



## **Implementation of the Data Seal of Approval**

The Data Seal of Approval board hereby confirms that the Trusted Digital repository CLARIND-UDS 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 17, 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:	<a href="#">DSA-booklet_2014-2017.pdf</a>
All Guidelines Documentation:	<a href="#">Documentation</a>
Repository:	CLARIND-UDS
Seal Acquiry Date:	Jun. 17, 2015
For the latest version of the awarded DSA for this repository please visit our website:	<a href="http://assessment.datasealofapproval.org/seals/">http://assessment.datasealofapproval.org/seals/</a>
Previously Acquired Seals:	Seal date: March 27, 2013 Guidelines version: 2010   June 1, 2010
This repository is owned by:	<b>Department 4.6, Applied Linguistics, Translation &amp; Interpreting, Universität des Saarlandes</b> Campus A 2.2 66123 Saarbrücken Germany  T +49 (0)681/302-7007 F +49 (0)681 302 7007 E a.ziegler@mx.uni-saarland.de W <a href="http://fr46.uni-saarland.de/index.php">http://fr46.uni-saarland.de/index.php</a>

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

# Assessment

## 0. Repository Context

### Applicant Entry

*Self-assessment statement:*

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

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

Within CLARIN-D this resource centre is a certified centre of type B (<https://www.clarin.eu/content/checklist-clarin-b-centres>). 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. A list of centre requirements can be found under <https://www.clarin.eu/node/3542>.

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). 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."

#### Data Seal of Approval Board

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

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

The repository makes use of the server virtualisation (<http://www.hiz-saarland.de/dienste/basisdienste/server-virtualisierung/>) and of the backup facilities (<http://www.hiz-saarland.de/dienste/basisdienste/zentrale-datensicherung/>) offered by the HIZ. The HIZ is the joint IT provider of Saarland University (The archive's hosting institution) and of the University of Applied Sciences of the Saarland (HTW Saar). This outsource partner offers relevant functionality for guideline 6: "The data repository applies documented processes and procedures for managing data storage."

### **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:*

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

*Self-assessment statement:*

The repository will include resources provided by CLARIN-D related institutions and other institutions and/or organizations that belong to the CLARIN-D extended community. The data in our repository contains sufficient information for others to assess the scientific and scholarly quality of the research data in compliance with disciplinary and ethical norms. We specifically rely on DFG ethical Codes of Conduct. Thus, our repository provides a quality assessment in that 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 that is related to each resource. Our repository does not (and cannot) systematically verify whether the data received are collected according to these quality standards.

We provide some guidance to depositors in terms of describing the full package of information that should be deposited to facilitate assessment at the repository's webpage:

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

Ethical rules

ALLEA (ALL European Academies) European Science Foundation, The European Code of Conduct for [http://www.allea.org/Content/ALLEA/Scientific%20Integrity/Code\\_Conduct\\_ResearchIntegrity.pdf](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/principles\\_dfg\\_funding/good\\_scientific\\_practice/index](http://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index).

Universität des Saarlandes, Richtlinie zu wissenschaftlichem Fehlverhalten <http://www.uni-saarland.de/campus/service-und-kultur/dienstleistungen-der-verwaltung/persona>

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## **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:*

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

Lists of recommended formats

CLARIND-UDS repository accepted formats, <http://fedora.clarin-d.uni-saarland.de/ressources/A>

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

—

### 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 is required to produce metadata accomplishing the formats provided by the repository. The repository requests metadata according to the Dublin Core standard. In the process of ingest, these metadata are semi-automatically converted to CMDI by means of an XSLT template and enriched with additional information, e.g., persistent identifiers (PIDs).

Dublin Core: <http://dublincore.org/>

CLARIN FAQ about Metadata: <http://www.clarin.eu/faq-page/267>

CMDI: <http://www.clarin.eu/content/component-metadata>

Conversion procedure from Dublin Core to CMDI: <http://www.clarin.eu/faq/how-can-i-convert-my>

### **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:*

We have an explicit mission to archive language resources especially multilingual corpora (parallel, comparable) and corpora including specific registers, both collected by associated researchers as well as researchers who are not affiliated with us. The mission goes together with the official possibility to store full copies of resources at Universität des Saarlandes. We are working together with the Hochschul-IT-Zentrum of Universität des Saarlandes to ensure long-term preservation. We have also established contact with the Saarländische Universitäts und Landesbibliothek in this regard.

As part of the CLARIN infrastructure, the repository is included in all promotional activities CLARIN-D as well as the European level of CLARIN.

Links:

<http://www.clarin.eu/content/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.

### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

The UdS CLARIN-D centre is not a legal entity of its own. It is part of Universität des Saarlandes, which is a legal entity. Deposits are handled in a case-by-case approach. There are individual contracts and different licences for each resource we have archived. The access to the items is also handled case-by-case, ranging from open access over restricted access requiring a contract to restricted access onsite.

The depositors themselves are responsible for compliance with any legal regulations in the area where the data is collected.

Where required by national regulations, the archive also signs contracts with national/regional institutions.

All ethical issues are dealt with by using the endorsed Codes of Conduct, read section 1 for more information.

We provide information in the repository's website about the applicable Terms of Use of the repository.

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

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

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## 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 highly available virtual servers hosted by the Hochschul-IT-Zentrum which provide a backup service including incremental backups on a daily basis as well as regular full and level backups using EMC NetWorker. Backups are written to hard disks and additionally to tapes, which are stored for three months. A data recovery from the backup tapes is possible using the EMC Network client. The repository makes use of checksums to verify the integrity of the data.

Documentation (in German)

Virtualisation <http://www.hiz-saarland.de/dienste/basisdienste/server-virtualisierung/>

Backup <http://www.hiz-saarland.de/dienste/basisdienste/zentrale-datensicherung/>

### 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 is encoded in Unicode to ensure future interpretability.

Before ingest, we do the following checks:

- Full validation of the metadata against their respective schemes (oai\_dc, CMDI)
  
  
  
  
  
  
  
  
  
  
- Check that provided XML data are well-formed (at least)

- Spot-check of character encoding

## **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:*

A minimal workflow for the ingestion procedure is defined by the archive management tool Fedora Commons, such as that no resource can be archived without metadata and that the resource has to conform to certain file formats and encodings. The responsibilities of the depositor are

to decide what kind of material is being archived,

- 

to assure that the material follows the technical criteria required by the repository,

- 

to decide who may access the material,

-

to protect the privacy of any subjects appearing in the recordings or texts

- 

There is an internal documentation on the preparation of resources and the corresponding metadata. For the time being there is no need to make these publically available as we do not intend to implement an automatic ingestion process.

A formal curation policy has not yet been developed. This will be done as soon as we have a real use case, where such a policy is required. We expect our resources to follow individual problems in this case. The data depositors grant the repository the licence to convert the submitted data to other formats.

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

## **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:*

All archived resources are available online, the access permissions are defined by the data producers/depositors themselves.

The crisis management plan relies on the technical solution described in section 6 of these guidelines.

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

### **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:*

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

Search facilities over metadata are available at our repository [1]; but a much more user-friendly search over our metadata is provided by the Virtual Language Observatory (VLO) [2]. Since we cooperate with the VLO within the framework of the CLARIN-D project, we don't plan any improvement of our local search interface.

Harvesting of metadata is implemented via OAI-PMH, which collects CMDI-metadata from all repositories run by CLARIN centres. The collected metadata are used in the back-end of web applications such as the VLO. Our OAI provider [3] offers such metadata for OAI-PMH harvesting in two formats: Dublin Core and CMDI.

The repository itself does not offer a persistent identifier service on its own but makes use of a common CLARIN PID [4] service based on the handle system [5]. We register handles from the handle service as persistent and resolvable identifiers for our resources.

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

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

## References

1. Search facility at Uds CLARIN-D Centre repository: <http://fedora.clarin-d.uni-saarland.de>
2. Uds CLARIN-D browsing facette at VLO:  
<http://catalog.clarin.eu/vlo/?fq=collection:Universität+des+Saarlandes+CLARIN-D-Zentrum,+Sa>
3. Uds OAI provider: <http://fedora.clarin-d.uni-saarland.de/oaiprovider/?verb=Identify>
4. CLARIN's PID short guide: <https://www.clarin.eu/sites/default/files/pid-CLARIN-ShortGuide>
5. Handle system: <http://www.handle.net>

## 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:*

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

*Self-assessment statement:*

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. The new version of a resource will contain a pointer to the older versions in its metadata.

We calculate MD5 and SHA1 checksums for the stored objects, and we check them on a regular basis.

### **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 in principle makes the original deposited objects available in an unmodified way, if the objects were in one of the accepted file types and encodings. In the case of changes by the data producer, the repository creates a new digital object with a new persistent identifier. In the case that the repository has to change the data, e.g., because a file format becomes obsolete and superseded, the original data are kept.

The repository only accepts works from the original data producers, who are acknowledged as such by means of the "dc:creator" or "creator" elements, in Dublin Core or CMDI metadata respectively.

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

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

Example

CMDI metadata record for the GRUG parallel treebank as delivered by the OAI provider <http://fedora.clarin-d.uni-saarland.de/oaiprovider/?verb=GetRecord&metadataPrefix=c>

—

### **Reviewer Entry**

#### **Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

*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 [1]. Besides, the repository is powered by Fedora Commons software, which is compliant with the Reference Model for an Open Archival Information System (OAIS) due to its ability to ingest and disseminate Submission Information Packages (SIPS) and Dissemination Information Packages (DIPS) in standard container formats.

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

A structure diagram of the repository is found under <http://fedora.clarin-d.uni-saarland.de/>

References

1. Reference Model for an Open Archival Information System (OAIS), Recommended Practice, CCSDS 650.0-M-2 (Magenta Book) Issue 2, June 2012 <http://public.ccsds.org/publications/archive/650x0m2.pdf>

#### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## 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:*

For open access data (currently in the minority on the repository) the terms of use apply.

Most of the data in the repository is protected; an account is necessary to get access to the data. For some data sets, explicit permission from the depositor is needed. For a large part of the data, the data consumer needs to agree with a code of conduct, which also contains licensing terms. Details are given on the landing page of the respective resources.

If the data consumer does not comply with the access regulations, the only thing that can be practically done is to deny him/her further access and to make the research community aware of the misuse.

Documents

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

-

## 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 ALLEA 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 deposited data are used.

Documentation

For codes of conduct endorsed by the repository, see section 1 of this document.

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

—

**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:*

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

References

See section 1 of this document for codes of conduct endorsed by the repository

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

—

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*