Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository HZSK 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 June 27, 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:	HZSK Repository
Seal Acquiry Date:	Jun. 27, 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: May 3, 2013 Guidelines version: 2010 | June 1, 2010
This repository is owned by:	**Hamburger Zentrum für Sprachkorpora** Max-Brauer-Allee 60 22765 Hamburg Hamburg Germany T 0049 (40) 42838-6425 E corpora@uni-hamburg.de W http://www.corpora.uni-hamburg.de/

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

The HZSK Repository (https://corpora.uni-hamburg.de/repository) is part of CLARIN-D (Common Language Resources and Technology Infrastructure Deutschland) - a web and centres-based research infrastructure for the social sciences and humanities. The aim of CLARIN-D and its service centres is to provide linguistic data, tools and services in an integrated, interoperable and scalable infrastructure for the social sciences and humanities. The research infrastructure is rolled out in close collaboration with expert scholars in the humanities and social sciences, to ensure that it meets the needs of users in a systematic and easily accessible way. CLARIN-D is funded by the German Federal Ministry for Education and Research.

CLARIN-D is building on the achievements of the preparatory phase of the European CLARIN initiative as well as CLARIN-D's Germany-specific predecessor project D-SPIN. These previous projects have developed research standards to be met by the CLARIN 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.

This repository offers resources such as mainly spoken multilingual corpora and several REST-based webservices for the visualization and conversion of such corpus data.

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). 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 attached document lists the services which were stipulated.

This outsource partner offers relevant functionality for guideline 10: „The data repository enables the users to utilize the research data and refer to them.“.

2) Forschungszentrum (FZ) Juelich GmbH

The CLARIN-D Ticketing System as part of the CLARIN-D Helpdesk (http://support.clarin-d.de/) which is operated by the HZSK is hosted on a virtual machine by the Forschungszentrum Juelich. CLARIN-D has a contractual relationship with FZ Juelich concerning long-term storage and hosting of workspaces and virtual machines.

3) Regionales Rechenzentrum (RRZ) - Universität Hamburg

The HZSK Data Repository as part of the CLARIN-D Infrastructure is hosted on a virtual machine by the RRZ Hamburg. For the respective policies of the RRZ see section 6.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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:

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

Self-assessment statement:

The HZSK-repository contains mainly spoken language corpora. For projects just starting to compile corpora, the HZSK provides assistance in all areas related to data formats, legal issues and compliance to regulations and requirements given in CLARIN.

For projects that are finished with their data collection, the HZSK supervises a curation process to ensure that data and metadata comply with the repository requirements.

The minimal requirements for corpora before the curation process starts are:

- data provided in a standardized format or a comprehensive documentation of a proprietary format
- metadata in a well documented and accessible format
- contact information on the data depositor / data producer
- a statement on the legal status of the resource

- a contract about the means and procedure of accessing the data

- compliance with internal and CLARIN guidelines concerning scientific and scholarly quality

- information on how the data was originally created

The depositor is required to deliver metadata according to CMDI profiles compliant with the requirements of the HZSK Repository. The HZSK provides several basic flexible profiles that can be adjusted - as new profiles - to project specific description needs. The specification of the exact metadata schema is agreed upon between the depositor and the HZSKand in cooperation with the depositor the HZSK creates a description of the corpus. We encourage depositors to provide further references to work describing the resource and its creation in detail.

The depositor is alone responsible for the consent of the participants in the data and for compliance with ethical codes of conduct and national and international legal regulations.

In the curation process ...

- ... if necessary, data is converted into suitable standardized formats (EXMARaLDA, ELAN/EAF, TEI)

- ... metadata is converted into standardized formats (EXMARaLDA corpus metadata (where applicable), CMDI metadata)

Data sharing and reuse is promoted by providing access to the data (download, webservices) and metadata (via the OAI-PMH protocol and the repository itself) free of charge. The CLARIN infrastructure contains software components such as the VLO (http://www.clarin.eu/vlo/) which enable users to browse and search through combined catalogs (metadata of all CLARIN repositories and further institutions and archives). New resources are promoted through announcements on relevant mailing lists.

For most resources, it is necessary to request (free) access to the corpus and state the nature of the intended work.

Example: http://www.corpora.uni-hamburg.de/sfb538/bipode_nutzungsvereinbarung.pdf

For these resources, citations are requested in publications using these resources. All resources are described in a comprehensive way with references to publications describing the resources.

Example: http://www.corpora.uni-hamburg.de/sfb538/en_h9_hacaspa.html

Since the HZSK hosts very heterogeneous resources, no claim can be made about its general reusability. Many of the resources are requested regularly.

See: https://corpora.uni-hamburg.de/drupal/en/resources and https://corpora.uni-hamburg.de/drupal/en/corpora-sfb538

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:

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

Self-assessment statement:

All spoken language corpora hosted at the repository have been converted into best practice formats during the curation process outlined in DSA guideline 1.

The XML-based EXMARaLDA format is a widely spread format for transcriptions with (mainly manually created) annotations and can be considered best practice. For all EXMARaLDA transcription files in a hosted corpus, the HZSK provides corresponding files automatically converted into the best practice formats TEI, EAF and sometimes also in the CHAT format depending on the original transcription guidelines used. Corpus metadata and metadata on communications (sessions), speakers and all the relationships and features of physical files is stored in the XML-based EXMARaLDA metadata format Coma, which is also part of the EXMARaLDA system. We store audio recordings in the uncompressed WAV format. All main corpus files are stored in open, standard or best practice formats.

To accept further corpora to the repository, we will require the corpus to either be in the EXMARaLDA formats, or to be convertible into the above mentioned formats in a corpus curation process as described in DSA guideline 1, or to use other standardized best practice formats for the transcription data (depending on data complexity etc.) and provide the metadata directly in the CMDI format. Before the HZSK decides on conducting a curation process, detailed information about the file formats and the tools and methods by which the files were created is requested.

The cost-benefit analysis leading to the decision on whether to perform the curation is based on a publicly available guideline.

The HZSK does not allow corpora to be deposited into the repository without previous curation. Since the EXMARaLDA software tools require valid files, we ensure compliance with the format requirements by using them in the curation process, for other best practice formats, corresponding validation of the data is required.

References:

The EXMARaLDA format: http://www.exmaralda.org/

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

The corpora hosted at the HZSK (information page, not part of the repository):
http://www.corpora.uni-hamburg.de/sfb538/en_overview.html

Guideline for the cost-benefit analysis for corpora curation (in German):
https://corpora.uni-hamburg.de/pdf/Leitfaden_Aufbereitungsaufwand_und_Nachnutzbarkeit_von_Korpora.pdf

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:

Since all corpora that are to be ingested into the HZSK repository undergo a manual curation process supervised by the HZSK, there is only little automated guidance for data depositors.

As part of that curation process, we try to obtain as many missing metadata as possible from the data depositors.

For guidance for projects just starting with their data compilation, please refer to DSA guideline 1. We also provide a metadata schema with recommended metadata elements, the EXMARaLDA core metadata schema, and more general recommendations through the CMDI metadata profiles used for the data in the HZSK repository.

For depositors, metadata including standardized DC and OLAC metadata for resource description can be created using EXMARaLDAs Corpus Manager tool (Coma).The Coma format can then be converted into CLARIN-compliant CMDI metadata. It is also possible to provide the metadata directly in the CMDI format or, depending on the specific project context and agreements between the depositor and the HZSK, in other automatically processable formats from which the required CMDI metadata can be generated.

References:

OLAC metadata:
http://www.language-archives.org/OLAC/metadata.html

CMDI component metadata:
http://www.clarin.eu/cmdi

Guideline for the cost-benefit analysis for corpora curation (german draft):
http://www.corpora.uni-hamburg.de/documents/leitfaden_draft.pdf

EXMARaLDA core metadata schema:
http://www.corpora.uni-hamburg.de/documents/HZSKcoremetadataset.pdf

EXMARaLDA Corpus Manager:
http://www.exmaralda.org/en_coma.html

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 repository is to serve as the repository of a CLARIN-D resource center of type B. The mission of CLARIN-D is to provide “linguistic data, tools and services in an integrated, interoperable and scalable infrastructure for the social sciences and humanities“ (http://de.clarin.eu/en/home-en.html). Therefore a repository in which data, tools and according metadata is archived on a long term basis has to be operated by such a resource center.

The "Satzung" (articles of association) of the HZSK include explicit references to the tasks of archiving as well as making language resources available and to the underlying methodology.

§2 of the Satzung states:

§2 Goals and Mission

1. *The HZSK promotes and coordinates computer based research and teaching in linguistics and related disciplines at the University of Hamburg. Its main aims are:*

a. Ensuring sustainability, i.e. long-term usability and availability of empirical digital linguistic data, created and used at the University of Hamburg for research and teaching purposes. (...)

The HZSK presents its activities on a regular basis, organizes workshops and training courses to introduce people to the underlying methodology.

References:

HZSK-Satzung:
https://corpora.uni-hamburg.de/pdf/Satzung.pdf

CLARIN-D B Centre requirements:
https://www.clarin.eu/content/checklist-clarin-b-centres

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:

The repository is part of the infrastructure of the Hamburg Centre for Language Corpora at Hamburg University, which, like all public German universities, is a corporation under public law ("Körperschaft des öffentlichen Rechts").

The access restrictions for language resources hosted by the HZSK are based on the CLARIN License Categories: CLARIN PUB, CLARIN ACA or CLARIN RES (see https://www.clarin.eu/content/license-categories and https://corpora.uni-hamburg.de/drupal/en/corpus-enquiries-licenses)

Due to privacy and IPR restrictions, access to most of the resources stored in the repository is restricted. For the purpose of making them accessible to data consumers, individual contracts implementing individual access restrictions with rights holders (in most cases data producers) are made for every single resource. Based on these contracts, one of four levels of access control applies for a resource
(see https://corpora.uni-hamburg.de/pdf/CorpusReleaseGuidelines.pdf for details).
All contracts used with the data consumers are based on one model contract and altered as requested by the rights holder.

The conditions of use are not repository-wide but corpus-specific and therefore published on each corpus description page and further specified in the contract/conditions of use signed by the data consumer.

All data is managed, stored and distributed with great care. Access to the repository is only possible for those corpora for which access has been granted as described above.

The means of distribution of sensitive data is at the discretion of the rights holder, who is also responsible for the consent of the participants in the data and for compliance with ethical codes of conduct and national and

international legal regulations. The avovementioned model contract for the conditions of use, contains a part where the data consumer consents to not reveal the identity of any participant, nor publish data or part thereof in a manner that would make the reconstruction of a person's identity possible. As a part of the curation process it is also possible to anonymize files if requested by the rights holder.

The HSZK staff has received training and guidelines to handle requests for access according to our four-level-access-system.

Personal data that is retrieved from data consumers in the framework of the abovementioned process or from users' home institutions via Shibboleth Login is processed according to the HZSK Privacy Policy: https://corpora.uni-hamburg.de/drupal/en/privacy-policy

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 runs on a virtual machine hosted by the *Computing Centre of the Hamburg University* (*Regionales Rechenzentrum, RRZ*).

Daily incremental backups allow for a restore of the whole server within a duration of 24 hours.

Bitstream-conservation is provided by the Computing Centre of the Hamburg University, for a timespan not less than 10 years.

Documentation (in German):

- Virtualization: https://www.rrz.uni-hamburg.de/services/virtuelle-server/daten/virtuelle-server-ordnung-rrz220307.pdf
- Backup: https://www.rrz.uni-hamburg.de/services/datenhaltung/backup.html
- Backup-Policy of the RRZ: https://www.rrz.uni-hamburg.de/services/datenhaltung/backup/daten/tsm-nutzungsbedingungen.pdf

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:

The digital assets of the HZSK (as well as the actual repository) are stored at the computing centre of the Hamburg University (Regionales Rechenzentrum, RRZ). See also guideline 6.

The RRZ commits itself to ensure bitstream-conservation for 10 years from the moment assets are first stored*.

All main corpus files in the HZSK repository are XML-based, allowing for easy conversion into other formats.

References:

[1] Requirements for the long-time archiving of database applications and repositories of the *Faculty of Humanities* at Hamburg University (in German):

https://www.gwiss.uni-hamburg.de/service/it-support/datenbanken-repositorien.html

* As long as the requirements in [1] are met, a long-term (10 years) archiving commitment is given. The HZSK-Repository meets the requirements in [1] by using a virtual machine for the repository. However, the requirements would also be met for the plain data in the repository (possibility to export the data, Unicode-encoding, xml-based modeling, documentation).

In practice, the ten years period starts from the moment when there is a change in the terms and regulations that is not met by the repository, or the HZSK goes out of service.

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:

While the goal of the curation process for the resources at the HZSK is always the same (a consistent high quality corpus of EXMARaLDA or other standardized best practice formats with CMDI metadata for ingestion into the repository), the process is highly individual depending on the source data.

The procedural documentation for archiving data is not published. As many steps are performed (semi)automatically, documentation of the curation process and the ingest into the repository is sometimes only available in the form of commented data processing programs or scripts.

The cost-benefit analysis leading to the decision on whether to perform the curation is based on a publicly available guideline (see answers to DSA guidelines 2 and 3).

We plan to further develop and standardize our workflows and will document these versions of our archiving procedures accordingly.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Developing standardized workflows is a good goal.

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 repository has contracts with the rights holder/data producer of each corpus. The contracts are all based on one model contract with details on access adapted according to the requirements of the rights holder. The repository will not allow any deposit of data without a signed agreement specifying the handling of the data and access to it in detail.

A copy of the model contract (in German) containing English comments can be downloaded here: http://corpora.uni-hamburg.de/pdf/DELA_HZSK_EN.pdf

There is a backup procedure for the virtual server hosting the repository. The stored version can be used in case of any severe issues with the current system (see also DSA guideline 6).

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:

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

Self-assessment statement:

Metadata for the corpora hosted at the HZSK can be (and is) harvested via OAI/PMH.

As an example see the DiK corpus in the VLO:

http://catalog.clarin.eu/vlo/record?q=dik&fq=collection:Hamburger+Zentrum+f%C3%BCr+Sprachkorpora+%28HZSK%29&fq=resource

The collected metadata is used in the back-end of web applications such as CLARIN's Virtual Language Observatory, which provide a central starting point when searching for resources in the CLARIN infrastructure.

To use the actual corpus data, users have to sign an end-user-agreement and will receive password-protected access to a corpus. All corpora are provided in different tool formats that are common in the research community dealing with spoken language corpora (EXMARaLDA, Folker, Praat, ELAN) as well as exchange and presentation formats (TEI, MS-Word, PDF).

The EXMARaLDA system provides the search and analysis tool EXAKT that allows deep search in EXMARaLDA corpora (transcription data, annotations and metadata).

A search endpoint for CLARINs federated content search is available, though due to access restrictions, most corpora can not be provided via this interface..

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, please refer to guideline 0.

The usage of PIDs is mandatory for resources in CLARIN, thus all resources added to the repository can be referenced using PIDs.

References:

The CLARIN Virtual Language Observatory:

http://catalog.clarin.eu/vlo/

EXMARaLDA search and analysis tool (EXAKT):
http://www.exmaralda.org/en/tool/exakt/

CLARIN PID-Service description:
https://www.clarin.eu/sites/default/files/pid-CLARIN-ShortGuide.pdf

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

Handle-System Implementation:
http://handle.gwdg.de:8080/pidservice/

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 HZSK repository uses Fedora Commons' ability to automatically generate checksums for ingested resources.

The repository will only allow manual versioning of corpora. The HZSK controls what is deposited and will release new versions if major changes such as major corrections or completions, further annotation layers etc. are made to the data. Previous versions remain available on request using their PIDs.

We provide means for the Nagios monitoring of the repository, consistent with the other CLARIN-D centers.

Reference:

CLARIN-D infrastructure status:
http://clarin-d.de/de/aktuelles/status-infrastruktur.html

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:

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

Self-assessment statement:

The repository does not allow any uncontrolled changes to deposited data. New versions will only be created and ingested manually by the HZSK staff when data changes require a new version. Data producers of the current corpora have deposited their data in the repository as an archive and did not wish to further develop their data. For future data collections, versioning issues will be discussed and individual strategies agreed upon when data producers sign the contract.

The repository maintains information on data provenance and versions ingested into the repository.

The Coma format is the part of the EXMARaLDA system used to manage metadata on corpora, communications/sessions, speakers and physical files of the corpus. Since all spoken corpora are based on a Coma file, the existing metadata is always included in the resources.

Since versioning is controlled and conducted by the HZSK, all changes to the data can be documented for each version of the data and since EXMARaLDA data is XML, more detailed information on differences between versions can be gained by using common XML editors or tools.

The repository does not allow anonymous depositing of resources. Before a corpus is added to the collection, the HZSK will meet with the rights holder in person to discuss the curation process and the details on the access to the deposited resources.

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:

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

Self-assessment statement:

The repository complies with the OAIS reference model's tasks and functions.

The repository builds on the Fedora Commons software, which is compliant with the Reference Model for an Open Archival Information System (OAIS) due to its ability to ingest and disseminate Submission Information Packages (SIPS) and Dissemination Information Packages (DIPS) in standard container formats.

The data consumer has direct access to the archived objects via the web, provided that access requirements have been met.

The repository is part of the CLARIN infrastructure and will fulfill current and future requirements decided on by the CLARIN board.

References:

- 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/

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:

The End User Licence(s) used with data consumers depend on the rights holders' requirements. If desired, right holders can require data consumers to sign contracts specifying the conditions of use for a particular resource in detail as defined by the rights holder.

Some resources are free to use for academic and teaching purposes after registering, and these restrictions are recognized by accessing the resources.

For an example please refer to the license agreement for the 'Hamburg Corpus of Argentinean Spanish (HaCASpa)':
https://corpora.uni-hamburg.de/sfb538/h9_terms_of_use.pdf

In case of misuse, the user is denied further access to the repository. Further legal measures remain reserved to the data depositors.

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 selected resource, as soon as he agrees to the license agreement of that resource.

In case of misuse, the user is denied further access to the repository. Further legal measures remain reserved to the data depositors.

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:

For all corpora, specific restrictions are given by their respective license.

According to our four level access guidelines, explicit statements might need to be made by the data consumer about the usage of the data before he/she gets access. The depositor then decides on whether access is granted or not.

In case of misuse, the user is denied further access to the repository. Further legal measures remain reserved to the data depositors.

References:

Levels of access restriction at the HZSK Repository:
https://corpora.uni-hamburg.de/pdf/CorpusReleaseGuidelines.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: