Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository IMS Repository complies with the guidelines version 2014-2017 set by the Data Seal of Approval Board.
The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2013 on September 2, 2015.

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

Yours sincerely,

The Data Seal of Approval Board

Assessment Information

Guidelines Version:	2014-2017 | July 19, 2013
Guidelines Information Booklet:	DSA-booklet_2014-2017.pdf
All Guidelines Documentation:	Documentation
Repository:	IMS Repository
Seal Acquiry Date:	Sep. 02, 2015
For the latest version of the awarded DSA for this repository please visit our website:	http://assessment.datasealofapproval.org/seals/
Previously Acquired Seals:	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. Repository Context

Applicant Entry

Self-assessment statement:

The IMS Repository of the CLARIN-D Resource Centre Stuttgart (http://clarin04.ims.uni-stuttgart.de/repo/) 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 (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. 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.

The IMS Repository offers language resources (corpora, lexical and tools) via pertinent metadata. Furthermore several REST-based webservices are provided for a variety of different NLP-relevant tasks.

Within CLARIN-D 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.

Within CLARIN-D the following requirements hold for centres of type B (https://www.clarin.eu/node/3542)and are fulfilled by this resource centre:

- Centres need to offer useful services to the CLARIN community and to agree with the basic CLARIN principles (own architecture choice, explicit statement about quality of service, usage of persistent identifiers,

adherence to agreed formats, protocols and APIs).

- 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. In case all resources at a centre are open, setting up a Service Provider is optional.

- 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).

List of outsource partners:

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, 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:
https://clarin04.ims.uni-stuttgart.de/repo/resources/GWDG_PID.pdf

This outsource partner offers relevant functionality for guidelines 7-11.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

This is a good and comprehensive desription of the context of the repository.

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

The IMS repository is located at the Institute for Natural Language Processing (IMS) at the University of Stuttgart. The repository's focus is on language resources provided by the IMS and other CLARIN-D related institutions such as the local Collaborative Research Centre 732 (SFB 732) as well as persons, institutions and/or organizations that belong to the CLARIN-D extended scientific community (digital humanities and social sciences). Comprehensive guidelines and workflows for dealing with submissions by external contributors have been compiled based on the experiences of archiving in-house resources.

Generally, our archiving service is only provided for resources described by a full set of CMDI (http://www.clarin.eu/cmdi) metadata which have to contain relevant information in order to assess the scientific and scholarly quality of the resource, e.g. in terms of data format, annotation guidelines, peer-reviewed publications describing or using the resource. All CMDI metadata are made publicly available via our repository's web frontend (http://clarin04.ims.uni-stuttgart.de/fedora/objects) and can be harvested via OAI-PMH (http://clarin04.ims.uni-stuttgart.de/oaiprovider/oai?verb=Identify). Within CLARIN, this information is aggregated and used by the Virtual Language Oberservatory (http://www.clarin.eu/vlo/) and the WebLicht platform (https://weblicht.sfs.uni-tuebingen.de/). Furthermore, we usually provide access to the resource data themselves for means of download or online usage via web services (either freely available, restricted to members of academic institutes or on an individual basis).

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.

Additional links:

- Deutsche Forschungsgemeinschaft: Rules of Good Scientific Practice
http://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/

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

- CLARIN Licenses, Agreements, Legal Terms. http://clarin.eu/content/licenses-agreements-legal-terms

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

2. The data producer provides the data in formats recommended by the data repository.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

The IMS repository requires compliance to the CLARIN standards recommendations for the LRT domain. These recommendations go beyond mere formats and call for standards in the following areas, please note the non-exhaustive lists of examples given in parentheses: general standards (XML, XML Schema, RelaxNG, URIs, Handles as persistent identifiers, ISO 639-3 language codes, ISO 3166 country codes), protocols (OAI-PMH, WSDL, SOAP, REST), terminology / ontologies (ISOcat, EAGLES/ISLE, GOLD), metadata (Dublin Core,OLAC, TEI, CMDI), media formats (MPEG1/2/4, JPEG, MP3), general (HTML, PDF, RTF, CSV) and LRT-specific text formats (LMF, (X)CES, TEI, EAF, LAF) and, finally, text encoding (Unicode, ASCII).

We check for compliance to these recommendations when reviewing the (meta-)data submitted for archiving. Metadata have to be provided in CMDI compliant form, other additional formats are possible. In case of non-compliant formats, we may provide advice about conversion to a recommended standard where applicable or deny the request.

Links:

- CLARIN standard recommendations document.
http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf

- CLARIN overview on standards and formats. http://www.clarin.eu/content/standards-and-formats

- CLARIN standard guidance. http://clarin.ids-mannheim.de/standards/

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

3. The data producer provides the data together with the metadata requested by the data repository.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The data producer has to deposit not only the research data, but also metadata in a format that complies to our regulations (see above). Specifically, the metadata have to be in the CMDI format or in a format that can be automatically transformed to CMDI. For some established formats, CMDI profiles and pertinent XSLT stylesheets are already available, e.g., for the conversion from Dublin Core to CMDI. Where this is not the case, we may assist in the creation of such profiles and stylesheets. However, this has to be decided on a case-by-case basis.

The CLARIN initiative provides exhaustive documentation (http://www.clarin.eu/cmdi) on how to create CMDI compliant metadata profiles and instances. Additionally, tools are provided that allow data producers to easily create or adapt metadata to the CMDI standard.

The compliance of the submitted metadata to a specific CMDI profile (-> XML schema) is validated as part of the ingest procedure. As an additional option, metadata in other CLARIN-endorsed formats (cf. the CLARIN standards recommendation document: e.g., TEI Headers) can be provided as the content of additional datastreams of Fedora Digital Objects in the repository. A minimal set of Dublin Core metadata, needed in order to adhere to the OAI-PMH protocol for disseminating the metadata, will be created automatically during ingest.

Additional Links:

- Dublin Core: http://dublincore.org/

- CMDI: http://www.clarin.eu/cmdi

- Conversion procedure to CMDI: http://www.clarin.eu/faq/how-can-i-convert-my-dc-or-olac-records-cmdi

- CLARIN, standard recommendations. http://www.clarin.eu/recommendations

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

4. The data repository has an explicit mission in the area of digital archiving and promulgates it.

Minimum Required Statement of Compliance:

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

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 (http://clarin04.ims.uni-stuttgart.de/repo) 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://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“ (http://www.clarin-d.de/en/home-en.html).

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. clarin.eu, 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.

Links:

- http://clarin04.ims.uni-stuttgart.de/repo

- http://www.clarin-d.de

- http://www.clarin.eu

- http://www.clarin.eu/files/centres-CLARIN-ShortGuide.pdf

- http://www.clarin.eu/system/files/CE-2012-0037-centre-types-v07.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

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 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 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., Shibboleth).

Links:

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

- CLARIN Licenses, Agreements, Legal Terms. http://clarin.eu/content/licenses-agreements-legal-terms

- 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.rtf

- 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:

6. The data repository applies documented processes and procedures for managing data storage.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

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 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).

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/

http://www.tik.uni-stuttgart.de/dienste/datensicherung/backup/Datensicherung_mit_TSM.html

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. The data repository has a plan for long-term preservation of its digital assets.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

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 should be encoded in Unicode to ensure future interpretability.

Access to data and metadata is provided via widely used open source software stacks (MySQL, Tomcat, Fedora Repository) that are installed on virtual machines. This maximizes the probability of long term support (updates, security fixes) for the tools being used and improves the ability to run installations of these software stacks independent from the underlying hardware/operating system.

Many parts of the CLARIN infrastructure do address the migration of data from one resource center / repository to another. Since the usage of these infrastructure services (e.g. a PID system, CMDI) is obligatory, every CLARIN center is, to a certain extent, ready to move it's digital assets to another center. This is of paramount importance in case a center/repository would be unable to continue offering its services. The virtual machines could be hosted by other centres, for example.

Links:

- Local Repository Home. http://clarin04.ims.uni-stuttgart.de/repo

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

8. Archiving takes place according to explicit work flows across the data life cycle.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

The IMS Repository uses Fedora Repository as its base. Hence, our 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.

Access to the research data has to be determined in accordance with the license chosen by the depositor. Metadata always have be publicly available.

Before starting the technical ingest procedure, a human reviewer probes the data submitted by external providers for basic compliance to the depositor’s description. There is currently no formal curation policy regarding when to deprecate data formats and how to deal with such

data.

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.

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:

9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.

Minimum Required Statement of Compliance:

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

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.

Crisis management is based on the technical solutions described above. In addition, 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.

Links:

- http://clarin.eu/content/licenses-agreements-legal-terms

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. The data repository enables the users to discover and use the data and refer to them in a persistent way.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

Local search facilities are provided on the basis of the search interface of Fedora Commons. In addition, the DC and CMDI metadata are provided via the OAI-PMH protocol. They are collected by the OAI-PMH harvester of the virtual language observatory (http://www.clarin.eu/vlo/) and WebLicht (https://weblicht.sfs.uni-tuebingen.de/) among others.

The search interface of the VLO provides a central starting point for searching among the aggregated resources of all CLARIN partners that

offer their data in this way. The VLO also features a very useful facetted browsing functionality. For some resources “deep search” (of distributed textual content contained in the primary data) is supported by means of the CLARIN Federated Content Search (http://www.clarin.eu/fcs) interface for some resources.

Unique persistent identifiers according to the Handle system (http://www.handle.net) are provided for the resources. This makes them citable, even if the URLs should change at some point. Where available, useful reference publications are included in the CMDI metadata.

Restricted access to resources is implemented as web-based Single-Sign-On (SSO) via Shibboleth/SAML2 (http://shibboleth.net/about/index.html) and membership in associated federations of trust.

Since usage of widely accepted standards and recommended formats is enforced, the data will be usable by the community.

Links:

- Search interface. http://clarin04.ims.uni-stuttgart.de/fedora/objects

- OAI-PMH Provider. http://clarin04.ims.uni-stuttgart.de/oaiprovider/oai?verb=Identify

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

11. The data repository ensures the integrity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

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 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 manual probing is still a good idea 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.

Link:

- Local Repository Search. http://clarin04.ims.uni-stuttgart.de/fedora/objects

- Local Repository Home. http://clarin04.ims.uni-stuttgart.de/repo

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

12. The data repository ensures the authenticity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

The repository in principle makes the originally 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 in the resource data, a new data stream or digital object with a new persistent identifier will be created. When new versions are stored in the repository, previous versions are maintained by a version control system built into the repository back end. In the case that changes have to be made to the data, e.g., because a file format becomes obsolete and superseded, the original data would also be kept. For updates of the metadata only, however, we do not create a new digital object with a new persistent identifier.

Generally, the repository only accepts works from the original data producers, who are acknowledged as such by means of the element in the CMDI metadata. The data producers should also contribute the exact date and time when the resource was prepared, submitted as part of the CMDI metadata.

Currently there is no explicit check of the identity of depositors since especially in the first phase of CLARIN only data that were provided by well-known partners have been added to the repository. For the time being, we will often know the depositors personally from the scientific community and will be in close contact with them during the ingest process. Nevertheless, no external resource will be ingested without the

depositors having signed a depositor’s agreement.

A limited number of authorized and trained data managers at our repository ensure the safety of both data and repository. Access to the administration facilities of the repository is restricted to these persons only.

Link:

- Local Repository Home. http://clarin04.ims.uni-stuttgart.de/repo

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

The IMS Repository is powered by the Fedora Repository software which is compliant with the OAIS reference model due to its ability to ingest and disseminate Submission Information Packages (SIPS) and Dissemination Information Packages (DIPS) in standard container formats. Archive Information Packages (AIPs) together with descriptive and technical metadata are represented in the form of Fedora Digital Objects, which are serialized in the Fedora XML format (FOXML) and stored as self-contained XML files (Archival Storage). Hence, the IMS Repository complies with the OAIS reference model's tasks and functions to the best of our knowledge.

The data consumer has direct access to the archived objects via the web, provided that access requirements have been met. For metadata we rely on the group of emerging standards around CMDI (ISO-CD 24622-1). The repository is part of the CLARIN infrastructure and will fulfill current and future requirements decided on by the CLARIN board.

Links:

- 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

- Fedora Commons: http://fedora-commons.org

- CMDI. http://clarin.eu/content/component-metadata

- Standards and Formats. http://clarin.eu/content/standards-and-formats

- CLARIN. http://clarin.eu

- CLARIN-D. http://clarin-d.de

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

14. The data consumer complies with access regulations set by the data repository.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

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.

Links:

CLARIN Terms of Use. https://kitwiki.csc.fi/twiki/pub/FinCLARIN/ClarinSA/CLARIN-TOS-2014-10.rtf

CLARIN EULA. https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/ClarinEULA

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

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.

Links:

- Deutsche Forschungsgemeinschaft: Rules of Good Scientific Practice
http://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/

- CLARIN Terms of Use. https://kitwiki.csc.fi/twiki/pub/FinCLARIN/ClarinSA/CLARIN-TOS-2014-10.rtf

- 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:

16. The data consumer respects the applicable licences of the data repository regarding the use of the data.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

Our general workflow does not allow for the integration of data into the repository without the specification of access criteria and without providing an appropriate license. If applicable, the data consumer is made aware of usage restrictions for the data s/he has gotten access to. These license conditions are available to the users of the repository, e.g. via the CMDI metadata. Generally, the usage restrictions are already described in the pertinent codes of conduct. For some data, explicit statements may need to be made by the data consumer about the usage of the data in terms of agreeing to a special licence agreement before s/he can get access to the data. The depositor then decides on whether access is granted or not.

In case of misuse, the only thing that can be practically done is to deny the user further access to the repository. Further legal measures remain reserved to the data depositors.

Links:

CLARIN Licenses, Agreements, Legal Terms. http://clarin.eu/content/licenses-agreements-legal-terms

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: