



Implementation of the CoreTrustSeal

The CoreTrustSeal board hereby confirms that the Trusted Digital repository IMS Repository complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board.

The afore-mentioned repository has therefore acquired the CoreTrustSeal of 2016 on December 2, 2018.

The Trusted Digital repository is allowed to place an image of the CoreTrustSeal logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the CoreTrustSeal website.

Yours sincerely,

The CoreTrustSeal Board

Assessment Information

Guidelines Version: 2017-2019 | November 10, 2016
Guidelines Information Booklet: [CTS Requirements 2017-2019](#)
All Guidelines Documentation: [Documentation](#)

Repository: IMS Repository
Seal Acquiry Date: Dec. 02, 2018

For the latest version of the awarded
EqgVtwUgcnlqt this repository please visit: <https://www.coretrustseal.org/why-certification/certified-repositories/>

Previously Acquired Seals:

Seal date:	September 2, 2015
Guidelines version:	2014-2017 July 19, 2013
Seal date:	March 12, 2013
Guidelines version:	2010 June 1, 2010

This repository is owned by:

- University of Stuttgart, Institute for Natural Language Processing**
Forschungszentrum Informatik
Pfaffenwaldring 5b
70569 Stuttgart
Baden-Wuerttemberg
Germany

T 0049 711 685 81357
F 0049 711 685 81366
E ims@ims.uni-stuttgart.de
W <http://www.ims.uni-stuttgart.de/>

Assessment

0. Context

Applicant Entry

Self-assessment statement:

Repository Type:

- Institutional repository

- Research project repository

The IMS Repository of the CLARIN-D Resource Centre Stuttgart (<http://clarin04.ims.uni-stuttgart.de/repo/>) is one of currently eight resource and service centres 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 (<http://clarin-d.de>) 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. 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 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.

CLARIN-D is building on the achievements of the preparatory phase of the European CLARIN initiative (<http://clarin.eu>) as well as CLARIN-D's Germany-specific predecessor project D-SPIN (<http://www.d-spin.org>). These previous projects have developed research standards to be met by the CLARIN services 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.

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/>).

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>

A work plan describes the [division of labour of the CLARIN-D centres](#).

As data for inclusion into the IMS Repository, the CLARIN-D centre Stuttgart accepts language resources (corpora, lexical and tools) via pertinent metadata. Furthermore, several REST-based webservice are provided for a variety of different NLP-relevant tasks.

This is explained on the repository web page (<http://clarin04.ims.uni-stuttgart.de/repo/>) and on the CLARIN-D project web page (<https://www.clarin-d.net/en/preparation/find-a-clarin-centre>) - each centre accepts specific kinds of data.

Repository's Designated Community:

Users of computational linguistics software, such as corpora and tools, parameter-based tools, and web services. More precise, data providers to the CLARIN-D infrastructure, and our centre in particular, are individual researchers from academic, non-commercial organisations.

Level of Curation:

Basic Curation (addition of metadata about the resource as a whole)

Outsource Partners:

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

CLARIN-D

The repository in one of currently eight Resource and Service Centres of CLARIN-D.

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.

•

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>, see the linked PDF document on that page, in particular: <http://hdl.handle.net/11372/DOC-138>

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

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.

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.

•

- 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, <http://www.pidconsortium.eu/>). 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 were stipulated: http://clarin04.ims.uni-stuttgart.de/repo/resources/GWDG_PID.pdf

Other Relevant Information:

CLARIN (<https://www.clarin.eu/>) is a European research infrastructure working in the field of archiving and processing of language-related resources in the humanities and social sciences. CLARIN is an acronym for “Common Language Resources and Technology Infrastructure”. Coordinated by the CLARIN-ERIC (European Research Infrastructure Consortium), 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.

- Centres need to have a proper and clearly specified repository system and participate in a quality assessment procedure as proposed by the Data Seal of Approval or MOIMS-RAC approaches.

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

- Centres need to associate PIDs records according to the CLARIN agreements with their objects and add them to the metadata record.

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

- Each centre needs to make explicit statements to the CLARIN boards about its technological and funding support state and its perspectives in these respects.

- Centres need to employ activities to relate their role in CLARIN to the research community in order to guarantee a research based status of the infrastructure and allow researchers to embed their services in their daily research work.

- Centres that are offering infrastructure type of services need to specify their services for CLARIN and the terms of giving service.

- Centres are advised to participate in the Federated Content Search with their collections by providing an SRU/CQL Endpoint. This content search is especially suitable for textual transcriptions and resources.

A short overview of all requirements for centres of type B is also given in the form of a checklist (<https://www.clarin.eu/content/checklist-clarin-b-centres>).

Reviewer Entry

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

Accept or send back to applicant for modification:

Accept

Comments:

The previous concern was about the origin of data held, which is now covered by the brief response "More precise, data providers to the CLARIN-D infrastructure, and our centre in particular, are individual researchers from academic, non-commercial organisations.". But, seeing that it is not compulsory to comply with a level of performance, we should ask them to amend, but not hold certification back.

1. Mission/Scope

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The mission of the IMS Repository is to serve as the repository of the University of Stuttgart's CLARIN centre of Type B. "CLARIN Type B centres offer services that include the access to the [language] resources stored by them and tools deployed at the centre via specified and CLARIN compliant interfaces in a stable and persistent way" (<http://clarin04.ims.uni-stuttgart.de/repo/>, see also: <http://www.clarin.eu/system/files/CE-2012-0037-centre-types-v07.pdf>).

The general mission of CLARIN-D, the German national CLARIN initiative, is to provide "linguistic data, tools and services in an integrated, interoperable and scalable infrastructure for the social sciences and humanities" (see <http://www.clarin-d.de/en/home-en.html>, at the bottom of the page, and <https://www.clarin-d.net/en/about/overview>).

CLARIN's mission is to "Create and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences." Furthermore, CLARIN's value proposition states: "CLARIN offers long-term solutions and technology services for deploying, connecting, analysing and sustaining digital language data and tools" (see <https://www.clarin.eu/content/vision-and-strategy>, and for details: <http://hdl.handle.net/11372/DOC-138>). 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

As part of the CLARIN infrastructure, the IMS repository usually does not carry out promotional activities on its own but participates in such activities on both the national and the European level. These activities do include but are not limited to:

- Providing exhaustive information on the CLARIN mission through websites (e.g. <https://www.clarin.eu>, <http://de.clarin.eu>).
- Operation and maintenance of the Virtual Language Observatory ([VLO](#)) which provides means to search for data/tools to the end user (based on the metadata provided by the resource centers/repositories that are part of CLARIN).
- Presenting data, tools and services provided by CLARIN on conferences.
- Organization of and participation in dissemination conferences that aim at getting in touch with the user communities of CLARIN.
- Organization of pertinent summer schools, training courses, tutorials and workshops.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

Preservation is not explicitly indicated as a mission component in the response, but implied. For example, the requirements placed on centres of Type B by Clarin states that they must meet DSA criteria (amended now to CTS in some documents) – which presumably includes preservation criteria. So our criteria are an explicit requirement from Clarin.

2. Licenses

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The data provider retains all intellectual property rights to their data. The depositor must grant distribution rights to the repository and choose an access model (public, academic, individuals). Access models are provided by the repository and distribution rights are specified in the distribution and license agreement.

There is no guarantee that resources are distributed, that is, the IMS reserves the right to restrict the distribution for ethical or technical reasons. In general it is the IMS' policy to only accept resources that are available for scientific usage.

All CMDI metadata are provided without access restrictions according to CLARIN-D policies. However, for all deposited primary data, depositors need to choose an appropriate licence when they sign the depositor's agreement. Some resources will have restricted access (academic or restricted to individuals vs. public) accordingly. This is supported by the repository, e.g. by Shibboleth-based means.

Data users have to adhere to the licences of individual resources which they use/download via the repository. The users agree to this before access to the data is granted, cf. the Terms of Use and the End-User License Agreements in the CLARIN Model Contracts.

If the data consumer should not comply with the access regulations, the only thing that can practically be done is to deny him/her further access to the IMS repository and to make the research community aware of the misuse. Further legal measures would be reserved to the data depositors. Access to the server host and the web-based administration interface of our Fedora Commons repository is restricted to trained employees of our institute, of course.

Neither the CLARIN-D resource center nor the repository run by it, are legal entities on their own. This also holds for the Institute for Natural Language Processing ("Institut für Maschinelle Sprachverarbeitung", IMS) where they are located. All are part of the University of Stuttgart which is a legal entity - specifically, like all public German universities, "eine Körperschaft des öffentlichen Rechts", an institution governed under public law.

Depositors must sign an agreement stating that they respect IPR (Intellectual Property Rights) and privacy issues and that they own all necessary rights required to deposit the data. In particular, data must be anonymised when applicable. Users must confirm that they will use resources only in the intended way. The depositor can choose to make the data publicly available. Alternatively, he can restrict access to the academic community. Data depositors are held responsible for compliance with any national or international legal regulations.

Pertinent regulations and model contracts are provided for both, depositors and users on basis of the Clarin Model Contracts,

please see the link below and the additional documents provided on the IMS repository home page.

In case a violation of conditions is observed, the original data provider is contacted. In case the violator can be identified, further access by this person/institution will be prevented if technically possible (e.g., via Shibboleth).

Links:

IMS Terms of Use: <http://clarin04.ims.uni-stuttgart.de/repo/resources/terms-of-use.pdf>

Intellectual Property Rights Agreement: http://clarin04.ims.uni-stuttgart.de/repo/resources/data_ipr.pdf

Available CLARIN licences: <https://www.clarin.eu/content/licenses-agreements-legal-terms>

Report on CLARIN Model Contracts: http://weblicht.sfs.uni-tuebingen.de/Reports/D-SPIN_R7.2.pdf

CLARIN License Categories. <https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/ClarinLC>

CLARIN Terms of Service (TOS).

<https://kitwiki.csc.fi/twiki/pub/FinCLARIN/ClarinSA/CLARIN-TOS-2014-10.rt>

CLARIN End-User License Agreements (EULA). <https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/ClarinEULA>

CLARIN Deposition License Agreements (DELA). <https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/ClarinSA>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

3. Continuity of access

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

CLARIN centres commit to ensuring long-term availability, access to and preservation of datasets submitted to their repositories, as set out in their Mission statements. CLARIN centres are set up 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 repository's content would be transferred to another CLARIN centre.

The legal aspects of the process of relocating data to another institution is addressed by the [Depositor's Agreement](#). Although there is no formal agreement concerning the transfer of repository content to another CLARIN center, yet, CLARIN-D has set up a [memorandum of understanding](#) to confirm that all CLARIN-D centers are willing to take over each others repository contents in case a center ceases to exist. Also, the [cooperation agreement of the CLARIN-D centers](#) is publicly accessible now (currently in German only).

Depositing agreements make 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 centre. This is especially important as there may not be any other contractual relationship between a depositor and a data centre stepping in for another centre. Therefore the IMS Repository archives all meta data and primary data in such a way that they can be easily migrated and mirrored at other CLARIN resource centers. All metadata and data have a registered persistent identifier (PID, handle system) and are stored as self contained XML files.

The repository software runs on its own virtual machine hosted on a server at the IMS Stuttgart. The local hard disks of the host system are organized as a RAID array for improved performance and safety. Individual parts are replaced at irregular intervals, depending on the technical requirements which are internally monitored (e.g., S.M.A.R.T. data).

Database dumps and file system backups are performed automatically to dedicated project directories on another IMS server. This latter server is included in the IMS backup plan, i.e. backups are run on a daily basis via the TVS (Tivoli Storage Manager) system provided by the University of Stuttgart's computing services TIK (Technische Informations- und Kommunikationsdienste).

- Backup & Archiving services at TIK (only German information available):

<http://www.tik.uni-stuttgart.de/dienste/Datensicherung/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

4. Confidentiality/Ethics

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Data depositors must sign an agreement stating that they respect IPR (Intellectual Property Rights) and privacy issues

and that they own all necessary rights required to deposit the data

(http://clarin04.ims.uni-stuttgart.de/repo/resources/data_ipr.pdf).

In particular, data must be anonymised when applicable. Users must confirm that they will use resources only in the intended way. The depositor can choose to make the data publicly available. Alternatively, he can restrict access to the academic community or individual

users. Data depositors are held responsible for compliance with any national or international legal regulations.

Pertinent regulations and model contracts are provided for both, depositors and users on the basis of the CLARIN Model Contracts.

In case a violation of conditions is observed, the original data provider is contacted. In case the violator can be

identified, further access by this person/institution will be prevented if technically possible (e.g., Shibboleth).

The resource description is provided by the depositor, or by the IMS in collaboration with the depositor.

The depositor is required to sign a depositor agreement stating that their resource meets disciplinary and ethical norms as specified in the DFG's Rules of "Good Scientific Practice" and the University of Stuttgart's pertinent guidelines ("Richtlinien zur Sicherung der Integrität wissenschaftlicher Praxis"). Additionally, we will review samples of the data before ingest.

There are a number of specific codes of conduct that are applicable to parts of the repository, e.g. the DFG code of conduct. The codes of conduct are in line with generally accepted codes of conduct for research data in Germany.

Any data user is bound by the terms and conditions of use of the repository, as soon as repository services or data deposited are used. In case of misuse, the user is denied further access to the repository. Further legal measures remain reserved to the data depositors. Data providers need to make sure that IPR and personality rights are respected in their deposited data.

Furthermore, the IMS Repository implements the GÉANT Data Protection Code of Conduct. The Data protection Code of Conduct describes an approach to meet the requirements of the EU Data Protection Directive in federated identity management.

Deutsche Forschungsgemeinschaft.

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

GÉANT Data Protection Code of Conduct.

<http://www.geant.net/uri/dataprotection-code-of-conduct/V1/Pages/default.aspx>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Include a link to typical data provider contract for completeness (this was subsequently found here:
http://clarin04.ims.uni-stuttgart.de/repo/resources/data_ipr.pdf).

In a future re-certification through CTS, please include link within the response.

5. Organizational infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

CLARIN centres are hosted by scientific institutions - their repository staff members (see the Stuttgart centre's CLARIN-D project web page: <http://www.ims.uni-stuttgart.de/forschung/projekte/Clarind.en.html>) 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.

In addition, our 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. The [work plan](#) illustrates the division of labour of the CLARIN-D centres.

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.

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. As an alternative to project based funding, CLARIN-D currently pursues a permanent continuation of funding.

The individual CLARIN-D centres are also partly supported by the hosting institutions. By this funding model, based on matching funds from the federal ministry and local institutions, it is possible for each CLARIN-D centre to ensure operation of the repository independent of project funding. The CLARIN-D centre at Stuttgart is supported by the Institute of Natural Language Processing of the University of Stuttgart.

Changes to this repository may be made only by a limited number of authorized and trained data managers, ensuring the safety of both data and repository.

The repository staff consists of scientists with solid knowledge of and experience in the field of the digital humanities data management. The Center staff consist of part-time appointees such that the work force sums up to 1 FTE position: One 0,5 FTE position is financed by CLARIN-D (funded by BMBF, secured until 09/2020), one 0,5 FTE position is financed by MWK (Ministerium für Wissenschaft, Forschung und Kultur, the government department for science, research and culture). The staff's responsibilities are organized into three functional groups: administration, data management, technology.

Administration:

Project leader, board reporting, assessments (Computational Linguist)

Data Management:

Data Manager, quality checks (Computational Linguist)

Technology:

Software developer, linguistic tools infrastructure, web administration, ingest of data (IT Specialist)

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Add explicit link to staff roster or complement for completeness. Not clear from the Clarin-D website what their staffing level is.

6. Expert guidance

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

On a larger scale, CLARIND-IMS is embedded in the CLARIN-D consortium that 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-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: Helen Aristar-Dry, Christiane Fellbaum, Björn Granström, Helge Kahler, Jan Christoph Meister, John Nerbonne, Heike Renner-Westermann and Achim Streit.

The joint Technical Advisory Board (TAB) of CLARIN-D and DARIAH-DE is a committee supports collaboration on the fundamental technical level between 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: Jonas Beskow (University of Stockholm), Carol Goble (University of Manchester), Jan Hajic (Head of the Prague CLARIN Centre), Ed Hovy (University of Southern California), Michael Lautenschlager (German Research Centre for Geosciences, Potsdam), Gerhard Schneider (University of Freiburg), Toma Tasovac (Digital Humanities Centre, Belgrade), Melissa Terras (University College London) and Claire Warwick (University College London). The TAB is currently restructured and its new composition will be announced soon.

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 overcome 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 (<https://www.clarin-d.net/en/clarin-d/work-packages/wp-4-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 100 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 monthly basis. Contents of the curation projects and activities of the WG are published on the CLARIN-D Website [4]. For communication, mailing lists and wiki contents are maintained.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. Data integrity and authenticity

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

After a first discussion about the planned data deposit in close personal contact with the data depositor, data provided by the data depositor is only accepted for ingestion

(1) if it matches the IMS repository's focus on "corpora and corpus tools, adaptable tools and web services" (see the [repository's description](#); otherwise she is referred to [corresponding CLARIN centres](#)) and

(2) after the identity of the depositor could be verified.

Furthermore, only non-proprietary, text-based data formats are accepted (as is usual in this field) - this facilitates long-term readability and preservation of the data.

The data provided by the data depositor is considered to be fixed and immutable - in case of format transformations (by the data depositor), the modified version of the data is treated like a new submission, with a link to the previous version.

The integrity of the data is fostered by using checksums (MD5) in Fedora. There is also a version control

mechanism in the Fedora Commons backend. CMDI metadata (according to [ISO 24622-1](#)) are represented as a data stream within Fedora

Digital Objects, and as such they can be version-controlled like all other object data.

It should be noted that we decided to do strict versioning with respect to the assignment of PIDs only for primary

(=research) data, not for the metadata. That is, changes to metadata will generally not result in a new PID being

registered. In contrast, changes to primary data will always result in a new data stream or digital object and,

accordingly, a newly registered and associated persistent identifier. However, we make use of the built-in

Fedora-internal versioning mechanism in order to keep track of changes to the CMDI metadata files. Hence,

respective changes can still be traced and old versions remain accessible at least in principle.

Part of the archiving workflow consists in an integrity and quality check of the data and the metadata. This is

brought about semi-automatically, e.g. well-formedness and validity can be checked automatically for XML

metadata, but they are also manually probed in order to check that descriptions actually make sense. The

object data are tested for syntactic correctness if possible, depending on the data type and format. For this purpose, a front-end

is used that helps creating valid CMDI metadata using components and profiles stored in the [Component Registry](#).

The [CMDI creation workflow](#) is described on a wiki page.

These best practices are summarised on a [public wiki page on data integrity and authenticity](#).

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

8. Appraisal

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The focus of the IMS repository is on corpora and corpus tools, adaptable tools and web services (<http://clarin04.ims.uni-stuttgart.de/repo/>).

Depositors are encouraged to use non-proprietary formats listed in the [CLARIN standard recommendations](#) 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 (by the depositor) before it is archived. Data in accessible formats and with the necessary access permissions may be made available to content-based analysis tools such as the [Federated Content Search](#). It can also be used as the input for automatic natural language processing tool chains such as [WebLicht](#). The list of accepted data formats may be extended to include new, widely-used formats in the field.

The depositor, with assistance from a data manager of the CLARIN-D Resource Centre Stuttgart if necessary, creates [CMDI metadata](#) using components (description building blocks including field definitions) and profiles (components grouped into a ready-made description blueprint) stored in the [Component Registry](#). Predefined CMDI components can be combined into a set forming a CMDI profile. This profile is then available to other users (stored in and shared via the Component Registry to promote reuse) which might want to describe similar resources, and it may be extended using additional CMDI components. Each metadata record is then expressed as an XML file, including a link to the profile on which it is based.

The data manager tries to help with questions concerning the CMDI metadata specifications. Furthermore, she tries to make an informed decision about the inclusion of the research data into the repository on the basis of their relevance to research purposes in the field. The data might also be accepted if it is not in a preferred format (and is not easily convertible by the depositor) - in this field, researchers work on a great amount of differing data formats based on the tools used to generate and further process them. For that reason, quality checks of the data are not easily feasible. However, so far data is only accepted if the depositor can be trusted, e.g. if she is personally known or demonstrates to be affiliated with a familiar academic institution or if a paper on the data has been peer-reviewed and published.

Reviewer Entry

Accept or send back to applicant for modification:

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

Accept

Comments:

9. Documented storage procedures

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Data deposited to the IMS repository remains under the ownership and responsibility of the data depositor (described in the [depositor's agreement](#)), who guarantees that data content is in accordance with German law. The data depositor decides if data should be made publicly accessible or if access should be restricted to academic access (login with user's existing institutional credentials via CLARIN's Service Provider Federation).

Metadata in CMDI format is created and optimised in close collaboration with the depositor to ensure an adequate description of the data; it consists of a series of steps described on the [metadata workflow page](#).

The repository is implemented as a setup of the Fedora Commons Repository Architecture. Resource representations are stored within both a database and the file system for improved disaster recovery.

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 of the resource will be restored.

Database dumps and file system backups are performed automatically to dedicated project directories on another server. This latter server is included in the IMS backup plan, i.e. backups are run on a daily basis via the TVS (Tivoli Storage Manager) system provided by the University of Stuttgart's computing services TIK (Technische Informations- und Kommunikationsdienste).

Documentation:

Fedora Commons Repository Software: <http://www.fedora-commons.org/>

Backup & Archiving services at TIK (only German information available):

<http://www.tik.uni-stuttgart.de/dienste/Datensicherung/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

The key question of Data Quality must be more completely addressed in any re-submission; quality does not consist of the integrity of checksum hash value verification. ACCEPTED AT COMPLIANCE LEVEL 3 ONLY

10. Preservation plan

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The repository contains (and accepts) only non-proprietary, text-based data formats, making for low curation requirements to ensure long-term preservation considering data content.

Depositors must sign a [Depositor's agreement](#) with the University of Stuttgart, 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 or academic). 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 depositor's agreement informs about the repository's responsibilities in the context of long-term preservation:

- storage of the data (sustainable, legible, accessible)
- data preservation in unchanged form (as far as possible)
- access to third parties (public or restricted academic access)
- remove data from the archive (if necessary)

- move data to a cooperating CLARIN centre or a similar infrastructure (in case of cessation of the hosting service)

Technical preservation arrangements:

- integrity tests of data stored
- periodical local and distributed backups (located in dedicated computing centers with strict access control)
- repeated testing of reinstalling the repository from backup
- administrator access to the repository is limited to a small group of trained experts

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.

Please see our public [document listing these preservation measures](#).

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

11. Data quality

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The IMS 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 IMS repository are invited to develop a [data management plan](#) in coordination with the staff of the CLARIN centre, which is offered as a free 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 Stuttgart CLARIN centre - in any case, it will be thoroughly checked and discussed by experienced centre staff.

The metadata portal CLARIN [Virtual Language Observatory](#) harvests ISO 24622-1 conformant metadata (CMDI) 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, 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 so-called [discipline-specific working groups](#) 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 IMS repository 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

Accept or send back to applicant for modification:

Accept

Comments:

The key question of Data Quality must be more completely addressed in any re-submission. ACCEPTED AT COMPLIANCE LEVEL 3 ONLY

12. Workflows

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The [standard data deposition workflow documented on a public wiki page](#) consists of the following steps:

Ingestion requirement checks:

Does the data content match the repository's mission?

Is the data in one of the acceptable formats (non-proprietary, text-based) or can it be converted?

The handling of requests to deposit data that do not fall within the CLARIN mission of the IMS repository (as described above) has to be decided on a case by case basis, but prospects will usually be negative. Data that conform to our mission statement will be prioritized in any case. But, the depositor may be referred to an appropriate CLARIN centre.

1.

If the data can not be provided in an acceptable format, it will not be stored in the repository. However, appropriate metadata with a link to the original data source can still be accepted.

Depositor's agreement terms:

Are scientific and ethical norms considered?

Are personal data contained in the data, and if so: is privacy protection ensured?

Does the depositor hold all rights to publish the data?

Which access license to end-users shall be granted? - Public or academic?

To sign the agreement, the depositor will have to meet above preconditions contained in that contract. Access to the research data has to be determined in accordance with the license chosen by the depositor. Metadata always have to be publicly available.

2.

3. Metadata creation:

In close cooperation with the depositor, appropriate metadata are created and reviewed. A human reviewer probes the data submitted by external providers for basic compliance to the depositor's description.

4. Signing of depositor's agreement.

In this step, the identity of the data depositor has to be verified, too. Moreover, data from depositors are preferred who show that they published a paper about their data or submitted it to a peer-reviewed journal.

Persistent storage and assignment of persistent identifiers:

Metadata (and data) are stored in the repository.

There is currently no formal curation policy regarding when to deprecate open, text-based data formats (accepted exclusively by the repository) and how to deal with such data.

The IMS Repository uses Fedora Repository as its base. Hence, our technical workflows are developed on top of the

provided batch utilities for ingest and the API REST interfaces for access and management provided by the

system.

A big picture of the steps involved: packaging/updating of the resource, creating or transformation of the metadata

(where necessary), quality check of the data and metadata (e.g. validation, where applicable), registering PIDs

(Persistent Identifiers, handle system) and inserting them in the CMDI metadata records.

5.

Links:

- Fedora Repository. <http://fedorarepository.org/>

- Documentation of the Fedora REST API. <https://wiki.duraspace.org/display/FEDORA38/REST+API>

- Documentation of the Fedora Digital Object Model.

<https://wiki.duraspace.org/display/FEDORA38/Fedora+Digital+Object+Model>

- Handle System. <http://www.handle.net/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

13. Data discovery and identification

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

All [CLARIN centres](#) provide their metadata according to ISO 24622-1 (CMDI) via OAI-PMH. The [Component MetaData Infrastructure](#) (CMDI) 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 [Component Registry](#). 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](#)). Since different communities use different names for the same concepts, linking CMDI elements to concepts enables communities to retain their terminology while enabling users to find concepts independent of the naming.

A strict requirement for CLARIN centres is to make their metadata available through the established and well documented [Open Archives Initiative Protocol for Metadata Harvesting](#) (OAI-PMH). This standard enables harvesting of the metadata from the repository via http(s).

The CLARIN [Virtual Language Observatory](#) (VLO) harvests the metadata in CMDI format from all CLARIN centres via OAI-PMH. 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.

All datastreams in the repository are assigned a persistent identifier, which can be used for citing and otherwise referring to the data. The repository itself does not offer a persistent identifier service on its own but makes use of a common [CLARIN PID service](#) based on the [handle system](#), in cooperation with the European Persistent Identifier Consortium (EPIC), implementing ISO 24619. The PID registration and resolution services are provided by the GWDG, as described in section 0. The usage of PIDs is mandatory for resources in CLARIN, thus all resources added to the repository may be referenced using PIDs.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

Comments:

14. Data reuse

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

All [CLARIN centres](#) provide their metadata according to ISO 24622-1 (CMDI) via OAI-PMH. The [Component MetaData Infrastructure](#) (CMDI) 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 [Component Registry](#). 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](#)). Since different communities use different names for the same concepts, linking CMDI elements to concepts enables communities to retain their terminology while enabling users to find concepts independent of the naming.

The designated community in this field uses a wide variety of data formats. However, data depositors are encouraged to use non-proprietary, text-based formats, preferably those listed on the [CLARIN standard guidance website](#) and the [CLARIN standard recommendations FAQ](#). Use of these formats will ensure that the data is interoperable within the CLARIN infrastructure, and thus, readability for future conversions (if needed) are ensured. 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.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

: It is not clear from the response if the centre makes any attempt to standardise data for improved re-use or whether data is offered in standard services (although looking at the range of standards listed in the website it is clear that they do - <https://www.clarin.eu/content/standards-and-formats/>). Provide explicit evidence of such efforts in any future re-certification.

15. Technical infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

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:

- 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.

With the use of the Fedora-Commons system and the defined workflow supported by the repository's interface, the repository aims to be as conformant to OAIS as possible. Provisions for the main functional entities described in OAIS are summarized as follows:

Ingest: Archive managers upload data into the system and assign a PID to the digital object. During the ingest process, CMDI metadata (ISO-CD 24622-1) appropriate for the type of resource must be created.

Archival Storage: The Fedora Commons system is used for permanently storing the data. Backups of the system are created and disaster recovery services are in place. The Fedora Commons built-in version control system is used.

Data management: The standard Fedora Commons tools, in combination with a custom administration application are used for data management. Metadata is distributed via the OAI-PMH protocol, supporting selective harvesting as well. Both the OAI-PMH supplied metadata and the Fedora Commons tools are used to report on the status of the data.

Administration: Using a local Authentication, Authorization and Access Infrastructure in Fedora Commons, data managers conduct administrative tasks. The hardware is securely stored in locations with highly restricted access.

Preservation Planning: A technology audit is used to evaluate the state of technology, long time efficiency and test migration procedures when new platforms become available. The migration tests are conducted routinely to different hardware even if the productive environment is not migrating. The cooperation with partner projects supports the preservation activities. The open format used by Fedora Commons guarantees the long-term accessibility of the data.

Access: The digital objects are available for reading access via their PID for authorized users, based on the AAI infrastructure of the CLARIN Service Provider Federation and a local user management. The PIDs are available in the metadata, which can be harvested via OAI-PMH (e.g. by the VLO).

The technical infrastructure and processes described here have been tested and evaluated and have been determined to be fully functional for the needs of our repository. State of the art firewalls block unauthorized access to the systems on which the repositories are being operated, including access to administrative tools and backends from unauthorized workstations.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

16. Security

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The IMS repository is certified as a CLARIN-B centre.

There are strict criteria to become a CLARIN B-Centre: it should be based on a stable [technical](#) infrastructure. The [Assessment Committee](#) checks these [requirements](#) during an [assessment procedure](#), while the the centres takes place in the [Centre Committee](#).

One security criterion is the existence of an SSL certificate for all servers involved that the repository is guarded against unauthorised access by a two step firewall (central firewall and local firewall provided by scientific linux)that exposes only http and https to the outside world. It runs on a dedicated virtual machine (scientific linux) with restricted access from the internet.

Furthermore, the University of Stuttgart provides a Computer Emergency Response Team (RUS-CE) for computer and network security in the University's IT infrastructure [2]. They report security suspicious activities) to our repository administration team.

Disaster recovery and planning is outsourced to TIK (<http://www.tik.uni-stuttgart.de/dienste>) service continuity is ensured by agreements with the other centres in the context of the CLA

[1] (German) <http://www.tik.uni-stuttgart.de/dienste/itsicherheit/firewall/index.html>

[2] <https://cert.uni-stuttgart.de/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

17. Comments/feedback

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

0. N/A: Not Applicable.

Self-assessment statement:

Please note that the text for Requirement 0 was extended in coordination with the other CLARIN-D centres currently applying for DSA/CTS.

As for Requirement 8:

We checked the link to the repository website and it should keep working (we are not aware of the server being down for several days) - in the reviewer's comment, though, a minus was missing in the link. Here's the correct link: <http://clarin04.ims.uni-stuttgart.de/repo/>

In doubt, the link is also accessible from our CLARIN project page:

<http://www.ims.uni-stuttgart.de/forschung/projekte/ClarinD.en.html>

As for Requirement 10:

We are unsure how an implemented preservation plan for the needs of this specific repository might look like other than described. We feel it is detailed and complete, but we'd appreciate suggestions on what might be missing or examples for more suitable descriptions.

As for Requirement 5:

In accordance with the other CLARIN-D centres and the laws of privacy protection, staff members' names and their exact part-time activities are not mentioned.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Please note comments that should be observed in any future application for re-certification by CoreTrustSeal.