



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

The Data Seal of Approval board hereby confirms that the Trusted Digital repository CLARIN Center INL 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 23, 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:	CLARIN Center INL
Seal Acquiry Date:	Jun. 23, 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>INL</b></li><li>INL Matthias de Vrieshof 3 2311 BZ Leiden The Netherlands</li></ul> <p>T +31 (0)71 – 5141648 F +31 (0)71 - 5272115 E <a href="mailto:info@inl.nl">info@inl.nl</a> W <a href="http://www.inl.nl/">http://www.inl.nl/</a></p>

# Assessment

## 0. Repository Context

### Applicant Entry

*Self-assessment statement:*

The INL CLARIN Center repository gives access to language resources and tools from INL and other CLARIN members.

Data and tools can be found at: <https://portal.clarin.inl.nl/>

All data also has metadata in place in CMDI format. The metadata can be harvested by OAI/PMH: <http://repository.dev.clarin.inl.nl/oai/>.

All data and metadata have persistent identifiers based on handle.net with prefix 10032

### 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 data sets are supplied with all the information that is essential for sustainable data management and future use. A minimum set of metadata has to be provided by either the data producers or is extracted by INL from the data and documents provided by the data producers before the data can be archived.

Data producers are encouraged to supply additional data description documents or links to publications (using persistent identifiers) about the data. The publications are stored in the repository (given that permissions are granted), while the data description documents are archived in the OAI/PMH accessible data repository and accessible through the tools and applications.

Preliminary guidelines for deposition are available at <https://portal.clarin.inl.nl> (Information about deposition).

**Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

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

*Minimum Required Statement of Compliance:*

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

### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

We have a policy to provide data as much as possible in open standards (<http://hollandopen.nl/> and <http://www.clarin.eu/node/2320>). The data from other producers is accepted if it complies to these standards. It will be rejected if it does not comply.

### **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 depositor has to provide data with metadata in valid CMDI format. CMDI profiles used are published in <http://catalog.clarin.eu/ds/ComponentRegistry/#>. We can provide some assistance in creating the right profile and metadata. Creating from CMDI profile and valid metadata is the responsibility of the depositor.

#### **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 statement is available at <https://portal.clarin.inl.nl> (Information about deposition), which is also the portal mentioned in the statement. The CLARIN center is integrated with the INL and uses all of its facilities: website, newsletters, twitter and so on.

#### **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 not a legal entity on its own, but is part of the Instituut voor Nederlandse Lexicologie (Institute for Dutch Lexicology), which is a legal entity, an Institute according to Dutch law. It is controlled by a board. As umbrella organization functions the Dutch Language Union. Both INL and the Dutch Language Union are under the supervision of the Committee of the (relevant) Ministers of the Netherlands and Flanders.

The data in the repository is a combination of INL products and the results of projects in which the INL was involved. If data is owned by third parties like publishers, the project agreement makes explicit all obligations and restrictions on use.

The access to the data is restricted in several ways. Sometimes data is only accessible for the Clarin community, through the Clarin authorization and authentication mechanism. Sometimes data has open access (data regulated through web applications which allow querying for useful, but limited amounts of data). Moreover, all web applications carry a description of the terms of use. The general terms of use are available at <https://portal.clarin.inl.nl> (End User License Agreement).

As to contracts, please see also the documents that will be sent separately to the reviewer in relation to Item 9 of the assessment.

The INL repository does not contain data which involve disclosure risk other than copyright infringement.

Breaches of contract will be handled by the legal representative of the INL.



Compliance to national and international laws is checked periodically by the legal representative of the INL, who is specialized in IPR-issues, trademark rights and database rights.

Data are stored in a state-of-the art LAN-DMZ set-up, according to best practices, including INL's system of authentication and authorization (e.g. Active Directory).

As stated before, data is made available (searchable) primarily through web applications. Distribution of the data as such takes place, but is exceptional and is restricted to historical material without any privacy issues. It is realized by the standing INL organization, under the supervision of management that has received instructions from experts in consortia such as CLARIN-NL and from legal experts.

## **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 management is taken care of by the IT department of the INL and takes place according to extensive documentation and numerous Standard Operating Procedures and checklists, stored in an internal wiki.

The procedures include:

- Weekly full back-ups to tape. The full back-ups are followed by incremental back-ups on a daily basis.
- Quarterly full back-ups to tape, to be stored at another location (<http://backupned.nl/>). These tapes will eventually be reused, but one set of each year is retained for at least seven years. A restore can be carried out upon request.
- Installation of security patches and updates on a monthly basis.
- Daily and automated monitoring of systems and applications.

The information in the internal wiki is confidential. The INL internal Wiki contains a vast number of documents, concerning both ongoing projects and project results (applications, data). The tables of contents of the most relevant sections have been sent separately to the reviewer. These concern Clarin as an umbrella, software development and system management. The documents consist for the most part of documentation, standard operating procedures, checklists and best practices. Internal INL standards are applied also to external data and applications (metadata, security etc.).

## **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:*

As much as possible open file formats are used. Small volume conversions due to obsolescence of file formats will be handled. 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.

The INL has the capability to convert when needed and follows the guidelines published by CCSDS (<http://public.ccsds.org/publications/MagentaBooks.aspx>), in the '*Reference Model for an Open Archival Information System (OAIS)*' ('CCSDS 650.0-M-2').

For converting images and video we make an appeal to archives which are specialized in these kind of resources.

### **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 entire archival process from ingest through to access is or will be based on the guidelines published by CCSDS (<http://public.ccsds.org/publications/MagentaBooks.aspx>), in the *Reference Model for an Open Archival Information System (OAIS)* ('CCSDS 650.0-M-2').

The ingest procedures in use by data producers and staff are documented. Other processes, due to long-term preservation (data conversions, version control...) still need to be described in detail and are currently in their draft form.

As stated above, all data is made available through tools and applications on the web. Periodic tests of the software and therefore data take place, typically when a new version of Internet Explorer is released. The tests also involve a number of other browsers.

The documentation of the ingest procedures (in English) will be sent separately to the reviewer.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

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

*Minimum Required Statement of Compliance:*

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

### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

The data producer, i.e. the depositor will always remain the proprietor. INL does in fact get a copy of the data of which it must take good care, according to the terms of the license contract and the terms and conditions for use.

INL also makes copies, for example for the benefit of back-up and looks after them well.

In case of an emergency we are able to build up an entirely new database composed of all files we backed up and stored safely at another location. Quarterly and yearly tapes are stored at <http://backupned.nl/>

Preliminary guidelines for responsibilities are available at <https://portal.clarin.inl.nl/> (Information about deposition).

License agreements or contracts (e.g. with the heirs of a deceased scientist) vary from product to product and are not made public. Four examples will be sent separately to the reviewer:

- Contract between the Flemish Publishing Company and INL, concerning the delivery and use of digital material from the Flemish newspaper De Standaard.

- Contract between PCM Publishers and INL, concerning the delivery and use of digital material from the Dutch newspaper NRC Handelsblad.

**Data Seal of Approval Board**

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

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

- License agreement between INL and a second institute, concerning the consultation and use of the product  
Brieven als Buit/Letters as loot.

- License agreement between a data supplier and INL, concerning the consultation and use of textual material.

### **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:*

Research data are currently available in formats suggested, required and frequently used by the data consumers or in formats in which INL has the highest confidence with regard to durability (cf. <http://hollandopen.nl/> and <http://www.clarin.eu/node/2320>).

The resource browser (<https://portal.clarin.inl.nl/>) allows access to all data collections available. All visible metadata and some metadata extracted from data files are indexed and searchable. Queries can contain special operators, fieldnames, wildcards etc. and results can be refined using facets by the user (<http://catalog.clarin.eu/vlo/>).

All objects maintained by the INL CLARIN Center are assigned a unique identifier (<http://hdl.handle.net/>) with prefix 10032.

All metadata can be harvested via the OAI-PMH protocol.

### 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:*

To ensure the integrity of the data sets, for every deposited file a checksum (md5 type) is made which allows us to check the files for defects in later years.

Once deposited, files in data sets are never changed and only minor changes to the metadata are allowed. For example: correction of spelling, minor changes in documentation, additional documentation added, etc. Changes to the data themselves will be issued as a new version of the dataset and will obtain a new persistent identifier. These changes are only made in narrow collaboration with the producer of the dataset.

Preliminary guidelines for preservation are available at <https://portal.clarin.inl.nl/> (Information about deposition).

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

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

*Minimum Required Statement of Compliance:*

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

### Applicant Entry

*Statement of Compliance:*

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

*Self-assessment statement:*

We do not change data, but may add metadata if required. Data producers hand over the material to us and so far trust that we are looking after their material properly.

If applicable, we create collection-level objects which provides a context for the data within the collection.

The repository maintains links to other relevant materials (e.g. article, thesis, documentation, data elsewhere) and to metadata of measuring instruments etc. whenever applicable.

The repository doesn't compare the essential properties of different versions of the same file.

The unique identity of a depositor is ensured by the required login using the CLARIN Service Provider Federation for identification (<http://www.clarin.eu/content/service-provider-federation>).

Preliminary guidelines for authenticity control are available at <https://portal.clarin.inl.nl/> (Information about deposition).

### 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)

### **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 OAIS presents a functional model consisting of six functional entities. A number of interactions are possible between those entities. We will present a description of these entities within the INL CLARIN Center.

1. Ingest. This entity receives data from producers. Special tasks are: receiving data, performing quality assurance, checks on documentation, description and formats. Establish metadata and prepare for archiving and data management. Implications for INL CLARIN Center: There is a Standard Operating Procedure for ingest of data (acquisition) which includes all the tasks mentioned.

2. Archival Storage. This entity is responsible for the systematic storage, maintenance and retrieval of the data. It further performs routine checks on media quality (refresh if necessary), errors and disaster recovery capabilities. Implications for INL CLARIN Center: The INL CLARIN Center distinguishes two separate functions: First, data management, which is responsible for storage of the data, error detection and retrieval. Second, system management, which is responsible for media quality and recoverability.

3. Data Management. This entity is responsible for content integrity of the data. It sees that data and descriptive information is connected and is responsible for version management. Implications for INL CLARIN Center: These responsibilities are part of the previously mentioned data management function.

4. Administration. Oversees all archiving operations. Negotiates submission agreements with Producers. Establishes policies for maintenance, standards and hardware and software planning, customer support, etc. Implications for INL CLARIN Center: The head of INL CLARIN Center, together with management of INL is responsible for developing these policies.

5. Preservation planning. Evaluates the quality of the content and the quality of the service in context of the user community. Signals developments in technology and use patterns and provides policies to upgrade the archive

**Data Seal of Approval Board**

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

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

service accordingly. Also provides migration planning. Implications for INL CLARIN Center: The INL collaborates in projects with many national and international parties. The services of the INL CLARIN Center are continuously updated according to the needs of those collaborations. Furthermore, the INL CLARIN Center participates in spearheading projects like CLARIN and CLARIAH that intend to provide a digital infrastructure for research.

6. Access. This entity is responsible for the interaction with consumers. It provides information about the available products and is responsible for communication with consumers. Implications for INL CLARIN Center: Information about products is disseminated through a number of portals like OLAC, ELRA Universal Catalogue and CLARIN. This information points interested parties to our products. Questions can be directed to the service desk.

## **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:*

These matters are dealt with at <https://portal.clarin.inl.nl> (End User License Agreement).

All Open Access data stored in INL are freely accessible thus following the principles of Open Access.

Some data is restricted to Academic Use. The CLARIN Service Provider Federation environment is used for identification of the user.

### **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:*

These matters are dealt with at <https://portal.clarin.inl.nl> (End User License Agreement).

The general terms and conditions incorporate the principles of “Open Access” and The Netherlands Code of Conduct for Scientific Practice, see:

<http://www.vsnu.nl/Media-item-1/Code-of-conduct-for-scientific-practice-2004.htm>).

**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:*

These matters are dealt with at <https://portal.clarin.inl.nl> (End User License Agreement).

The general terms and conditions incorporate the terms and conditions of the license agreement.

In the metadata of individual datasets additional licenses can be specified (e.g. GNU, Creative Commons, etc.).

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*