



## Assessment Information

[CoreTrustSeal Requirements 2020–2023](#)

Repository: Clarin Center at the Arni Magnusson Institute for Icelandic Studies  
Website: <https://clarin.is>  
Certification period: 22 June 2023 - 21 June 2026  
Requirements version: CoreTrustSeal Requirements 2020-2022

This repository is owned by: **Arni Magnusson Institute for Icelandic Studies**

## CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

### Background Information

#### Repository Type

Please provide context for your repository. You can select one or multiple options.

#### Compliance level:

Not Applicable - 0

#### Response:

- Domain or subject-based repository
- Research project repository

#### Reviews

##### Reviewer 1:

#### Compliance level:

Not Applicable - 0

#### Comments:

##### Reviewer 2:

#### Compliance level:

Not Applicable - 0

#### Comments:

### Description of Repository

Provide a short overview of the repository.

#### Compliance level:

Not Applicable - 0

#### Response:

CLARIN-IS[1] is the Icelandic node of the European Common Language Resource and Technology Infrastructure, CLARIN-ERIC[2].

CLARIN is a distributed digital infrastructure, with participating institutes all over Europe, such as universities, research centres, libraries and public archives. All participating organisations have in common that they provide access to digital language data collections, digital tools to work with them, and expertise for researchers to work with them.

CLARIN's governance and coordination body is an ERIC, that is, a European Research Infrastructure Consortium. An ERIC is an international legal entity, established by the European Commission in 2009.

Iceland joined CLARIN-ERIC in February 2020, after having been an observer since November 2018.

The Ministry of Education, Science and Culture assigned The Arni Magnusson Institute for Icelandic Studies (AM)[3] the role of leading partner in the Icelandic CLARIN consortium[4].

AM is an independently funded academic research institution at the University of Iceland and hosts and manages the CLARIN-IS Repository [5].

The aim of the CLARIN-IS repository is to provide linguistic data, tools and services in an integrated and interoperable infrastructure for research in the humanities, social science and language technology.

[1] CLARIN-IS: <https://clarin.is/en>

[2] CLARIN-ERIC: <https://clarin.eu/>

[3] Arni Magnusson Institute for Icelandic Studies: <https://arnastofnun.is/en/about>

[4] CLARIN consortium: <https://clarin.is/en/participants/>

[5] CLARIN-IS Repository: <https://repository.clarin.is/repository/xmlui/>

## Reviews

### Reviewer 1:

#### Compliance level:

Not Applicable - 0

#### Comments:

### Reviewer 2:

#### Compliance level:

Not Applicable - 0

#### Comments:

## Designated Community

### Provide a clear definition of the Designated Community

#### Compliance level:

Not Applicable - 0

#### Response:

The designated community of users of the Clarín-IS repository is particularly the research community where language plays a central role.

The repository welcomes varieties of data including corpora, dictionaries, lexical resources, texts of many formats, multimedia files (sounds and videos) as well as tools and services that can further enhance research and growth in the humanitarian fields concerned with languages. The tool services can include simple Python scripts to more advanced pretrained translation models based on Artificial Intelligence.

Students, researchers and Software developers working in the fields of Humanities and Language technology who are interested in using language tools or analysing language data are the typical users of the repository assets.

With open licences[1] such as CC-BY the repository is also of value for institutes, companies and individuals developing applications in the language technology.

The community is encouraged to deposit their research material and suggests the use of open licences.

[1] Available Licences: <https://repository.clarin.is/repository/xmlui/page/licenses>

## Reviews

### Reviewer 1:

#### Compliance level:

Not Applicable - 0

#### Comments:

### Reviewer 2:

#### Compliance level:

Not Applicable - 0

#### Comments:

## Level of Curation

Select all relevant types of curation.

- Content distributed as deposited
- Basic curation – e.g., brief checking, addition of basic metadata or documentation
- Enhanced curation – e.g., conversion to new formats, enhancement of documentation
- Data-level curation – as above, but with additional editing of deposited data for accuracy

**Compliance level:**

Not Applicable - 0

**Response:**

- B. Basic curation – e.g. brief checking; addition of basic metadata or documentation

**Reviews**

**Reviewer 1:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Reviewer 2:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Level of Curation - explanation**

Please add the description for your Level(s) of Curation.

**Compliance level:**

Not Applicable - 0

**Response:**

Users are guided through the deposit process provided by the repository system[1]. The workflow assists the depositor step-by-step in creating the necessary metadata for the curated item. The workflow follows the DuraSpace/DSpace submission workflow.

Every submitted item in the repository is checked for whether the submission and the accompanying metadata meet the principles provided by the standards and guidelines for the CLARIN-IS Repository [2]. If not, they will be rejected and must be resubmitted when necessary amendments have been made.

[1] How to Deposit: <https://repository.clarin.is/repository/xmlui/page/deposit>

[2] Standards and Guidelines for Submission to the CLARIN-IS Repository: <https://repository.clarin.is/repository/xmlui/page/submission-guidelines>

**Reviews**

**Reviewer 1:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Reviewer 2:**

**Compliance level:**

Not Applicable - 0

## Clarín Center at the Arni Magnússon Institute for Icelandic Studies

### Comments:

### Insource/Outsource Partners

If applicable, please list them.

### Compliance level:

Not Applicable - 0

### Response:

The Division of Information Technology (UTS) at the University of Iceland[1] is an outsource partner.

UTS is responsible for backups and surveillance of the system.

Continuous integration and software maintenance is a joint venture between the IT department, and the AM Institution.

CLARIN-IS uses WAYF[2] and Clarín Service Provider Federation (CLARIN SPF[3]) to authenticate users to the repository system.

The repository software is a fork of DSpace and is developed and maintained by a group of skilled software developers from the Clarín Centres [4].

[1] University of Iceland: <https://english.hi.is/>

[2] Where Are You From (WAYF): <https://wayf.dk/>

[3] CLARIN SPF: <https://www.clarin.eu/content/service-provider-federation>

[4] UFAL- clarín-dspace: <https://github.com/ufal/clarin-dspace>

### Reviews

#### Reviewer 1:

#### Compliance level:

Not Applicable - 0

#### Comments:

#### Reviewer 2:

#### Compliance level:

Not Applicable - 0

#### Comments:

### Significant Changes

#### Summary of Significant Changes Since Last Application if applicable.

#### Compliance level:

Not Applicable - 0

#### Response:

-

### Reviews

#### Reviewer 1:

#### Compliance level:

Not Applicable - 0

#### Comments:

#### Reviewer 2:

## Clarín Center at the Arni Magnússon Institute for Icelandic Studies

### Compliance level:

Not Applicable - 0

### Comments:

### Other Relevant Information

You may provide other relevant information that is not covered by the requirements.

### Compliance level:

Not Applicable - 0

### Response:

The application is a part of a certification as an Icelandic CLARIN B-type Centre at the Arni Magnússon Institute. The Institute will pay the current fee for Core Trust Seal certification.

### Reviews

#### Reviewer 1:

### Compliance level:

Not Applicable - 0

### Comments:

#### Reviewer 2:

### Compliance level:

Not Applicable - 0

### Comments:

## Organizational Infrastructure

### R1 Mission/Scope

The repository has an explicit mission to provide access to and preserve data in its domain.

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Response:

CLARIN-IS is a part of CLARIN-ERIC. It is organised as a consortium, which comprises all the main institutions that are involved in the production and use of language resources in Iceland [1].

The ultimate objective of CLARIN-ERIC is to advance research in the humanities and social sciences by giving researchers unified single sign-on access to a platform which integrates language-based resources and advanced tools at a European level. This is implemented by the construction and operation of a shared distributed infrastructure that aims at making language resources, technology and expertise available to the humanities and social sciences research communities at large.

CLARIN-IS is committed to the long-term care and preservation of items deposited in its repository and aims to adopt the current best practices in digital preservation [2], and in particular, to obtain and operate as a certified CLARIN B Centre [3].

The national research infrastructure CLARIN-IS was established especially in relation to The language technology programme in Iceland (2019-2023) [4] and its focus has mainly been on harvesting the output of the programme, and other resources related to language technology, but the aim is to reach out to a wider group of users and domains. CLARIN-IS focuses on the Icelandic language resources. The resources can be deposited by associated researchers as well as researchers who are not affiliated with us.

[1] <https://clarin.is/en/participants/>

[2] <https://repository.clarin.is/repository/xmlui/page/about?locale-attribute=en>

[3] <https://www.clarin.eu/content/certified-b-centres>

[4] <https://aclanthology.org/2020.lrec-1.418.pdf>

## Reviews

### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

## R2 Licenses

**The repository maintains all applicable licenses covering data access and use and monitors compliance.**

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Response:

When submitting a new item the submitter accepts a "Distribution Licence Agreement" [1], in which the repository's rights and duties are described, while the submitter acknowledges that they have the right to submit the data and gives the repository centre the right to distribute the data on their behalf. Each submission is reviewed by a CLARIN-IS staff member.

When depositing a dataset to the repository, data producers must select a licence governing its access and usage from a set of licence options [2].

Everyone who downloads data is bound by the licence assigned to the item. A licence can be either free or restricted. If the data is licensed with a licence that requires signing, the user is asked to electronically sign the licence before downloading the data. The process is carefully monitored to ensure that authenticated users are real people and that the signed licence agreements are stored.

The data consumer is made aware of usage restrictions using clear visual indicators (see [3] for an example) with colour-coded "openness" of the licence ("public", "academic", "restricted"). In case of misuse, the user is denied further access to the repository and the research community is made aware of the breach of the licence agreement.

[1] Licence Agreement: <https://repository.clarin.is/repository/xmlui/page/contract>

[2] Licences: <https://repository.clarin.is/repository/xmlui/page/licenses>

[3] Handle to a deposited item: <http://hdl.handle.net/20.500.12537/215>

1. <https://repository.clarin.is/repository/xmlui/page/licenses>

## Reviews

### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### R3 Continuity of access

**The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.**

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Response:

In the year 2018 The Ministry of Education, Science and Culture commissioned The Árni Magnússon Institute for Icelandic Studies (AM) to operate a CLARIN centre in Iceland in the years 2018 - 2022 in order to fulfil Iceland's conditions to be a member of the European language technology research infrastructure CLARIN-ERIC, in the years 2018 - 2022.

At the beginning of 2023, The Ministry of Higher Education, Science and Innovation made a new contract with AM in order to fulfil Iceland's conditions to be a member of CLARIN-ERIC in the years 2023 - 2026.

As long as funding is available, the CLARIN-IS centre will ensure preservation and access. In case of a lack of funding, CLARIN-IS has measures to preserve data access.

With the expertise and knowledge of the University IT department, the CLARIN-DSpace repository platform will be monitored and maintained continuously in the future by the IT department. The group members that administer the repository have a permanent position at AM.

Thus, if CLARIN-IS funding would be interrupted, the hosting institute would be able to keep the repository running without dedicated funding for a substantial time, while most likely continuing to accept new submissions as well.

[1] <https://aclanthology.org/2020.lrec-1.418.pdf>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R4 Confidentiality/Ethics

**The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.**

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Response:

The Árni Magnússon Institute for Icelandic Studies aims to ensure that datasets that are uploaded to and made available by the repository comply with confidentiality and ethics guidelines in [1].

Each dataset submitted is monitored by CLARIN-IS staff members, and if they are in doubt about the compliance of the dataset with applicable laws or regulations, they may request more information from the submitter, and help them modify the data (by for example anonymising the texts or reshuffling the sentences) or refuse to publish the submission. It is also possible to create a new distribution licence custom-made for a particular item.

Those who download resources from the repository must agree to the conditions stated in the end-user licence associated with the data, which can include statements relating to the infringement of copyright, intellectual property and redistribution of content.

The CLARIN-IS Center is bound by the University of Iceland data protection policy [3], which includes rules for processing and handling sensitive personal data, for technical and organisational safeguards to be in place and for prevention of unauthorised access.



It should be noted that the CLARIN Legal and Ethical Issues Committee [2] organises training sessions in the legal and ethical management and distribution of text data

[1] <https://repository.clarin.is/repository/xmlui/page/about?locale-attribute=en#privacy-policy>

[2] <https://www.clarin.eu/governance/legal-issues-committee>

[3] <https://english.hi.is/node/58449/>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R5 Organizational infrastructure

**The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.**

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Response:

The CLARIN centre is hosted at The Árni Magnússon Institute for Icelandic Studies (AM) in Reykjavík.

The Institute is an independently funded academic research institute at the University of Iceland, operating under the auspices of the Ministry of Culture and Business Affairs (formerly the Ministry of Education, Science and Culture). Its role is to conduct research in the field of Icelandic studies and related scholarly disciplines, in particular, Icelandic language and literature; to disseminate knowledge in these fields, and to preserve and augment the collections entrusted to its care.

Other participants in the CLARIN-IS consortium are listed here [1]. And the CLARIN-IS staff members are listed on this page [2].

Two staff members of The Department of Lexicography, a research body within AM (whose function is lexicography, lexicology, language technology, and the study of Icelandic lexis in general), are responsible for running and managing the repository, curation of data and general service for the users. They both hold permanent positions and have been working for the centre since 2017. The CLARIN National Coordinator uses between 30% and 40% of his time on CLARIN-related tasks, while the technical staff member uses around 80% of his time on the development and daily service of the system. These numbers can vary though from time to time. The National Coordinator and the software developer work closely together.

The software specialist is a member of the Standing Committee for the CLARIN Technical Center [3] and the CLARIN Centre Assessment Committee [4] and the National Coordinator is a member of the CLARIN User Involvement Committee.

The University IT Department is responsible for the maintenance of the underlying technical infrastructure such as security and server system upgrades, while the AM Institute's technical responsibility is in the form of the development and upgrade of the repository software, as well as curation of data and observation of data soundness.

The repository system is an integrated part of the University's technical infrastructure and as such undergoes regular and systematic updates that are out of the scope of the AM Institute.

The University IT Department takes care of system security, backups and daily maintenance of the system. This service is part of a general technical service agreement between the University and the AM Institute.

[1] <https://clarin.is/en/participants/>

[2] <https://clarin.is/en/about/>

[3] <https://www.clarin.eu/governance/standing-committee-clarin-technical-centres>

[4] <https://www.clarin.eu/governance/centre-assessment-committee>

### Reviews

### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### R6 Expert guidance

**The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).**

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Response:

CLARIN-IS is a member of the CLARIN-ERIC infrastructure and, as such, is in regular touch with its experts. CLARIN-IS is a member of the CLARIN-ERIC Standing Committee of Technical Centres and User Involvement Committee [1]. In regular teleconferences and face-to-face meetings in the scope of these committees, the latest developments and subsequent necessary modifications to the workings of the CLARIN repositories are also discussed.

The coordinator of the CLARIN C-centre, The Árni Magnússon Institute for Icelandic Studies (AM), is also the national coordinator for CLARIN-IS and as such participates in meetings with other national coordinators on a monthly basis. In addition to the fixed meetings, the employees of the Centre regularly take part in CLARIN-ERIC conferences, workshops and other meetings to ensure the exchange of expert knowledge and stay up-to-date with recent developments.

AM's department of Lexical Studies & Lexicography is mostly responsible for running the CLARIN-IS centre and is thus responsible for the management of the repository. The department has as its staff members experts in both lexical studies and language technology. The language technology branch of the department has mainly been focusing on the creation and maintenance of various corpora, wordlists and wordnets, as well as the creation and usage of related tools (Pos-tagger, lemmatizer, UD-tagger etc.) and the use of different formats, such as TEI.

Since we are using a modified version of DSpace, called CLARIN-DSpace, developed by LINDAT/CLARIN, the Czech national node of CLARIN-ERIC, we take advantage of the support that LINDAT/CLARIN offers.

The CLARIN-ERIC Standing Committee Technical Centres (SCCTC)[1] and CLARIN task force offer technical advice and share relevant knowledge.

The repository server is managed and hosted at The Division of Information Technology (UTS) at the University of Iceland[2]. UTS is responsible for backups, system updates and surveillance. The IT department has a system administrator at our disposal in any technical circumstances regarding the operation of the system when needed.

[1] SCCTC: <https://www.clarin.eu/governance/standing-committee-clarin-technical-centres>

[2] University of Iceland: <https://english.hi.is/>

### Reviews

### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

**Comments:**

**Digital Object Management**

**R7 Data integrity and authenticity**

**The repository guarantees the integrity and authenticity of the data.**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Response:**

The depositor's identity is verified either by a Shibboleth account (providing their name and email address and attributes) or a local account. The dataset is then linked to the Data Producer's account.

Once a dataset is published, the dataset version is fixed and immutable. If the Data Producer wants to edit the bitstream, a new version has to be created. It is possible to make minor changes to the metadata, for example in the case of misspellings, minimal corrections or adding further information. Those changes have to be done through the help desk since a Data Producer cannot change the dataset in any way after submission. Any changes to metadata are recorded in provenance information. A timestamp is recorded in the following internal metadata field, dc.date.accessioned, to hold the last updated timestamp. Log files are stored including appropriate messages regarding changes made to data and/or metadata.

An item has to pass through submission where certain metadata fields are mandatory. A curator also verifies the data before it is accepted to the repository. The curator has several curator tools available that he can use to verify the data. Some of these tools are part of the weekly checks (e.g., whether all required metadata are present, URLs are working). The results of weekly checks are automatically sent to the repository staff.

The repository uses MD5 checksums to verify the data integrity of all digital objects. The checksum is attached as metadata to the created bitstream, together with other file information, during the submission.

To verify that a digital object has not been altered or corrupted we periodically (on a weekly basis) verify the md5 checksums of the objects.

The submission workflow ensures that the Submission Information Package (SIP) is validated in compliance with established policies. All submissions follow the lifecycle policy [1].

[1] <https://repository.clarin.is/repository/xmlui/page/item-lifecycle>

**Reviews**

**Reviewer 1:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

**Reviewer 2:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

**R8 Appraisal**

**The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Response:**

The repository accepts any language data, including but not limited to language corpora, text and speech collections, machine-readable dictionaries, computational lexicons and language processing tool models, as well as language processing software. As for now, the main emphasis has been on language technology due to the ongoing language technology programme already mentioned and all the products of that programme are harvested in the

## Clarín Center at the Arni Magnusson Institute for Icelandic Studies

repository. The data we accept are divided into four categories: corpora, language descriptions, lexical-conceptual resources and tools.

The submission workflow requests that the depositor provides all metadata necessary to interpret the data prior to acceptance and publication in the repository. We are using a modified version of DSpace, called CLARIN-DSpace, developed by LINDAT/CLARIN, the Czech national node of CLARIN ERIC, and the system ensures that all basic metadata are provided (type of data, authors, publisher, date of submission etc.) while other information is encouraged (funding, related articles etc.). The metadata provided through the system is automatically valid with respect to the CMDI metadata schema [1]. Thorough instructions concerning the submission workflow are given here [3].

Usually, the repository does not accept entries without data (i.e. without the bitstreams attached to the entry). We encourage the use of one of several preferred formats for the bitstream [2] and users have to have a valid reason not to use one of those formats, and then thorough documentation of the format is required.

[1] CMDI: [http://media.dwds.de/clarin/userguide/text/metadata\\_CMDI.xhtml](http://media.dwds.de/clarin/userguide/text/metadata_CMDI.xhtml)

[2] Submission guidelines: <https://repository.clarin.is/repository/xmlui/page/submission-guidelines>

[3] Submission workflow: <https://repository.clarin.is/repository/xmlui/page/deposit>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R9 Documented storage procedures

#### The repository applies documented processes and procedures in managing archival storage of the data.

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Response:

The University of Iceland IT department is (UTS) responsible for storage of media and monitoring of servers.

All deposited resources are checked manually by a human validator and through XML schema validation in the ingestion process. The curation of the deposited resources follows an instruction that guides the submitter in the process [1].

The ingestion is done through a web interface. When the deposited item has been submitted our curators check the metadata and modify it if necessary. Any attached data and bitstream is checked for compliance with the repository submission guidelines [2], and tested as thoroughly as possible. In the final step of the ingestion, when a resource has been approved by the curators, the resource automatically gets a persistent identifier (PID) [3]. The submitter can not do any changes to the submitted resource after that. If an error is discovered in the resource metadata after it has been deposited, the curators can update the metadata.

The repository system and the repository asset store is hosted in a dedicated virtual environment running on Security Enhanced (SELinux) Redhat operating system with strict access control, where only system administrators have access to the server. All curated resources are MD5 checksum checked on a regular basis to detect possible corruptions in the asset store data. In case of any disturbance in the operation of the system, the IT department can restore any daily snapshot of the installation itself, going three months back in time.

The repository database management system is hosted on a dedicated database server. The databases are backed on a daily basis and can be restored on demand.

The source code and documentation of the installed repository system can be found on GitHub [3] and any local modifications on the source are kept in AM private GitHub repository.

The backup system is Veeam. Backups are both on-site and off-site. Backups are end-to-end encrypted. All hosts are backed up at least once daily.

Advania [3] is UTS outsource partner. Backups are stored in Advania Data Centres.

All of our data is open access. There is thus no concern about data breaches regarding access to the deposited resources.

## Clarín Center at the Arni Magnusson Institute for Icelandic Studies

[1] Deposit: <https://repository.clarin.is/repository/xmlui/page/deposit>

[2] Submission guidelines: <https://repository.clarin.is/repository/xmlui/page/submission-guidelines>

[3] Clarín-DSpace repo: <https://github.com/ufal/clarin-dspace>

[4] PID: <http://handle.net> <https://repository.clarin.is/repository/xmlui/page/faq#what-is-the-handle-good-for>

[4] Advania: <https://www.advania.com/>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R10 Preservation plan

**The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.**

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Response:

The curator of the deposited item ensures that the depositor accepts the agreement between the two parties before depositing the item. For every deposit, there is a standard contract [1] describing the rights and duties of the curator and acknowledgement of the submitter that they have the right to submit the data and gives the repository the right to distribute the data on their behalf. In case of a violation of the contract, steps will be taken to withdraw the item until the issue will be resolved. The repository accepts only widely used and well-known file formats [2] which should increase the lifespan of the deposited items. In case of obsolete or unsupported file formats, the curator has the right to convert the files to a new format, but that has never been necessary in our case.

The Árni Magnusson Institute (AM) is committed to long-term preservation and access to the data contained in the repository. In case of withdrawal of funding that might make it impossible for AM to continue managing the repository, the repository content can be transferred to another CLARIN centre as stated in the preservation policy [3].

All the deposited assets will be long-term preserved at the University of Iceland IT department, where the repository is hosted. The IT department guarantees regular backups of data and its consistency. The IT Department is in the process of certifying its activities according to ISO/IEC 27001:2005 standard.

[1] Distribution Licence Agreement: <https://repository.clarin.is/repository/xmlui/page/contract>

[2] Guidelines for Data Submission: <https://repository.clarin.is/repository/xmlui/page/submission-guidelines#accepted-binary-formats>

[3] Preservation Policy: <https://repository.clarin.is/repository/xmlui/page/about#preservation-policy>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

#### R11 Data quality

**The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality- related evaluations.**

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Response:

Data producers must submit their data through a web interface where they are led through the submission process in a few clear steps. The depositors are authenticated through WAYF [1] or other national federations via CLARIN Service Provider Federation (CLARIN SPF [2]).

Sufficient completeness and quality of metadata are assured, during the submission process, by requiring certain fields (without them being filled in, the submission cannot be completed) and by filling in certain fields automatically (e.g., the PID and date of entry into the repository).

Data producers must submit their data through a web interface where they are led through the submission process in a few clear steps. The depositors are authenticated through WAYF [1] or other national federations via CLARIN Service Provider Federation (CLARIN SPF [2]).

Sufficient completeness and quality of metadata are assured, during the submission process, by requiring certain fields (without them being filled in, the submission cannot be completed) and by filling in certain fields automatically (e.g., the PID and date of entry into the repository).

The web interface is part of the CLARIN-DSpace system, developed by LINDAT/CLARIN, the Czech national node of CLARIN ERIC. DSpace aims to follow the OAIS (Open Archival Information System) reference model that seeks to entrust the quality and durability of all data. The system follows the standard principles of a high-quality digital repository such as the usage of Persistent Identifiers via handle.net, authorization and authentication using Shibboleth authentication and sharing of metadata and binaries.

The assessment of the quality of the data and metadata is thus mainly automated. But the metadata, the data format and the data soundness are also checked and approved by the repository curators before the resources are published in the repository's collection. The curators have access to several curation tools the repository offers to check the quality of the data and the metadata. The type of attached bitstream must follow the submission guidelines [3]. If submitted metadata is inadequate, documentation is missing or the data is in some way flawed, the submitted item is kept in a task pool and the submitter is informed about the necessary adjustments. If the submitter is asked to supply a new bitstream, the curators test the newly supplied bitstream before replacing the rejected bitstream in the task pool. The submission cycle is complete when the supplied bitstream and metadata have passed all tests.

Submissions can be replaced by newer versions. In that case, the older version is marked as such with a pointer to the replacing versions [5].

The repository metadata is harvested by CLARIN Virtual Language Observatory VLO [4] and tested on a regular basis at the Clarín Curation Dashboard [5] for soundness. In case of an XML validation failure, the metadata will be updated by the curators. No other alteration of the submission is allowed.

[1] WAYF: <https://www.wayf.dk/>

[2] CLARIN SPF: <https://www.clarin.eu/content/service-provider-federation>

[3] Submission Guidelines: <https://repository.clarin.is/repository/xmlui/page/submission-guidelines>

[4] VLO: <https://www.clarin.eu/content/virtual-language-observatory-vlo>

[5] Example of version update of assessment: <http://hdl.handle.net/20.500.12537/107>

### Reviews

#### Reviewer 1:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

#### Reviewer 2:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

#### R12 Workflows

**Archiving takes place according to defined workflows from ingest to dissemination.**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Response:**

The repository accepts four kinds of data types (corpora, language descriptions, lexical-conceptual resources and tools) and the same workflow applies to all of them [1]. The workflow can be divided into the four following steps:

- 1) After login the depositor submits their data to the repository using a defined submission workflow internally configured in our repository [2]. During the submission process all mandatory, as well as some optional, metadata fields are entered, the bitstream is uploaded, a licence for the dataset is chosen [3] and the Distribution Licence Agreement is signed electronically.
- 2) The curators receive a notification via e-mail and perform an assessment of the submission. The repository offers the curators some automatic tools that help them to validate metadata and the integrity of the submitted data. In some cases, a curator will contact the depositor and request re-submission. The curator then works closely with the depositor in the case of insufficient data or metadata quality, and/or licensing issues.
- 3) The submission information package (SIP) is approved, published and receives a new persistent identifier (PID) [4].
- 4) It becomes immediately visible and retrievable via the repository web interface and harvestable to the CLARIN Virtual Language Observatory (CLARIN VLO [5]) at the next scheduled OAI-PMH harvesting.

[1] Deposit: <https://repository.clarin.is/repository/xmlui/page/deposit>

[2] Submission Guidelines: <https://repository.clarin.is/repository/xmlui/page/submission-guidelines>

[3] List of Available Licences: <https://repository.clarin.is/repository/xmlui/page/licenses>

[4] Persistent Identifier: <http://handle.net/>

[5] CLARIN VLO: <https://vlo.clarin.eu>

**Reviews**

**Reviewer 1:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

**Reviewer 2:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

**R13 Data discovery and identification**

**The repository enables users to discover the data and refer to them in a persistent way through proper citation.**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Response:**

When a dataset has been approved by the curator it receives a unique persistent identifier (PID). The repository issues PIDs via a local handle.net server running on the repository's server with a global PID prefix, 20.500.12537, registered by the handle.net Registry (HNR) [1].

The repository has browse- and search capabilities and provides faceted search and filter queries on the metadata [2]. All the metadata, as well as text files, are indexed for full-text search.

To enhance the discoverability of content, the repository supports metadata harvesting via OAI-PMH protocol using DSpace's OAI v2 server. The CLARIN Virtual Language Observatory (CLARIN VLO) harvests metadata from our OAI endpoint into the CLARIN VLO, where language resources may be discovered using a facet browser [3]. Through VLO and the CLARIN Component Metadata Initiative [4], the potential user can search among larger collections of materials harvested from various data providers within CLARIN and related communities.

For each dataset, a citation is provided and users are asked to acknowledge and cite data sources properly in all publications and outputs [5].

[1] HNR: <http://handle.net/>

[2] Discoverability: <https://repository.clarin.is/repository/xmlui/discover>

[3] CLARIN VLO: <https://vlo.clarin.eu/search>

[4] Component Metadata Initiative: <https://www.clarin.eu/content/component-metadata>

[5] Citation: <https://repository.clarin.is/repository/xmlui/page/cite>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R14 Data reuse

**The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.**

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Response:

The CLARIN-IS repository requires a set of metadata to be provided [1]. Required metadata complies with Dublin Core (DCMI)[2] and together with additional custom metadata fields, combine to fulfil the requirements of the CMDI metadata schema [3]. The user can see this descriptive metadata, as well as licensing information covering intellectual property, conditions and others on the item view page in the repository.

CLARIN-IS repository accepts well-known and widely used formats [4] suitable for long-term preservation, that ensure that all digital objects are clearly understandable and usable by members of the Designated Community.

The metadata information combined with a well known and widely used format for digital objects ensures the assets' understandability over time.

The curator has the right [5] to migrate the files to a newer format if necessary, keeping the original submission intact.

If users have problems understanding metadata or licensing they can contact the help desk.

Re-submission of data is allowed, as a new version.

[1] <https://repository.clarin.is/repository/xmlui/page/metadata>

[2] DCMI: <https://www.dublincore.org>

[3] CMDI Metadata Schema: <https://www.clarin.eu/content/component-metadata>

[4] Accepted Binary formats: <https://repository.clarin.is/repository/xmlui/page/submission-guidelines#accepted-binary-formats>

[5] Contract: <https://repository.clarin.is/repository/xmlui/page/contract>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:



The guideline has been fully implemented in the repository - 4

**Comments:**

### Technology

#### R15 Technical infrastructure

**The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Response:**

Like so many CLARIN centres, this repository system is the CLARIN-DSpace[1] fork of the DuraSpace institutional repository application [2]. The repository system is maintained at the Czech national node of CLARIN ERIC. The system is actively being developed and a Slack channel is open for members of CLARIN-ERIC for DSpace development discussion. Here members can follow the development progress and participate in the dialogue. CLARIN-IS is a member of the Standing Committee for CLARIN Technical Center [3].

The CLARIN-IS repository [4] runs on a virtual server maintained by the University of Iceland's IT department. The server runs a security Enhanced (SELinux) Redhat operating system (See R16).

**Hardware:**

The hardware is currently being upgraded from four-year-old Lenovo FLEX to brand-new Lenovo ThinkSystem

**Backup:**

All hosts are backed up at least once daily. Backups are stored in Advania Data Centres [5]. The backup system is Veeam. Backups are both on-site and off-site. Backups are end-to-end encrypted. In case of system failure, the IT department's failover procedures guarantee that the system will be up and running on demand. The system can be restored from an arbitrary point, 93 days back in time (See R16).

[1] Ufal/Clarín-dspace: <https://github.com/ufal/clarin-dspace>

[2] Dspace: <https://dspace.lyrasis.org/>

[3] SCCTC: <https://www.clarin.eu/governance/centre-assessment-committee>

[4] CLARIN-IS Repository: <https://repository.clarin.is>

[5] Advania: <https://www.advania.com/>

### Reviews

**Reviewer 1:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

**Reviewer 2:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

#### R16 Security

**The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Response:**

## Clarín Center at the Arni Magnusson Institute for Icelandic Studies

The hosting and management of the virtual machine, storage system, operating system, network storage, database maintenance and backup system for CLARIN-IS at Árnastofnun are managed by the IT department of the University of Iceland (UTS).

The University of Iceland finished implementing ISO27001 Information Security Standard in Q2 2022 and was officially certified in Q4 2022. UTS has incident management as well as change management according to ISO27001.

**Backup & Recovery:** The CLARIN-IS virtual machine is backed up once every day to the UTS backup system (Veeam), as well as a separate backup of the database to disk. The backup system stores multiple copies of the backups at an offsite location. The disks used in the CLARIN-IS virtual machine are configured to RAID6 in the network storage system to guarantee no data loss in case of a disk failure.

Current backup and recovery plan ensures that the system can be recovered to one point in time each day. The backup system stores the last 93 days of daily backups.

**Security:** Physical security is controlled by access control where authorised personnel must have a physical key and a pin code to gain access. An offsite password solution is used to store necessary passwords.

The CLARIN-IS virtual machine uses its own firewall accepting only HTTPS and SSH connections. Virtual machine authentication methods are local login with an SSH key.

The repository authentication system is Shibboleth [1], a web-based single sign-on system based on the SAML2.0 protocol where users can log in with their institutional credentials. No credentials are kept on the server. CLARINs Federated Authentication is through the CLARIN Service Provider Federation (SPF) [2]. The authentication method is Edugain through Wayf [3]. University of Iceland is a member of the EduGAIN federation.

Extensive user rights in the CLARIN-IS repository system are restricted to the least persons necessary. UTS are responsible for the monitoring of the website and other IT systems as part of the Service Level Agreement.

[1] Shibboleth: <https://shibboleth.atlassian.net/wiki/spaces/CONCEPT/overview>

[2] SPF: <https://www.clarin.eu/content/service-provider-federation>

[3] Wayf: <https://www.wayf.dk/en>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### Applicant Feedback

#### R17 Applicant Feedback

**We welcome feedback on the CoreTrustSeal Requirements and the Certification procedure.**

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Response:

-

### Reviews

#### Reviewer 1:

## Clarín Center at the Arni Magnusson Institute for Icelandic Studies

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

-

**Reviewer 2:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

Generally speaking, this was a comprehensive application. I would recommend that the certification be approved.