



Assessment Information

[CoreTrustSeal Requirements 2017–2019](#)

Repository: CLARIND-UDS
Website: <http://fedora.clarin-d.uni-saarland.de/>
Certification Date: 15 February 2019

This repository is owned by: Universität des Saarlandes



CLARIND-UDS

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, Institutional repository, Research project 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 CLARIND-UDS centre (<http://fedora.clarin-d.uni-saarland.de>) is part of CLARIN-D (Common Language Resources and Technology Infrastructure Deutschland) - a web and centres-based research infrastructure for the social sciences and humanities. The aim of CLARIN-D and its service centres is to provide linguistic data, tools and services in an integrated, interoperable and scalable infrastructure for the social sciences and humanities. 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. CLARIN-D is funded by the German Federal Ministry for Education and Research.

In collaboration with other centres in the CLARIN-D consortium, the UdS CLARIN-D centre enables eHumanities by providing a service for hosting and processing language resources (notably corpora) for members of the research community.

The UdS CLARIN-D centre thus contributes to the overall CLARIN mission of lifting the fragmentation of language resources by assisting members of the research community in preparing language materials in such a way that easy discovery is ensured, interchange is facilitated and preservation is enabled by enriching such materials with meta-information, transforming them into sustainable formats and hosting them.

The designated community of the CLARIND-UDS repository consists of linguists (specially corpus linguists), philologists, and researches in digital humanities.

The UDS CLARIN-D centre hosts corpora and tools, specially multilingual corpora (parallel and/or comparable) and corpora including specific registers.

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 CLARIND-UDS centre is one of currently eight German CLARIN-D Resource and Service Centres 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 CLARIND-UDS centre 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, C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

Comments

We do a regular check of the provided data and metadata before accepting and ingesting a resource to our repository. We routinely enrich the metadata in cooperation with the data provider to enhance visibility and findability in the Virtual Language Observatory (VLO, <https://vlo.clarin.eu>)

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). CLARIN-D has a contractual relationship with GWDG concerning the provision of PID-services via EPIC API v2. This document lists the services which were stipulated: http://www.clarin-d.de/mwiki/images/0/0b/GWDG_PID.pdf.

This outsource partner offers relevant functionality for Requirement 13 “The repository enables users to discover the data and refer to them in a persistent way through proper citation.”

2) Hochschul-IT-Zentrum (hiz-saarland)

The repository makes use of the server virtualisation (<https://www.hiz-saarland.de/dienste/virtualisierung/> and <https://www.hiz-saarland.de/dienste/webservices/virtualisierungsarten/>) and of the backup facilities (<https://www.hiz-saarland.de/dienste/datensicherung/>) offered by the HIZ. The HIZ is the joint IT provider of Saarland University (The archive's hosting institution) and of the University of Applied Sciences of the Saarland (HTW Saar). This outsource partner offers relevant functionality for Requirement 9 “The repository applies documented processes and

procedures in managing archival storage of the data.”

English summaries of the German language web pages

<https://www.hiz-saarland.de/dienste/virtualisierung/>

The Hochschul-IT-Zentrum offers departments of the University highly available Virtual Servers. There are two types of Virtual Servers (see next page). Several Operating systems (Linux and Windows) are supported, daily backup is also offered.

<https://www.hiz-saarland.de/dienste/webservices/virtualisierungsarten/>

The Hochschul-IT-Zentrum (HIZ) offers two types of virtualisation: Virtual Web Servers (vWeb) and Virtual Servers (vServer).

[Skipped part on vWeb servers, because our repository does not use this offer].

On VServer: The virtualisation service of the HIZ is based on the VMWare virtualisation technology. Resources can be individually assigned to the virtual servers and flexibly extended, if necessary. The whole virtualisation environment is run by the HIZ, but the system administration of the servers is done by the repository owner. The HIZ also offers backup facilities.

[Skipped: Comparison of the two types of servers]

<https://www.hiz-saarland.de/dienste/datensicherung/>

The Hochschul-IT-Zentrum (HIZ) offers a central backup service for departments of the university based on the EMC Networker. Features of the service include

- Backup and restauration of data
- Daily backup (incrementell and full) on hard disks
- Tape backup of the data
- Storing time of the back-up for at least three months
- Status notification of the the backup
- The user is responsible for the configuration of the backup client
- The user is responsible for the restauration of data
- No archival is offered

3) CLARIN-D

The repository is 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 its 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.

4) CLARIN-ERIC

CLARIN-D is a member of CLARIN'S European Research Infrastructure Consortium (ERIC). CLARIN-ERIC offers central services to its 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 Observatory - 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.

re3data record: <https://www.re3data.org/repository/r3d100010384>

DFG RI sources: http://risources.dfg.de/detail/RI_00435_de.html

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:

We have an explicit mission [1] to archive language resources especially multilingual corpora (parallel, comparable) and corpora including specific registers, both collected by associated researchers as well as researchers who are not affiliated with us.

The mission is supported by the infrastructure of Saarland University and is integrated into the national and international CLARIN infrastructures. As part of the CLARIN infrastructure, the repository is included in all promotional activities carried

out at the

national level of CLARIN-D as well as the European level of CLARIN [2,3].

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.

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

1. <http://fedora.clarin-d.uni-saarland.de/index.en.html>

2. <https://www.clarin.eu/content/vision-and-strategy>

3. https://office.clarin.eu/v/CE-2017-1093-Value_proposition.pdf

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:

All CMDI metadata are provided without access restrictions according to CLARIN-D recommendations.

The actual data have individual licenses and access restrictions. While the repository prefers data to be Open Access under a Creative Commons license, this cannot be granted for all archived resources. In one case it was necessary to arrange an in-house stay to comply with the license restrictions of a certain resource.

If applicable, the data consumer is made aware of usage restrictions for the data she/he has gotten access to. Generally the general usage restrictions are already described in the codes of conduct (see Requirement R4 for details), specific restrictions are specified by the depositor (if applicable). For some data, explicit statements need to be made by the data consumer about the usage of the data before he/she gets access.

In case of misuse, the only thing that can be practically done is to deny the user further access to the repository and to make the research community aware of the misuse.

Terms of Use: <http://fedora.clarin-d.uni-saarland.de/termsfuse.en.html>

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:

Currently CLARIN-D is funded by the Bundesministerium für Bildung und Forschung (BMBF). The current project phase has a duration of 4 years and is funded until 30.09.2020. As an alternative to project based funding, CLARIN-D currently pursues a permanent continuation of funding. In the case that the funding is completely ceased, Universität des Saarlandes will provide access to the archived resources for a grace period.

In this time, a transfer of the resources to another archive can be negotiated. CLARIN centres commit to ensuring long-term availability, access and to preservation of datasets submitted to their repositories, as set out in their mission statements. CLARIN centres are setup as a distributed network, where each centre 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 funding, the repositories content would be transferred to another CLARIN centre.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

For future certifications, the reviewers hope to see evidence for more tangible plans and details about continuity of funding, in particular formal agreements between the repository and CLARIN-D / Universität des Saarlandes for the continuity of access.

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:

Deposits are handled in a case-by-case approach. There are individual contracts and different licences for each resource we have archived. The access to the items is also handled case-by-case, ranging from open access over restricted access requiring a contract to restricted access onsite. The depositors themselves are responsible for compliance with any legal regulations in the area where the data is collected. In particular, all depositors are required to follow the The European Code of Conduct for Research Integrity [1].

Before ingest, the data are spot checked for proper anonymisation when disclosure risks are imminent. A description of our workflow on acceptance and curation is found in Requirement R12.

The repository will include resources provided by CLARIN-D related institutions and other institutions and/or organizations that belong to the CLARIN-D extended community. The data in our repository contains sufficient information for others to assess the scientific and scholarly quality of the research data in compliance with disciplinary and ethical norms. We specifically rely on DFG ethical Codes of Conduct [2]. Our repository provides a quality assessment in that the data consumer can make some judgement about the level of trust or about the reputation of the depositor on the basis of the meta-information about the source institution/organization that is related to each resource. Our repository does not (and cannot) systematically verify whether the data received are collected according to these quality standards. We provide some guidance to depositors in terms of describing the full package of information that should be deposited to facilitate assessment at the repository's webpage [3,4]. The host of the repository, Universität des Saarlandes, also has a guideline on scientific misconduct [5, in German] that applies to our repository.

1. ALLEA (ALL European Academies) European Science Foundation, The European Code of Conduct for Research Integrity, Revised Edition

<http://www.allea.org/wp-content/uploads/2017/04/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf>

2. DFG, Rules of Good Scientific Practice

http://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html

3. <http://fedora.clarin-d.uni-saarland.de/termsfuse.en.html>

4. <http://fedora.clarin-d.uni-saarland.de/depositors.en.html>

5. <http://www.uni-saarland.de/campus/forschung/weitere-hinweise/wissenschaftliches-fehlverhalten.html>

Summary of German webpage:

<http://www.uni-saarland.de/campus/forschung/weitere-hinweise/wissenschaftliches-fehlverhalten.html>

Scientific misconduct (ombudsperson)

Universität des Saarlandes has issued guidelines and a code of conduct to secure good scientific practice and to avoid scientific misconduct.

This includes setting up an ombudsperson to advise member of the university on good scientific practice and to confidentially accept any allegations of scientific misconduct.

Examples of scientific misconduct (fake results, theft of intellectual property, impeding the research of other researchers, and destroying primary data violating legal restrictions) and of good scientific practice (promoting young researchers, adequate assessment of scientific achievements, respect of authorship) are given.

In the case of a reasonable suspicion of misconduct, the ombudsperson can be contacted confidentially and an investigation will be undertaken.

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:

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:

The CLARIND-UDS repository is part of CLARIN-D, a research infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences. CLARIN-D also offers information on a wide range of topics, including teaching material, help on data management plans and other, discipline-specific support. In particular, our 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 centres, which come from within their respective fields of digital humanities. By being part of the CLARIN-D consortium the repository gains access to funding for running and further developing a sustainable repository and resource centre to support these 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.

The repository staff consists of scientists with solid knowledge and experience in the fields of corpus linguistics and metadata management. The staff dedicated to the repository consists of the project leader and one FTE position for technology development, repository administration, and metadata curation. The staff members have access to training on data management, metadata curation, and long-term preservation offered by CLARIN-D and CLARIN-ERIC. There are regularly scheduled developer meeting within CLARIN-D where the repository staff participates.

<https://www.clarin-d.net/en/about/centres/division-of-labour-of-clarin-d-centres>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

For future certifications, the reviewers would like to see more details on the funding available via CLARIN-D and the host institution (Universität des Saarlandes).

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 CLARIND-UDS repository is run in the Language Science and Technology department of Saarland University by members of the designated user community. The repository team is present on relevant conferences and also organises own workshops (e.g., http://www.sfb1102.uni-saarland.de/?tribe_events=workshop-historical-corpus-linguistics-methods-and-applications)

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 Academy 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 CLARIND-UDS repository does not allow for anonymous deposition of resources and it also does not provide any way for automated deposition by trusted users. Before a resource is accepted, the repository administration will be in contact with the depositor to assess the data, the metadata, and the legal issues. The repository administration offers guidance on how to produce good metadata and checks the formal correctness of them.

We consider all objects deposited in our repository as fixed and immutable. We create new digital objects for updates and keep the old versions in our repository.

We monitor the integrity of all archived files using two kinds of checksums (MD5 and SHA1). The checksums are controlled on a daily basis, and in case of a checksum mismatch a backup copy of the resource will be restored.

The new version of a resource will contain a pointer to the older versions in its metadata. The repository in principle makes the original deposited objects available in an unmodified way, if the objects were in one of the accepted file types and encodings. In the case of changes by the data producer, the repository creates a new digital object with a new persistent identifier. In the case that the repository has to change the data, e.g., because a file format becomes obsolete and superseded, the original data are kept. The repository only accepts works from the original data producers, who are acknowledged as such by means of the "dc:creator" or "creator" elements, in Dublin Core or CMDI metadata respectively.

We use the Dublin Core field "relation" in the metadata to maintain relations to other datasets, tools, or publications. The relations given there reflect the time when the resource was prepared and submitted and are contributed by the data producers.

We know the authors of our contributions from the scientific community and we are in contact with them during the ingest process. We do not formally check their identity.

Examples:

CMDI metadata record for the GRUG parallel treebank as delivered by the OAI provider http://fedora.clarin-d.uni-saarland.de/oaiprovider/?verb=GetRecord&_metadataPrefix=cmdi&_identifier=oai:fedora.clarin-d.uni-saarland.de:clarind-uds:grug

An example of a versioned resource is the Royal Society Corpus that is currently archived in two versions, Version 2.0 (<http://hdl.handle.net/11858/00-246C-0000-0023-8D26-7>) and Version 4.0 (<http://hdl.handle.net/21.11119/0000-0001-7E8B-6>)

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:

The data are provided in the formats chosen by the data producers from a list of supported formats [1]. Metadata for each resource are always provided in both Dublin Core [2] and CMDI [3] formats. The CLARIND-UDS repository requests good Dublin Core metadata from the data provider and takes care for the conversion into CMDI.

The repository staff checks the metadata provided by the data depositor and asks for additional information when the metadata are too sparse. The repository staff is also aware of the mapping of metadata to the facets of the CLARIN Virtual Language Observatory (VLO) and suggests optimisations for better search results.

The repository did not experience any problems with non-preferred formats yet. We were always able to find a consensus on an accepted format with our data providers.

References

1. <http://fedora.clarin-d.uni-saarland.de/ressources/AcceptedFormats.en.pdf>
2. <http://www.dublincore.org>
3. <https://www.clarin.eu/content/component-metadata>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
For future certifications, a collection development policy (appraisal/selection document online) should be documented (or the relevant one pointed to as part of CLARIN-D). More documented procedures should be made publicly available on the UDS repository.

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 deposited data are stored on a running virtual machine. They are secured against

- * physical loss
- * change of the data
- * unauthorized access

Metadata in CMDI format is created and optimised in close collaboration with the depositor to ensure an adequate description of the data; for more information see our answer on requirement R14.

Protection against physical loss

The virtual machine run on a distributed hardware environment in two separate buildings. The storage is realised with RAID disks 1,2,in German]. In addition, we use the backup service by the HIZ [3, in German] for daily backups on EMC Networker. In case of data loss the repository management restores the original data.

Protection against change of data

The repository monitors the integrity of all archived files using two kinds of checksums (MD5 and SHA1). The checksums are controlled on a daily basis, and in case of a checksum mismatch a backup copy [3, in German] of the resource will be restored.

Protection against unauthorized access

The virtual machine of the repository is run behind a firewall. From the outside, only the standard http and https ports are accessible, and they only provide access to resources under free licences. A remote login to the machine is only possible from a trusted internal subnet, and it can be only done via Secure Shell [4].

1. <https://www.hiz-saarland.de/dienste/virtualisierung/>
2. <https://www.hiz-saarland.de/dienste/webservices/virtualisierungsarten/>
3. <https://www.hiz-saarland.de/dienste/datensicherung/>
4. <https://www.ssh.com/ssh/>

For English summaries of the cited webpages [1-3], see Requirement R0.

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:

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:

The CLARIND-UDS repository performs data curation on levels B (basic curation) and C (enhanced curation). Basic curation is applied before the ingest of every resource, enhanced curation is done on a case-by-case basis when a data format is becoming obsolete. The depositors grant the right to convert dataformats to the repository [1].

Measures are taken to enhance the chance of future interpretability of the data. The number of accepted file formats is limited, to make future conversions to other formats more feasible. As much as possible open (non-proprietary) file formats are used. For textual resources, XML formats are used whenever possible, to make future interpretation of the files possible even if the tool that was used to create them no longer exists. Text is encoded in Unicode to ensure future interpretability. Before ingest, we do the following checks:

- * Full validation of the metadata against their respective schemes (oai_dc, CMDI)
- * Check that provided XML data are well-formed (at least)
- * Spot-check of character encoding

All archived resources are available online, the access permissions are defined by the data producers/depositors themselves. The crisis management plan relies on the technical solution described in Requirement R9 of these guidelines.

[1] <http://fedora.clarin-d.uni-saarland.de/depositors.en.html>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

By the time of the next review, CLARIND-UDS should provide a link to a detailed preservation plan covering in more details the questions provided in the guidance (in particular evidence for managed documentation).

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 CLARIND-UDS repository is integrated into the Common Language Resources and Technology Infrastructure (CLARIN) which implements several channels through which members of the designated communities can give feedback on data and metadata hosted by its certified centres.

Researchers interested in including their resources into the CLARIND-UDS repository are invited to develop a data management plan [1] in coordination with the staff of the CLARIN centre, which is offered as a service already in early states of their projects. Otherwise, actual data currently is only accepted when data depositors can be trusted (see R8: Appraisal). Metadata may be submitted by the data depositor or will be created in coordination with the CLARIND-UDS centre - in any case, it will be thoroughly checked and discussed by experienced centre staff.

The metadata portal CLARIN Virtual Language Observatory [2] harvests the CMDI metadata of all CLARIN centres and

displays the large amount 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 [3], maintained by the CLARIN-D 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 harmonising metadata within the infrastructure.

Furthermore, the so-called discipline-specific working groups within the CLARIN-D project [4] 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.

Our metadata profiles and the quality assurance of our metadata is described in our answer to requirement R14.

1. <https://www.clarin-d.net/de/aufbereiten/datenmanagementplan-entwickeln>
2. <https://vlo.clarin.eu>
3. <https://www.clarin-d.net/en/help>
4. <https://www.clarin-d.net/en/disciplines>

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 CLARIND-UDS repository complies with the OAIS reference model's tasks and functions [1]. Besides, the repository is powered by Fedora Commons software, which is compliant with the Reference Model for an Open Archival Information System (OAIS) due to its ability to ingest and disseminate Submission Information Packages (SIPS) and Dissemination Information Packages (DIPS) in standard container formats. We provide a structure diagram of the repository [2] The ingestion workflow is defined by the archive management tool Fedora Commons, such as that no resource can be archived without metadata and that the resource has to conform to certain file formats and encodings. The responsibilities of the depositor are

- * to decide what kind of material is being archived,
- * to assure that the material follows the technical criteria required by the repository,
- * to decide who may access the material,
- * to protect the privacy of any subjects appearing in the recordings or texts.

There is an internal documentation on the preparation of resources and the corresponding metadata. For the time being there is no need to make these publically available as we do not intend to implement an automatic ingestion process. The data depositors grant the repository the licence to convert the submitted data to other formats for the purpose of data curation [3]. The data consumer has direct access to the archived objects via the web, provided that access requirements have been met. In the case of protected resources, special arrangements can be made on individual basis, including an in-house stay to access a resource that cannot be given out due to copyright restrictions. For details on the search interfaces for dissemination, see our answer to Requirement R13.

Verbal description of the Ingest workflow

For a suggested resource, the archive management decides whether the repository accepts it using the criteria given in the checklist [3] on this page. The repository does not support anonymous submissions or self-uploads of resources. We spotcheck the resource for well-formedness and accordance to our archival formats.

On acceptance of a resource, we compute SHA1 and MD5 checksums for the datastreams of the resource and add them to our own checksum checking tool. We fully validate the metadata using xmllint and we manually check that all relevant information is given and well structured. The repository management gets persistent identifiers for the resource data and metadata and sets up a landing page for the resource. The metadata are ingested in Fedora commons and published on our OAI provider.

When the license conditions allow this, we will set up a download page for the resource

References

1. Reference Model for an Open Archival Information System (OAIS), Recommended Practice, CCSDS 650.0-M-2 (Magenta Book) Issue 2, June 2012 <https://public.ccsds.org/pubs/650x0m2.pdf>
2. <http://fedora.clarin-d.uni-saarland.de/struktur.en.html>
3. <http://fedora.clarin-d.uni-saarland.de/depositors.en.html>

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:

The data are provided in the formats chosen by the data producers from a list of supported formats, see section 2 of this guidelines for the full list of supported formats. Metadata for each resource are always provided in both Dublin Core and

CMDI (Component MetaData Initiative [1]) formats.

Search facilities over metadata are available at our repository [2] but a much more user-friendly search over our metadata is provided by the Virtual Language Observatory (VLO) [3]. Metadata from all 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. Since we cooperate with the VLO within the framework of the CLARIN-D project, we don't plan any improvement of our local search interface.

Harvesting of metadata is implemented via OAI-PMH. Our OAI provider [4] offers such metadata for OAI-PMH [5] harvesting in two formats: Dublin Core and CMDI.

The repository itself does not offer a persistent identifier service on its own but makes use of PID service offered by GWDG [6] based on the handle system [7]. We register handles from the handle service as persistent and resolvable identifiers for our resources.

Furthermore, the repository provides a section for data users [8], where links to search interfaces, data user agreement and citation good practices are provided.

References

1. <https://www.clarin.eu/content/component-metadata>
2. <http://fedora.clarin-d.uni-saarland.de/fedora/objects>
3. UdS CLARIN-D browsing facette at VLO:
https://vlo.clarin.eu/?fq=collection:Universitä_t+des+Saarlandes+CLARIN-D-Zentrum,+Saarbr%C3%BCcken
4. UdS OAI provider: <http://fedora.clarin-d.uni-saarland.de/oaiprovider/?verb=Identify>
5. <https://www.openarchives.org/pmh/>
6. CLARIN's PID short guide: <https://www.clarin.eu/sites/default/files/pid-CLARIN-ShortGuide.pdf>
7. Handle system: <http://www.handle.net>
8. <http://fedora.clarin-d.uni-saarland.de/users.en.html>

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:

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:

The data producer is required to produce metadata accomplishing the formats provided by the repository. The UdS repository requests metadata according to the Dublin Core [1] standard. In the process of ingest, these metadata are semi-automatically converted to CMDI [2] by means of an XSLT template [3] and enriched with additional information, e.g., persistent identifiers (PIDs). The UdS repository provides a list of accepted formats [4], that include common multimedia-document formats as well as formats for binaries. For other file formats, we provide advice for conversion.

We deliberately use the freedom of Dublin Core metadata to leave fields blank or to provide them several times. We insist on metadata on the following topics

author of the resource

title of the resource

description of the resource

rights

We also ask for data on the other basic Dublin Core fields, whenever available.

We manually check the metadata provided by the data producer and ask for additions or clarifications when the metadata are not sufficient. We also offer help and guidance on metadata writing.

We currently use exclusively the CMDI profile clarin.eu:cr1:p_1288172614026 (OLAC-DcmiTerms) [5]

[1] <http://dublincore.org/>

[2] <http://www.clarin.eu/content/component-metadata>

[3] <http://www.clarin.eu/faq/how-can-i-convert-my-dc-or-olac-records-cmdi>

[4] <http://fedora.clarin-d.uni-saarland.de/ressources/AcceptedFormats.en.pdf>

[5] https://catalog.clarin.eu/ds/ComponentRegistry/#/?itemId=clarin.eu:cr1:p_1288172614026

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Additional information on measures to ensure the future usability and understandability of the data, evolution of formats, future migrations or conversions is needed for full compliance.

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 UdS repository runs on a LAMPJ (Linux, Apache (httpd and tomcat), MySQL, Python, Java) software stack. We maintain a software inventory in an internal Wiki and keep up with the necessary software updates. The repository complies with the OAIS reference model's tasks and functions [1]. Besides, the repository is powered by Fedora Commons software, which is compliant with the Reference Model for an Open Archival Information System (OAIS) due to its ability to ingest and disseminate Submission Information Packages (SIPS) and Dissemination Information Packages (DIPS) in standard container formats.

The hardware environment and backup procedure are described in the answer to R9. Saarland University has a high bandwidth connection provided by the DFN (Deutsches Forschungsnetz [2]).

The data consumer has direct access to the archived objects via the web, provided that access requirements have been met. We provide a structure diagram of the repository [3]. 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 are (<https://www.clarin.eu/content/clarin-technology-introduction>):

- * 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.

References

1. Reference Model for an Open Archival Information System (OAIS), Recommended Practice, CCSDS 650.0-M-2 (Magenta Book) Issue 2, June 2012 <http://public.ccsds.org/publications/archive/650x0m2.pdf>
2. <https://www.dfn.de/en/>
3. <http://fedora.clarin-d.uni-saarland.de/struktur.en.html>

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:

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:

The repository is run in the virtual machine environment provided by the Hochschul-IT-Zentrum (HIZ) [1,2,in German]. The virtual machine environment is installed redundantly on two hardware clusters located in two different buildings on the Saarbrücken Campus. Daily backups of the repository are taken using the EMC NetWorker, a professional backup solution [3, in German].

The repository is guarded against unauthorised access by a firewall that exposes only http and https to the outside. It runs on a dedicated virtual machine with restricted access from the the internal network.

The UdS repository is certified as a CLARIN-B centre [4,5].

There are strict criteria to become a CLARIN B-Centre: it should be based on a stable technical and institutional foundation. The Assessment Committee checks these requirements during an assessment procedure, while the technical coordination among the centres takes place in the Centre Committee.

One security criterion is the existence of an SSL certificate for all servers involved that provide a full trust chain.

The technical risks (like hardware failure) are outsourced to the Hochschul-IT-Zentrum and handled by them.

Recovery is done (and it has been actually done in the past) by the repository staff.

1. <https://www.hiz-saarland.de/dienste/virtualisierung/>
2. <https://www.hiz-saarland.de/dienste/webservices/virtualisierungsarten/>
3. <https://www.hiz-saarland.de/dienste/datensicherung/>
4. <https://www.clarin.eu/content/certified-centres>
5. <https://www.clarin.eu/node/3767>

For English summaries of the cited webpages, see Requirement R0.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
For future certifications, evidence of documented procedures and arrangements is needed for full implementation.

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:

We have no additional comments or feedback.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: