiki. Both the git repository and the wiki are backed up regularly by the local system administrator.

- Firewalls block unauthorized access to the systems on which the repository is operated, including access to administrative tools and backends from unauthorized workstations.

- Icinga (https://www.icinga.com/) is used by CLARIN-D (http://clarin-d.net/images/ap3/ap3-005-monitoring.pdf) to monitor infrastructure components, including repository probes. Repository probes are made every few minutes and all repository technical staff are notified by email if a problem occurs so that it can be resolved quickly.

- Metadata is created according to ISO 24622-1 and ISO/DIS 24622-2 standards, using either the web based editor Comedi (http://clarino.uib.no/comedi/page?page-id=repository-main-page) or standard XML editors such as Oxygen. A set of XQuery functions are used to test and generate reports on the quality of the CMDI metadata. These tests include consistency checks (use of PIDs, naming of persons and institutions, field names, etc), file record reference update notifications, completeness, etc.

- Metadata is disseminated using the OAI-PMH protocol, with the PROAI plugin of Fedora Commons.

- PIDs are aquired from, and resolved by, the Gesellschaft für Wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), with whom CLARIN has a contractual relationship.

- The repository bandwidth is 10 MB/sec.

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 repository's preservation policy (https://uni-tuebingen.de/en/137029) includes local and distributed backups, reinstalling the repository from backup and integrity tests of stored data. The repository and backups are located in dedicated computing centers with strict access control, and administrator access to the repository is limited to a small group of trained experts.

The Information, Communication, and Media Center (IKM) of the University of Tübingen is the central information center of the university. It is formed through cooperation between the university library (UB) and the university computing center (ZDV), and reports directly to the rectorate of the University of Tübingen. The computing center of the university provides all central IT-services, including data storage. Storage service is provided in cooperation with the Universities of Stuttgart and Hohenheim under the umbrella of a statewide concept for data. The repository makes use of this central infrastructure for backup and operating services.

The repository runs on a server housed, managed and maintained by the ZDV, who is also responsible for making daily backups of the data and system configurations to a remote location. The repository server is currently configured to perform daily backups to the University of Ulm data center using Bacula (https://uni-tuebingen.de/en/2944), and a detailed report is sent to the repository staff for each backup.

An additional backup mechanism is also in place, making weekly backups to a different remote location. These backups are performed with the B2SAFE (http://www.eudat.eu/b2safe) service, provided by the European Data Infrastructure project EUDAT (http://eudat.eu/). B2SAFE is built upon iRODS (https://irods.org/) data management middleware.

In case of disaster, recovery will first be attempted through the ZDV backups, and then through the documented recovery

procedures of the alternative backup strategy.

The repository status and availability of resources are continually monitored within the CLARIN infrastructure. In case of any failure, the repository staff is notified immediately.

In order to maintain the integrity of archived data, checksums based on the MD5 algorithm are being calculated and the stored objects are assessed regularly. In addition, checksums are automatically computed each time a data stream is downloaded. Deviations are visible to the archive managers for taking immediate action.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

For future renewals, please provide links to evidence relating to information security and physical security, as well as data integrity and resilience.

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:

Thank you for reviewing our application.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: