



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

The Data Seal of Approval board hereby confirms that the Trusted Digital repository The Clarin centre at the University of Copenhagen 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 January 9, 2014.

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:	The Clarin centre at the University of Copenhagen
Seal Acquiry Date:	Jan. 09, 2014
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:	None
This repository is owned by:	<ul style="list-style-type: none"><li>• <b>University of Copenhagen</b> Copenhagen Denmark  T +45 35329090 E <a href="mailto:cst@hum.ku.dk">cst@hum.ku.dk</a> W <a href="http://info.clarin.dk/">http://info.clarin.dk/</a></li></ul>

# Assessment

## 0. Repository Context

### Applicant Entry

*Self-assessment statement:*

The CLARIN Centre at the University of Copenhagen, Denmark, hosts and manages a data repository, which is a digital research infrastructure for humanities and social sciences driven by international quality standards and financed with public national funding, through the national infrastructure collaboration DIGHUMLAB. Denmark is one of the founding members of the European Research Infrastructure Consortium, CLARIN, and the purpose of the repository is to be the Danish node in the European CLARIN-ERIC, and thus provide easy and sustainable access for scholars in the humanities and social sciences to digital language data (in written, spoken, video or multimodal form) and advanced tools to discover, explore, exploit, annotate, analyse or combine them, independent of where they are located.

The CLARIN Centre at the University of Copenhagen encourages data owners and producers to deposit data and their corresponding research material (documents and annotations) in the repository and provides data management consultation and support in connection with the deposit.

CLARIN-ERIC members are bound by a set of statutes and a membership agreement, which obliges each member to grant access to content to the other members' institutions through a federated catalogue and allows them to generate preservation copies.

The CLARIN Centre at the University of Copenhagen <http://info.clarin.dk/>

CLARIN-ERIC <http://www.clarin.eu/>

DIGHUMLAB, Digitalt Humaniora Laboratorium: <http://dighumlab.dk/>

### 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)

# **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:*

A data depositor is granted permission to deposit data through his or her CLARIN compliant, single sign-on authentication system. In Denmark, all Danish research institutions have the option to connect through the WAYF identity federation. WAYF is supported by the Danish Research Network (Forskningsnettet) - a high-speed network connecting Danish universities and research institutions.

Data has to be prepared such that it complies, technically, with the validation requirements of the platform, and it has to be available on a web address such that it can be harvested by the platform. It also has to be accompanied by relevant metadata to enable users to assess the suitability and quality of the data (data type (e.g. text, audio, video), producer, language, year, size, domain are just a few examples). The preparation measures, including the validation requirements are specified on the platform webpage.

Before data can be deposited, a Deposition License Agreement has to be in place between the Depositor (the entity or person who owns the data or holds the copyright to the data) and the Repository (who will include the data and distribute it). Three different sets of conditions are offered: PUB (to give public access to the data), ACA (to give access only for research purposes), and RES (to give restricted access, to be specified in the agreement). When a user wants to access and use the data he or she must first accept the license conditions for use following from the chosen deposition license. The data depositor is responsible for the data's adherence to relevant legal requirements and ethical norms and standard in the discipline in question.

WAYF – Where Are You From <http://www.wayf.dk/>

Guide for depositors <http://info.clarin.dk/deponer-resurser/vejledning/>

Available licenses: <http://info.clarin.dk/overblik/licenser/>

**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 has a list of accepted file formats. See <http://info.clarin.dk/deponer-resurser/valideringskrav/>. Files are stored in their original file formats in the repository.

When data is uploaded, certain requirements must be observed as described in <http://info.clarin.dk/en/deponer-resurser/> and <http://info.clarin.dk/deponer-resurser/vejledning/>:

Before depositing data in the repository, the following requirements must be fulfilled:

- All data must be valid, as described in <http://info.clarin.dk/deponer-resurser/valideringskrav/>
- All data must be accessible at a WEB-address from where the repository can obtain it.
- All data must be packed in .zip or .tar.gz archives, not much bigger than 250 MB per archive.
- Relation files shall not be compressed and have a comma separated format
- Directory structure shall not have semantic meaning, because all files are unpacked in the same directory.

**Data Seal of Approval Board**

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

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

- File names may only contain the characters [a-z][A-Z][1-9] and '-' and '\_'.
- Optionally, the depositor can enter the MD5 checksum of the uploaded data.

At ingest metadata and file formats are validated.

### **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:*

We only permit deposits of data if sufficient metadata is available.

The Data Deposit Form requests that data producers provide all metadata necessary to interpret the data prior to data ingest. The metadata standards used by the repository are detailed in <http://info.clarin.dk/en/deponer-resurser/validationrequirements/>. Below is a summary of this page:

All XML-files, including metadata files, must be utf-8 encoded.

The metadata that has to be supplied together with data shall be valid with respect to different rng-schemas:

- Text resources in TEIP5DKCLARIN format (with or without accompanying pictures or illustrations) have metadata in accordance with <https://www.clarin.dk/schemas/tei/TEIDKCLARIN.rng>.
- Text resources in TEIP5 format have metadata in accordance with <https://www.clarin.dk/schemas/tei/TEIDKCLARIN-plain-tei.rng>. We recommend that texts be delivered in TEIP5DKCLARIN format.
- Text annotations shall be in TEIP5DKCLARIN\_ANNO format, in accordance with the schema [https://www.clarin.dk/schemas/tei/TEIDKCLARIN\\_ANNO.rng](https://www.clarin.dk/schemas/tei/TEIDKCLARIN_ANNO.rng).

**Data Seal of Approval Board**

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

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



- Lexicon resources shall contain one and only one metadata file, which shall be in accordance with the XML schema [https://www.clarin.dk/schemas/tei/TEIDKCLARIN\\_LEX.rng](https://www.clarin.dk/schemas/tei/TEIDKCLARIN_LEX.rng). The resource itself must be packed in one compressed archive (zip, tar.gz).
- Sound, video and media annotations can be deposited as IMDI session, described by an IMDI file. The IMDI file shall be in accordance with the IMDI-standard [http://www.mpi.nl/IMDI/schemas/xsd/IMDI\\_3.0.xsd](http://www.mpi.nl/IMDI/schemas/xsd/IMDI_3.0.xsd). Attached IMDI WrittenResource files must be UTF-8 encoded.
- Tools must be deposited with metadata that is in accordance with [https://www.clarin.dk/schemas/cmd/DKCLARIN\\_TOOL.rng](https://www.clarin.dk/schemas/cmd/DKCLARIN_TOOL.rng). Metadata in accordance with this schema can be generated automatically by registration of the tool as an integrated tool.
- Resources that are not of the types mentioned above, can be deposited as ‘data’. Data resources have to be provided together with metadata that is in accordance with the schema [https://www.clarin.dk/schemas/cmd/DKCLARIN\\_DATA.rng](https://www.clarin.dk/schemas/cmd/DKCLARIN_DATA.rng). The resource itself must be packed in one compressed archive (zip, tar.gz).

Documents describing how to create metadata are available at <https://www.clarin.dk/documentation/>

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

Thank you for the clarification!

#### **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 be the Danish node in the European CLARIN-ERIC, and thus provide easy and sustainable access for scholars in the humanities and social sciences to digital language data (in written, spoken, video or multimodal form) and advanced tools to discover, explore, exploit, annotate, analyse or combine them, independent of where they are located. Digital archiving and long-term preservation and easy and sustainable digital access to data resources and tools will offer new possibilities for the scholars to develop new research methods and ask new types of research questions, and it will support and enhance their participation in collaborative international research.

The CLARIN Centre at the University of Copenhagen promulgates all aspects of this mission through publications, conference attendance, organization of PhD courses and other courses and workshops, e.g. in collaboration with other Danish Universities through the national digital humanities collaboration, DIGHUMLAB\*. Employees at the CLARIN Centre at the University of Copenhagen are active participants in both national and international fora that aim to establish standards for best practices and infrastructures for digital archiving.

Mission statement: <http://info.clarin.dk/en/overblik/datamanagement/>

*\* The CLARIN Centre at the University of Copenhagen is financed with public national funding through the national infrastructure collaboration DIGHUMLAB (<http://dighumlab.dk/>).*

#### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

Thank you for the clarification!

**Data Seal of Approval Board**

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

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

## **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 CLARIN Centre at the University of Copenhagen is hosted by the Faculty of Humanities, which falls under the governance of the University of Copenhagen. Prior to depositing data, the data depositor is required to accept a Deposition License Agreement, which states that the data depositor, as owner or copyright holder, is responsible for the data's compliance with legal and ethical requirements. All users of material from the CLARIN-DK repository must also agree to the terms and conditions stated in the User License associated with the data, which includes statements relating to the infringement of copyright and intellectual property.

As part of the University of Copenhagen, the CLARIN Centre at the University of Copenhagen is also bound by University and Administrative Policies as well as the Information Technology Services Information Security Policy, which protects against the disclosure of sensitive information.

- Is the repository a legal entity? Please describe its legal/organisational status.

The repository is not a legal entity on its own. The repository belongs to the legal entity University of Copenhagen which is a public institution, in Denmark.

- Does the repository use standard/template/model contract(s) with data producers?

Data owners, data producers and other potential data depositors have to sign a standard contract with the repository.

- Does the repository use standard/template/model contract(s) with data consumers?

Data consumers have to accept standard licence conditions before they are allowed to download data.

- Are the repository's conditions or terms of use published?

Yes, see <http://info.clarin.dk/overblik/bestemmelser/>

- Are there measures in place if the conditions are not complied with?

If the conditions are not complied with, the measures in place are the legal consequences that may apply, according to national and international laws.

- Does the repository ensure knowledge of and compliance with national and international laws? How?

The licence conditions under which a data provider deposits the data, and the licence conditions under which a user downloads the data are rather comprehensive w.r.t. describing what the licence allows and what it does not allow. Users have to explicitly accept the licence conditions for each specific piece of data before it can be downloaded. The depositor agreement templates and the licence templates are prepared in collaboration with the legal department of the university administration.

- Are special procedures applied to manage data with disclosure risk?

The repository does not store research data with disclosure risk ('personhenførbare' data). Video and sound recordings, for instance, can only be stored if the depositor has secured permission from the persons involved.

- Are data with disclosure risk stored appropriately to limit access?

The repository does not store research data with disclosure risk.

- Are data with disclosure risk distributed under appropriate conditions?

The repository does not store research data with disclosure risk

- Are procedures in place to review disclosure risk in data and to take the necessary steps to either anonymise files or to provide access in a secure way?

The repository advises data depositors not to store research data with disclosure risk.

- Are staff trained in the management of data with disclosure risk

The repository does not store research data with disclosure risk.

## **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 stores its resources on servers owned by the University of Copenhagen. We don't have archival copies, but rely on backup/restore for the preservation of data. Backup is performed on a daily basis, using a TSM backup setup that ensures a full backup of all repository files and services and the backup is stored off-site. The backup is monitored both by the IT department and two persons from the technical management team of the repository, by inspecting the backup status messages, and by an automatic check of the runtime of the backup processes. The storage of backup data is in the hands of a third party with which the University of Copenhagen has a general agreement for storage of backup data. The backup process is documented in a wiki for the technical management team

The repository system, the technical requirements and configuration of the repository are also described in the wiki for the technical management team. The wiki is regularly backed up to another server.

The servers of the repository are monitored each 10 min from an external server for http-access, process load and memory usage, and two persons from the technical management team are notified in case of an alert.

The repository is using the Fedora Commons Repository Software with the eSciDoc (The Open Source e-Research Environment Processes) extension.

Information to users including processes to ingest new resources and access data are documented on [info.clarin.dk](http://info.clarin.dk).

Metadata information and deposited resources cannot be changed by the users. Updates of metadata and relations between resources can only be performed by the technical management team. Content is not changed after deposit. The repository enforces different levels of access to content according to licensing restrictions and data producer preferences. These levels restrict access to specified individuals or groups of authorized users.

**Data Seal of Approval Board**

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

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



The principles for backup can be found at <http://info.clarin.dk/en/overblik/datamanagement/>

## **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:*

Some measures are taken to enhance the chance of future interpretability of the data.

CLARIN-DK's primary techniques to address file format obsolescence is normalization before ingest. The number of accepted file formats is limited, to make future conversions to other formats more feasible. We are using open or industrial de facto standard formats, such as a diversity of XML formats and TIFF.

We have implemented a persistent identifier service using Handle System handles and are right now assigning PID's to all objects in our repository and expect to finish this process in 2013.

We are currently making plans for future service maintenance.

### **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 data of the repository are collected as part of a former project which had as the aim to collect resources and prepare metadata for them. New data can be added by researchers at a Danish research institution. The repository does not handle sensitive information that contains micro data that exposes the integrity of human subjects. The repository team goes into a dialogue with the researchers who want to deposit data, to clarify if the data are relevant for the purpose of the repository, and to guide about metadata creation, as well as correction of validation errors. See <http://info.clarin.dk/en/deponer-resurser/>.

Producers of data can read about the handling of their data here: See <http://info.clarin.dk/en/overblik/datamanagement/>

A team of employees are available for user assistance and repository maintenance. The team has expertise in the areas of validation of data, data processing, repository management, using of involved standards, and metadata formats. The employees are educated in the fields of natural language processing, linguistics, software development.

The types of data within the repository are text resources, text annotations, sound resources, video resources, annotations of sound and video, lexica and tools. These are listed here: <http://info.clarin.dk/kom-godt-i-gang/vis-resurser/>.

The repository implements an explicitly defined workflow described on our website in several pages.

The workflow consists of

## Deposition

Guidance: see <http://info.clarin.dk/deponer-resurser/vejledning/>

- You first need to obtain a depositor role. Contact address [admin@clarin.dk](mailto:admin@clarin.dk).
- Ensure that your data is valid. Resources need to be prepared in one of the formats that the repository accepts.
- Package data
- Make data available at a web-location from where the deposit service can download the data
- Log in (WAYF).
- Choose resource access (Public, Academic or Restricted)

•

•

## Archiving

- Description: see <http://info.clarin.dk/en/overblik/datamanagement/>

•

## Access

- Who can use the repository, see <http://info.clarin.dk/overblik/hvem/>
- How to search: <http://info.clarin.dk/soeg-resurser/soeg/> (CLARIN-DK), <http://info.clarin.dk/soeg-resurser/vlo/> (VLO)
- Overview of Licences for viewing and downloading data: <http://info.clarin.dk/overblik/licenses/>

•

## Reviewer Entry

*Accept or send back to applicant for modification:*

**Data Seal of Approval Board**

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

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

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 repository assumes responsibility as stated in the Deposition License Agreement:

<https://clarin.dk/clarindk/download-proxy.jsp?license=depositpublic>,

<https://clarin.dk/clarindk/download-proxy.jsp?license=depositacademic>.

Citation from these licences:

#### *4. The Repository*

*The Repository shall ensure, to the best of its ability and resources that the deposited Content is archived in a sustainable manner and remains legible and accessible.*

**Data Seal of Approval Board**

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

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

*The Repository shall, as far as possible, preserve Content unchanged in its original digital format, taking account of current technology and the costs of implementation. The Repository has the right to modify the format and/or functionality of Content if this is necessary in order to facilitate the digital sustainability, distribution or re-use of Content.*

*If the access categories "Restricted Access" or "Academic Access", as specified at the end of this Agreement, are selected, the Repository shall, to the best of its ability and resources, ensure that effective technical and other measures are in place to prevent unauthorised third parties from gaining access to and/or consulting the Content or substantial parts thereof.*

End of citation.

We have a contingency plan for computer services.

### **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:*

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. A metadata search utility is provided, as well as a deep search tool for textual content. The data repository enables the users to discover and use the data and refer to them in a persistent way. To enhance discoverability of content, the repository supports OAI harvesting; the repository's content is harvested and replicated by VLO using OAI-PMH.

For every item stored in the repository, a unique persistent identifier (Handle) is automatically generated and included in the metadata. The repository's Handle prefix is 11221. (In progress)

### **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:*

The repository utilizes MD5 checksums to verify data integrity.

The integrity of the data and metadata is monitored as it follows the work flow in a controlled environment. Once the metadata is in the repository, the access is available only for viewing and downloading; no user is able to modify its content.

We support versions of metadata and PID's for each version.

The depositing system ensures that resources are validated in compliance with established policies.

### **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 stores data but does not process or alter it in any way. All objects in the repository have metadata. No links are made from the repository to other data sets. When new versions are stored in the repository, previous versions are maintained by a version control system built into the repository back end. A new version of a resource will get a new persistent identifier; the old version will keep the original persistent identifier. The repository does not compare versions in any way, and there are no plans to implement that, because each version is regarded as an independent resource in its own right. Each data stream has an associated checksum, which is automatically computed by the repository. Metadata are updated by the repository staff if either the data depositor or the archive content manager sees a need for it. A limited number of authorized and trained data managers ensure the safety of both data and repository. The repository relies on WAYF to check the identity of depositors. The metadata of all objects in the repository contain provenance data.

### **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 repository develops plans for infrastructure development by participation in CLARIN ERIC (<http://clarin.eu/>). We have an overall plan for the period until mid-2017 and a detailed implementation plan for the current year.

The repository aims to be as conformant to the OAIS reference model's tasks and function as possible. However, due to the complexity of the OAIS reference model, the repository cannot guarantee that all tasks and functions will be implemented.

**Ingest:** The repository uses the national identity federation WAYF.dk to support single identity and single sign-on operation based on SAML2.0 and trust declarations. Those users that are defined as researchers by their home institutions can ingest a Submission Information Package (SIP) to the repository. To submit a SIP the user selects and accepts the licence under which the data will be deposited. The SIP has to fulfil a number of requirements to be accepted. The metadata format and content included in the SIP has to comply with a defined list of standards for which there are defined xml schemas, that will be used to evaluate the metadata contained in the SIP. There are also restrictions on the formats for content in the SIP. After validation of the SIP, the deposit service handles the transformation of the SIP to the Archival Information Package (AIP). The procedures for checking the SIP before creating the AIP will be extended in the future.

**Archival Storage:** The repository is using the Fedora Commons Repository Software with the eSciDoc (The Open Source e-Research Environment Processes) extension. Backup of the repository is carried out on a daily basis, and backup storage is done on an external location.

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

**Preservation Planning:** The metadata contained in the SIP is preserved unchanged. It is an important issue that the data should be preserved, but the procedures are not yet defined. This work is in progress.

**Administration:** Contract agreements with the Data Producer are created when the SIP's are ingested. Administration staff includes a content manager who is dedicated to issues about the content administration and validation.

**Access:** The Dissemination Information Packages and query responses are delivered to consumers, who have the rights to access the data. Metadata are publicly available, content data can require public, academic or restricted

access permissions. A user interface available on [clarin.dk](http://clarin.dk) allows the consumer to search metadata. The consumer can also online inspect some content types and download the content if the access requirements of the content have been met by the rights of the consumer. We do not handle sensitive information that contains micro data. The digital objects are in the process of being available for reading access via their Persistent Identifiers (PID) for authorized users, based on the national AAI infrastructure. The PIDs will be available in the metadata, which can be harvested via OAI-PMH (e.g. by the VLO <http://catalog.clarin.eu/vlo/>).

## **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 repository uses End User Licences with data consumers.

All metadata are openly accessible.

The repository restricts access to academic data by the requirement of a login from a WAYF affiliation (“Where Are You From”, the Danish national identity federation). Access to restricted data is restricted to the depositor of the data. These restrictions are clearly labelled in the repository and are enforced by WAYF mechanisms, which disallow restricted file access and download by unauthorized individuals.

Any user (whether or not logged in) who wants to download data from the repository, must agree to the terms of use as described in the licence pertaining to the data. These terms address issues related to the redistribution of content and copyright concerns. See

<https://clarin.dk/clarindk/download-proxy.jsp?license=downloadpublic>

<https://clarin.dk/clarindk/download-proxy.jsp?license=downloadacademic>

## Reviewer Entry

*Accept or send back to applicant for modification:*

**Data Seal of Approval Board**

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

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

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:*

The repository does not need to deal with any relevant codes of conduct. To comply with disciplinary and ethical norms, we have privacy policies <http://info.clarin.dk/overblik/privatlivspolitik/>, instructions for citations, as well as disclaimers for the use of the data <http://info.clarin.dk/overblik/bestemmelser/>.

The repository does not need to deal with codes of conduct specifically pertaining to protection of human subjects, since our data does not consist of micro data that exposes the integrity of human subjects.

If the requirements are not complied with, the measures in place are the legal consequences that may apply, according to national and international laws.

Users must agree to a public (<https://clarin.dk/clarindk/download-proxy.jsp?license=downloadpublic>) or academic (<https://clarin.dk/clarindk/download-proxy.jsp?license=downloadacademic>) licence agreement before they can download data from the repository. The terms therein address issues related to the redistribution of content and copyright concerns.

University of Copenhagen is the only legal entity involved, i.e. no other institutions or bodies are involved.

**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)



## **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:*

To proceed with any file download, users must agree to the terms of use in a licence. The data in the repository is of either public, academic or restricted character and the terms of use for other persons than the depositor of the data are specified in <https://clarin.dk/clarindk/download-proxy.jsp?license=downloadpublic> and in <https://clarin.dk/clarindk/download-proxy.jsp?license=downloadacademic>. Data with restricted access is only accessible by the depositor of the data.

If these licences are not complied with, the data user is subject to the legal consequences that may apply. At the time of this application, CLARIN-DK has not been alerted of any breach of the licences.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*