



Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository CLARIN Center BBAW complies with the guidelines version 2010 set by the Data Seal of Approval Board.

The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2010 on May 21, 2013.

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: 2010 | June 1, 2010
Guidelines Information Booklet: [DSA-booklet_2010.pdf](#)
All Guidelines Documentation: [Documentation](#)

Repository: CLARIN Center BBAW
Seal Acquiry Date: May. 21, 2013

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 8, 2015
	Guidelines version:	2014-2017 July 19, 2013
	Seal date:	May 21, 2013
	Guidelines version:	2010 June 1, 2010

This repository is owned by:

Berlin-Brandenburg Academy of Sciences and Humanities (BBAW)
Berlin-Brandenburg Academy of Sciences and Humanities (BBAW)
Jägerstr. 22-23 Zentrum Sprache 10117
Berlin Germany
Berlin
Germany

T +49 (0)30 20370 0
F +49 (0)30 20370 600
E clarin@bbaw.de
W <http://www.bbaw.de/>

Assessment

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

Minimum Required Statement of Compliance:

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

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository will include resources provided by CLARIN-D member institutions and other institutions and/or organizations that belong to the CLARIN-D extended community. The data in our repository contains sufficient information for others to assess the scientific and scholarly quality of the research data in compliance with disciplinary and ethical norms. We specifically rely on DFG ethical Codes of Conduct (e.g. laid down in the DFG Rules of Good Scientific Practice). Thus, our repository provides a quality assessment by which the data consumer can make some judgment about the level of trust or about the reputation of the depositor on the basis of the meta-information about the source institution/organization information associated with any given resource. Our repository does not (and cannot) systematically verify whether the data received have been collected according to these quality standards. Ethical rules

ALLEA (ALL European Academies) European Science Foundation, The European Code of Conduct for Research Integrity. http://www.allea.org/Content/ALLEA/Scientific%20Integrity/Code_Conduct_ResearchIntegrity.pdf

DFG, Rules of Good Scientific Practice
http://www.dfg.de/en/research_funding/legal_conditions/good_scientific_practice/index.html

BBAW, Richtlinien zur Sicherung guter wissenschaftlicher Praxis
<http://www.bbaw.de/die-akademie/aufgaben-und-ziele/sicherung-guter-wissenschaftlicher-praxis/RichtlinienundAusfuehrungsbestimmun>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository provides a list of accepted formats, including common multimedia-document formats as well as formats for binaries. For other file formats, we provide advice for conversion. Lists of recommended formats

CLARIN, standards recommendations. <http://www.clarin.eu/recommendations>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

3. The data producer provides the research 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.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

CMDI metadata (CLARIN link www.clarin.eu/cmdl) is uploaded or created during the archiving process. This step is required during the uploading process, since data without metadata is technically not accepted by the system. The front-end of the archiving system includes software to assist the depositor in creating valid CMDI metadata.

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.

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The mission of the repository is to ensure the availability and long-term preservation of german text corpora, lexical and other resources.

This mission is supported by the infrastructure of the Berlin-Brandenburg Academy of Sciences and Humanities and by the integration of the repository into the national and international CLARIN infrastructures.

As part of the CLARIN infrastructure, the repository is included by all promotional activities carried out at the national level of CLARIN-D as well as the European level of CLARIN.

see <http://fedora.deutschestextarchiv.de/mission>

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.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository is no legal entity in its own right. It is run by the Berlin-Brandenburg Academy of Sciences and Humanities which is an institution governed by public law. Deposits are handled in a case-by-case approach. There are individual contracts and different licences for each resource we have archived. The access to the items is also handled case-by-case, ranging from open access over restricted access requiring a contract to restricted access on-site. The depositors themselves are responsible for compliance with any legal regulations in the area where the data is collected. Where required by national regulations, the archive also signs contracts with national/regional institutions.

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.

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

Backups are performed when the data in the repository changes, and are stored in the form of disaster recoverable virtual machine images as well as file system and database dumps. The backups are copied to tape storage which is deposited in a locked safe in a separate fire safety zone of the building (in german: 'Brandschutzabschnitt') and are performed with open source software, so that they are recoverable also on a long-term basis.

For software backups, we dump databases to local storage, sync those dumps (via rsync software) and additionally local software daily to a another server. Weekly backups are performed to a Quantum LTO5 tape library via the backup software Amanda (see www.amanda.org), which decides independently when incremental and full dumps have to be made (but full dumps are done at least once per month). Amanda is open source software which is based on basic GNU backup software like tar, gzip and dump, which ensures the ability to recover backups also in the distant future.

On the other hand, the virtual machines are completely backed up as virtual machine image snapshots via Proxmox vzdump (see http://pve.proxmox.com/wiki/Backup_-_Restore_-_Live_Migration), which are then backed up to tape storage to ensure fast disaster recovery times and also live migration of virtual machines to another virtualization cluster node. Proxmox uses the open source kernel virtual machine (kvm) software internally, which again ensures the ability to recover or convert snapshots also in the distant future. The snapshots are performed prior to configuration updates on the machines.

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.

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

In addition to the measures mentioned under §6 above to ensure the preservation of the raw resource data, measures are taken to ensure the future interpretability of the data. The number of accepted file formats is limited, to make future conversions to other formats more feasible. Open (non-proprietary) file formats are used whenever possible. For textual resources, XML formats are used whenever possible, to ensure future interpretability of the files independent of the tool used to create them. Text is encoded in Unicode to ensure future interpretability.

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 its 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 can be hosted by other centres.

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.

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The online archive management tool Fedora Commons defines a workflow to a certain extent, because no resources can be archived without metadata being present. The depositor mainly decides what material is being archived; the archive only has technical requirements with regard to the file formats and encodings. The depositor determines who can access the material and is also responsible for protecting the privacy of any subjects appearing in the recordings or texts. Additionally quality checks of data and metadata including PID (Persistent Identifier) assignment are done by the repository software.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

We would hope that during the implementation process documentation is developed which can be referenced in future DSA submissions.

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.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

In general it is the BBAW policy to accept only resources that are available for scientific use (preferably under a Creative Commons License). All archived resources are available online, the access permissions are defined by the depositors. Crisis management is addressed on a technical level. Since a PID system is used in CLARIN, moving resources from one CLARIN resource center to another one is possible without affecting the validity of references (e.g. PID reference of a resource used in a research paper). Our setup consists of virtual machines which are implemented by a high-availability failover cluster.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. The data repository enables the users to utilize the research data and refer to them.

Minimum Required Statement of Compliance:

2. Theoretical: We have a theoretical concept.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository provides various ways of utilizing the archived data via online tools as well as by downloading the data in formats commonly used by the research communities. An advanced metadata search utility is provided, as well as a simple search tool for textual content. All metadata can be harvested via the OAI-PMH protocol. Unique persistent identifiers according to the Handle system are provided for each corpus and the each session within the corpora. Additionally, CLARIN provides search facilities like the VLO (<http://www.clarin.eu/vlo/>).

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.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The integrity of the data is ensured by the version control in the Fedora-Commons back-end by MD5 checksums. Checksum tests are done regularly, especially before performing backups. Metadata is a data stream within the digital object, and as such is version-controlled like object data. The availability of file, web, and application servers is monitored continuously. We consider all objects deposited in our repository as fixed and immutable. We create new digital objects for updates and keep the old versions in our repository. However, updates of metadata for existing resources are possible without considering the result to be a new version.

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.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository in principle makes the original deposited objects available in an unmodified way, if the objects are delivered in one of the accepted file types and encodings. New versions of archived resources can be deposited, in which case the old versions will be moved to a version archive. Different versions of the same resource are not compared; we assume the depositor has good reasons for depositing a newer version. A new version of a resource will get a new persistent identifier; the old version will keep the original persistent identifier. Metadata can change if the depositor or archivist sees the need for that, in the case of errors or missing information. Changes to the metadata are currently not logged. All archived objects are linked to their metadata descriptions and are organized in hierarchical (or multi-rooted) tree structures to indicate relationships between objects and sets of objects. The tree structures can change if the depositors decide that this is necessary. The identities of the depositors are checked by the repository staff when they hand over their data. Provenance metadata as to who made changes to the repository is currently only stored in log files and not shown to the data consumer.

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.

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

For metadata we rely on the group of emerging standards around CMDI (ISO-CD 24622-1). 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.

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.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

Most of the data in the repository have Creative Commons licenses applied to them. If the data consumer does not comply with the access regulations, the only measure that can be taken in practice is to deny him/her further access and to make the research community aware of the misuse. For some data sets, explicit permission from the depositor is needed. In that case a login is necessary.

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 higher education and scientific research 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.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

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 in the repository are used.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

If applicable, the data consumer is made aware of usage restrictions for the data to which she/he has received access. Generally, the usage restrictions are already described in the codes of conduct. For some data, explicit statements need to be made by the data consumer about the use of the data before he/she receives access. The depositor then decides whether or not access is granted. In case of misuse, the only thing that can be done in practice is to deny the user further access to the repository and to make the research community aware of the misuse.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: