



## Assessment Information

[CoreTrustSeal Requirements 2020–2023](#)

Repository:	CLARINO Bergen Centre
Website:	<a href="https://clarino.uib.no">https://clarino.uib.no</a>
Certification period:	18 April 2023 - 17 April 2026
Requirements version:	CoreTrustSeal Requirements 2020-2022

This repository is owned by: **University of Bergen Library**

## 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:

- Archive
- Domain or subject-based repository

#### Reviews

##### Reviewer 1:

#### Compliance level:

Not Applicable - 0

#### Comments:

Accept

##### Reviewer 2:

#### Compliance level:

Not Applicable - 0

#### Comments:

### Description of Repository

Provide a short overview of the repository.

#### Compliance level:

Not Applicable - 0

#### Response:

The CLARINO Bergen Centre is a center of the Norwegian national CLARINO research infrastructure, which in turn is Norway's contribution to the construction and operation of CLARIN ERIC [1]. The aim of the CLARINO Bergen Centre is in line with that of CLARIN ERIC, i.e. to make existing and future language resources easily accessible for researchers, and to bring eScience to humanities disciplines. Thus the CLARINO Bergen Centre is restricted to the domain of language data and tools; datasets may be any collections of language expressions (e.g. corpora, lexicons, termbases) or data about language (e.g. quantitative data obtained from field work or experiments).

The CLARINO Bergen Centre is operated by the University of Bergen (UiB) which is one of the partners in the CLARINO consortium carrying out the CLARINO project funded by the Research Council of Norway (RCN). The project has continued in the CLARINO+ project, aimed at updating the infrastructure, which is also funded by the RCN.

The CLARINO Bergen Centre runs a repository (henceforth called the CLARINO Bergen Repository or for short the Repository [2]) offering download and upload of resources. Furthermore the CLARINO Bergen Centre also provides access to deposited resources for online search through the following two online systems for exploration: Corpuscle, a corpus exploration tool, and INESS, the Norwegian Infrastructure for the Exploration of Syntax and Semantics. The CLARINO Bergen Centre also provides COMEDI, an online component metadata editor [3], and cooperates with Menota, the Medieval Nordic Text Archive. All data and services offered by the CLARINO Bergen Centre (the Repository, Corpuscle, Menota, INESS and COMEDI) adhere to CLARIN principles for metadata, PIDs, licensing and access. The further implementation of the Term Portal and making it accessible to CLARINO is also an ongoing task in CLARINO+.

## CLARINO Bergen Centre

[1] CLARIN ERIC, the European Research Infrastructure for Language Resources and Technology: <https://www.clarin.eu/>

[2] CLARINO Bergen Repository: <https://repo.clarino.uib.no>

[3] The additional online services Corpuscle, INESS, Menota and COMEDI are accessible through a top line menu at the main site of the Repository.

### Reviews

#### Reviewer 1:

##### Compliance level:

Not Applicable - 0

##### Comments:

Accept

#### 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 CLARINO Bergen Centre helps researchers working in the Humanities and the Cultural and Social Sciences who have a need for managing, preparing, storing, accessing or analyzing research data related to language. Typical producers of language data are researchers in corpus linguistics, translation, grammar, terminology, lexicography, sociolinguistics, psycholinguistics, philology, philosophy, historical linguistics, political sciences, anthropology, etc. Producers of language processing tools are researchers in natural language processing. Consumers of language data and tools in the CLARINO Bergen Centre are often the same groups as producers; they are interested in analyzing language data and using text processing tools available in the CLARIN infrastructure. Many resources in the CLARINO Bergen Centre originate from Norway, but the centre is open to all CLARIN members, contingent upon licenses.

### Reviews

#### Reviewer 1:

##### Compliance level:

Not Applicable - 0

##### Comments:

Accept

#### Reviewer 2:

##### Compliance level:

Not Applicable - 0

##### Comments:

### Level of Curation

## CLARINO Bergen Centre

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:**

- A. Content distributed as deposited
- B. Basic curation – e.g. brief checking; addition of basic metadata or documentation
- C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation

**Reviews**

**Reviewer 1:**

**Compliance level:**

Not Applicable - 0

**Comments:**

Accept

**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:**

The Repository performs basic curation of the submissions by checking and editing the metadata. The Repository uses the DSpace submission workflow which foresees several curation steps before completing a submission. Curation of submitted items is performed which implies assessment and revision of metadata and quality checks by experts affiliated with the CLARINO Bergen Centre. See checklist at R4. Automatic tools are helping the editors to verify and validate metadata and the integrity of the submitted data [1][2].

The basic curation process foresees the possibility of returning the submission to the submitter for additional changes before the dataset is published in the repository. On some occasions the repository assists in enhanced curation, for example, conversion to a different format, enhancement of metadata, or customization of licenses.

[1] CLARIN DSpace metadata info: <https://github.com/ufal/clarin-dspace/wiki/Metadata-info>

[2] DSpace curation system: <https://wiki.duraspace.org/display/DSDOC5x/Curation+System>

**Reviews**

**Reviewer 1:**

**Compliance level:**

Not Applicable - 0

## CLARINO Bergen Centre

### Comments:

Accept

### Reviewer 2:

### Compliance level:

Not Applicable - 0

### Comments:

### Insource/Outsource Partners

If applicable, please list them.

### Compliance level:

Not Applicable - 0

### Response:

The Repository is maintained at the University of Bergen Library, using cloud infrastructure from the Norwegian Open Research Cloud (outsource partner) [1], an Openstack [2] cloud offering, managed in collaboration by SIKT (outsource partner) [3], the University of Bergen and the University of Oslo. In addition, we use additional infrastructure from the University of Bergen IT-department, i.e., storage with backup. See R5 for more details.

The University of Bergen Library is responsible for provisioning, configuring, and maintaining the server, software maintenance of the repository, implementation of the repository policies (see below) and user support. NREC supplies the required infrastructure (self-managed by the University Library), and support via email and chat. The Department of Linguistic, Literary and Aesthetic Studies has the scientific coordinator and additional support. The IT department provides storage and back-up. The IT-department is a unit within the UiB, organised as a part of the central administration. The IT-department is one of the active partners in running NREC; however NREC is considered a national service with different terms of service [4]. The UiB Library is a unit organised at the level of the faculties and the University Museum. However, the University Library, like the IT-department and the rest of the Central Administration, reports directly to the University Director. A chart is available at [5]. The roles and responsibilities of the IT-department relative to the other departments ("system owners") are defined in the University of Bergen statutes [6]. This obviates the need for Service Level Agreements with each service. See R15 for more details. Further development of CLARIN DSpace software, of which the Repository is a fork, is performed by a strong community of developers in more than 10 CLARIN centres where features, updates and security measures are presented and solved [7]. The CLARIN Dspace community has a slack channel (#dSPACE) for communication [8].

[1] <https://nrec.no>

[2] <https://www.openstack.org/>

[3] <https://sikt.no/>

[4] <https://docs.nrec.no/terms-of-service.html>

[5] UiB organisational chart: <https://www.uib.no/en/about/organisationchart>

[6] Statutes (most relevant parts are in Norwegian only): <https://regler.app.uib.no/regler/Del-4-OEkonomi-eiendom-og-IKT/>

[7] CLARIN DSpace development: <https://github.com/ufal/clarin-dspace>

[8] <https://clarineric.slack.com>

### Reviews

### Reviewer 1:

### Compliance level:

Not Applicable - 0

### Comments:

Accept

### Reviewer 2:

### Compliance level:

Not Applicable - 0

## CLARINO Bergen Centre

### 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:

### 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 current application is part of the re-certification of the CLARINO Bergen Centre as a CLARIN type B centre. CLARIN will pay the fee for the current CoreTrustSeal certification.

The CLARINO Bergen Centre was certified as a CLARIN type B centre in 2019 [1], which confirms that resources are managed and made accessible through CLARIN compliant interfaces in a stable and persistent way. The CLARINO Bergen Repository was certified with the Core Trust Seal on 18/11/2019 [2].

[1] CLARIN B centre certificate: <http://hdl.handle.net/11372/DOC-116>

[2] Core Trust Seal: <https://www.coretrustseal.org/wp-content/uploads/2019/11/CLARINO-Bergen-Centre.pdf>

### Reviews

#### Reviewer 1:

### Compliance level:

Not Applicable - 0

### Comments:

#### Reviewer 2:

### Compliance level:

## CLARINO Bergen Centre

Not Applicable - 0

### Comments:

Accept

## 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:

The mission of the CLARINO Bergen Centre is to provide reliable archiving and documentation as well as easy access to language-related data, tools, services and associated metadata for research and development purposes (e.g. corpora, lexicons, termbanks, audio and video recordings, grammars, language models, parsers, format converters, lexicon extraction tools, etc.) to scholars in the fields of Social Sciences and Humanities and beyond. CLARINO implements the Norwegian contribution to CLARIN ERIC [1] which has the objective of advancing research in humanities and social sciences by giving researchers unified single sign-on cross-border access to a platform which integrates language-based resources and advanced tools. 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 [2]. This work has had national approval first through the national infrastructure project CLARINO funded by the RCN (2012-2019, grant number 208375) [3] and presently through the upgrade project CLARINO+, also funded by the RCN (2020-2026, grant number 295700) [4]. In turn, Norway's national contribution receives yearly approval by the CLARIN General Assembly. By accepting the RCN grant, UiB has approved of the work plan and mission statement. The mission statement is shown on the Repository website [5]

[1] CLARIN ERIC: <https://www.clarin.eu>

[2] CLARIN Short Guide: <http://www.clarin.eu/files/centres-CLARIN-ShortGuide.pdf>

[3] The CLARINO project and its funding by the RCN are mentioned in the following RCN documents:

<https://www.forskingsradet.no/siteassets/publikasjoner/1254023795613.pdf> and

<https://www.forskingsradet.no/siteassets/publikasjoner/1253995988917.pdf>

[4] The upgrade project is mentioned by the RCN on this webpage:

<https://www.forskingsradet.no/sok-om-finansiering/midler-fra-forskingsradet/infrastruktur/norsk-veikart-for-forskingsinfrastruktur/prosjektbeskrivelser/b.esfri-infras>

[5] Mission statement and policies: <https://repo.clarino.uib.no/xmlui/page/about#mission-statement>

### Reviews

#### Reviewer 1:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

Accept

#### 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:

## CLARINO Bergen Centre

The guideline has been fully implemented in the repository - 4

### Response:

Each resource in the CLARINO Bergen Centre is made available under the conditions attached to it. The relations between the involved parties are regulated as follows.

1) For every deposit, a standard contract is made between the CLARINO Bergen Centre and the submitter, the so-called "Distribution License Agreement" [1], in which the rights and duties of each party are described. This agreement acknowledges that the submitter has the right to distribute the resource and gives the CLARINO Bergen Centre the non-exclusive right to distribute the data on their behalf under certain conditions. Submitters can choose among standard licenses [2] or set a custom license. Submitters must be reliably authenticated (registered and logged in, see also R4).

2) Users who download or otherwise access the data are bound by the end user license assigned to the item. Restricted licenses require users to be reliably authenticated and to electronically accept the license before access to the resource is authorized. Only files licensed as publicly available can be downloaded without authentication. The Repository keeps a log of accepted licenses. We do not however log individual downloads, neither do we routinely inspect the activity of users as part of our workflow. So far, as before, our trust in the community of authenticated users to respect the licenses they formally have accepted, seems justified. No incidences have been reported to us. It is also beyond the possibility of the Repository to check all use of the resources, but if possible, abuse is signaled to us, we can retrieve which licenses have been accepted by which authenticated users.

3) Users are bound by the Repository's Terms of Service [3] which specify that users must acknowledge the deposited work in their publications, using the work's persistent identifier, its original author(s)/creator(s), and any publisher where applicable. Full items must not be harvested by robots except transiently for full-text indexing or citation analysis. Full items must not be sold commercially unless explicitly granted by the attached license without formal permission of the copyright holders. Furthermore, the user shall agree to observe best practices regarding research ethics. This includes treating colleagues, stakeholders, customers, suppliers and the public respectfully and professionally, taking into account confidentiality when appropriate, respecting cultural differences and having an open and explicit relationship with government, the public, the private sector and other funders.

Users who disrespect the license conditions or Terms of Service and fail to repair their transgression without delay will be excluded from using the services of the CLARINO Bergen Centre.

[1] Distribution License Agreement (also called "Deposition License Agreement"): <https://repo.clarino.uib.no/xmlui/page/contract>

[2] List of Available Licenses: <https://repo.clarino.uib.no/xmlui/page/licenses>

[3] Terms of Service: <https://repo.clarino.uib.no/xmlui/page/terms-of-service>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

Accept

#### 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:

As part of the nationally funded CLARINO infrastructure project, started in 2012, the CLARINO Bergen Centre was established as a permanent infrastructural offer. All its services are currently in the operational phase and are intended to be continued indefinitely. Under the current updating project CLARINO+, the University of Bergen has committed itself to maintaining its part of the CLARINO infrastructure at least until 2029. With a view to local organizational persistence, the Repository has been embedded in the University of Bergen Library, a unit with high stability which curates digital data as



## CLARINO Bergen Centre

well as printed publications.

The University of Bergen Library strategy 2016-2022 [1] states that the University Library is a central part of the research infrastructure of the University of Bergen, and that it shall "develop and utilize technology and infrastructure for retrieval and display of literature and research data" (p. 11, our translation). On p. 12 it is said that the Library shall "offer high level expertise for the production and exchange of metadata", and "follow local, national and international guidelines", furthermore that it shall "contribute to the development of guidelines and systems for storage and reuse of research data". The strategy thus comprises the Library's commitment to Clarino and the running of the Repository.

In the more distant future, continuity will also depend on regular updating in order to remain relevant and compatible in CLARIN; to that end the CLARINO consortium applies regularly for funding to the Research Council of Norway, but this is a competitive process. The continuity of the Repository's integration in the CLARIN ecosystem is therefore contingent on Norway's continued membership in CLARIN ERIC as well as on continued national funding of CLARINO.

In the unlikely case that the Government of Norway would cease its membership and would no longer support future updates, the UiB Library would maintain the Repository on its own funding, as confirmed by the Library director. In the unlikely case that the UiB library would not be in a position to maintain the repository in the long term, the Repository holdings would be transferred to one or more sister repositories in CLARIN [2]. Technically, such a transfer will be smooth thanks to the fact that many other CLARIN repositories are forks of the same CLARIN DSpace system. As far as licenses are concerned, many resources in the Repository are already cleared for distribution by any CLARIN centre (and in fact some are already duplicated at other CLARIN centres), while some other Distribution License Agreements would have to be renegotiated.

[1] University of Bergen Library strategy 2016-2022 (in Norwegian):

[https://www.uib.no/sites/w3.uib.no/files/attachments/ub\\_strategiplan\\_endelig\\_versjon\\_2016-2022\\_v4\\_web.pdf](https://www.uib.no/sites/w3.uib.no/files/attachments/ub_strategiplan_endelig_versjon_2016-2022_v4_web.pdf)

[2] Preservation policy statement: <https://repo.clarino.uib.no/xmlui/page/about#preservation-policy>. The UiB Library director has in an official letter approved of this policy.

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

Accept

#### 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:

During the submission process, the submitter accepts the repository policy called "Distribution License Agreement" [1] in which they confirm their responsibility regarding the submission, in particular regarding the right to grant the distribution right. The deposited item lifecycle is explained on the website [2]. Information about the requirements and handling of disciplinary and ethical issues is available in the text of the Terms of Service [3].

Any submitter must be authenticated through the federation of ID providers (eduGain and CLARIN Service Provider Federation), which ensures a high level of trust. No anonymous upload of items is allowed. In case of problems, Repository editors can retrieve the exact dates and users who have deposited the resources; for restricted resources, also authenticated users who have accessed (downloaded) the resources are logged. This means that, if needed, users who do not comply to the license or the Terms of Service can be blacklisted: they can be denied further access to the repository and the research community may be made aware of any abuse.

Each submission is validated automatically and manually by an expert Repository editor, whose task includes a check if privacy and ethical issues are respected. If the editor detects problems with the submission, he/she will interact with the depositor, and request or suggest modifications. The

## CLARINO Bergen Centre

submission will not be approved until all problems are resolved. Repository editors are data experts who are careful not to make changes which in themselves may create privacy or ethical issues.

This is the manual checklist used by the administrator for each submission:

"On the basis of the submitted data and metadata for new submissions, check that:

1. There are no GDPR/personal data issues represented by publishing the datasets in Clarino Bergen Repository. The Repository does not accept data sets containing personal data. Anonymised data sets are generally accepted. If in doubt, check that the submitter has cleared the publication procedure with the Norwegian Centre for Research Data (NSD)[4].
2. There are no IPR/Copyright issues represented by publishing the datasets in Clarino Bergen Repository. Assess the submitter's acknowledgement that (s)he is the holder of the rights to the data. If in doubt, check that depositing rights are cleared with other rightholders that might be implied.
3. Data formats are compliant with CLARIN recommendations (LRT). If not: Ask the submitter to convert the data. Assist in data conversion if necessary and possible.
4. CMDI metadata file is provided. If not: Ask the submitter to provide CMDI metadata. Assist in producing metadata if necessary.

Remember for old submissions that any changes to data need to take the form of submitting a new version of the data as a separate data set. Submitted data sets can not be changed."

All Norwegian research producing data with potential privacy issues must clear their procedures with the Norwegian Centre for Research Data NSD. As stated in the checklist above, the Repository does not accept data sets containing overt sensitive personal data. Anonymized (or pseudonymized) data sets are generally accepted, as well as sign language data obtained with consent and released under a restricted license. If in doubt, we check that the submitter has cleared the publication procedure with the NSD. Since even anonymized or pseudonymized data may carry residual risks of revealing personal information, such data are normally released under a restricted license (CLARIN RES license with the additional PRIV condition) [5].

All interaction with the Repository is via secure protocols (https).

The CLARIN Legal Information Platform [6], set up and run by the CLARIN Legal and Ethical Issues Committee (CLIC) of CLARIN ERIC, provides relevant advice and organises periodic training sessions in management of data with a disclosure risks.

[1] Distribution License Agreement: <https://repo.clarino.uib.no/xmlui/page/contract>

[2] Deposited Item Lifecycle: <https://repo.clarino.uib.no/xmlui/page/item-lifecycle>

[3] Terms of service: <https://repo.clarino.uib.no/xmlui/page/about#terms-of-service>

[4] <https://www.nsd.no/en/data-protection-services> (NSD is now part of SIKT; services are still found at this link but may migrate to a different site in the future)

[5] CLARIN restricted licenses: <https://www.kielipankki.fi/support/clarin-eula/>

[6] CLARIN Legal Information Platform: <https://www.clarin.eu/content/legal-information-platform>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

Accept

#### 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:

## CLARINO Bergen Centre

The Repository is hosted by UiB, a medium-sized academic institution which is appropriate to the Repository's main audience of academic researchers and students. The University of Bergen is a member of the CLARINO consortium, which in turn is implementing Norway's contribution to the CLARIN ERIC. The CLARINO Bergen Centre is a certified centre type B in CLARIN ERIC [1], the governance of which strongly contributes towards guiding the CLARINO Bergen Centre in carrying out its mission.

Through the CLARINO+ contract with the Research Council of Norway and the consortium agreement with the partners, UiB has committed itself to updating and maintaining its part of the CLARINO infrastructure from 2020 until at least 2029. During the implementation phase until the end of 2023, UiB is receiving 6.9 M NOK from the RCN for this purpose. In the operational phase until the end of 2029, UiB has budgeted 7.3 M NOK of its own funds towards the centre.

The UiB Library, which is directly responsible for the operation of the Repository, is a highly stable unit with a long-term mission that includes the curation of digital data. IT-resources for the Repository are also provided by the University's IT Department through their involvement in NREC, and in providing storage (see R0). The Repository is located at the Library's Section for Education and Research Support, and is run by the Digital Development team there in cooperation with the Department of Linguistic, Literary and Aesthetic Studies. At the Digital Development team, the following highly experienced persons are assigned to CLARINO tasks: Chief Engineers Øyvind Liland Gjesdal and Tone Merete Bruvik, Senior Academic Librarian Rune Kyrkjebø and Senior Executive Officer Paul Meurer. Liland Gjesdal is responsible for technical maintenance, upgrades and development. He keeps in touch with LINDAT/CLARIN in Prague, which is the leading technical among the CLARIN centres running DSpace repositories. Liland Gjesdal has the main responsibility for co-operation with the University IT-department and the NREC support channels, and he provides the day-to-day maintenance and technical support for our repository. Bruvik represents a doubling of this essential function within the team. The two Chief Engineers conduct larger tasks, like software upgrades, together. Meurer maintains the additional services (Corpuscle, Menota, INESS and COMEDI) and their interaction with the Repository. Kyrkjebø is assigned by the library to work with CLARINO on the scientific and administration and user support levels. Administration of submissions is mainly handled by Kyrkjebø.

Academic Director Juliane Marie-Thérèse Tiemann heads the Digital Development team and coordinates the tasks based on regular status meetings where progress is reported and challenges are discussed. The team belongs to the Section for Education and Research Support, with Karin Cecilia Rydving as Section Head. Tiemann reports to Rydving. Rydving reports to the Library Director Maria-Carme Torras. They answer for the Library's strategy and its long-term and short-term commitments. All that are mentioned so far are permanent full-time staff, except Meurer who works 60% at the library and 20% at the faculty. Also readily available is the advice and assistance of Professor Koenraad De Smedt at the Department of Linguistic, Literary and Aesthetic Studies, with whom the library team works closely together. As National Coordinator and leader of the CLARINO project and the Bergen Centre, Prof. De Smedt also supervises the library team, and ensures that the library is always informed about CLARINO and CLARIN activities and possibilities to gain new knowledge, acquire contacts and share experience within the community.

The Repository website has contact information [2].

[1] Certificate for type B status: <http://hdl.handle.net/11372/DOC-116> This certification is up for renewal as soon as the CoreTrustSeal (which is part of the criteria) is granted.

[2] Contact information: <https://repo.clarino.uib.no/xmlui/page/contact>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

Accept

#### 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

## CLARINO Bergen Centre

### Response:

The Repository is networking in CLARIN ERIC, which is a large group of consortia [1]. It is managed, operated and maintained by members of the Digital Development team at the University of Bergen Library (see R5), in cooperation with the Department of Linguistic, Literary and Aesthetic Studies and with additional advice by the Norwegian School of Economics at Bergen. The members of the Digital Development team are employees selected for their high-level knowledge and skills and extensive experience in digital data management. The team is a part of the Section for Education and Research Support. The responsibilities for the Section for Education and Research Support are stated on internal, Norwegian webpages as the following:

- Guidance and instruction in information literacy and academic writing - Collection development and curation
- Open science: Digital publications and research data
- Bibliometry
- Copyright
- Development of digital services [the overall task of the Digital Development team]
- Metadata - Outreach

Some of the main digital resources run by the Digital Development team can be accessed from the link page "Writing and Research" at the University Library homepage [2]. The CLARINO Bergen Centre Repository is directly accessible from that page, as are several resources under "Digital Text Archives", and the "Manuscripts and Photographies" overview which leads on to the MARCUS homepage (direct link here: [3]). MARCUS is a Linked Open Data portal for the University Library's Special Collections, developed and run by the team.

The current University Library Staffing Plan ("Bemanningsplan\_endelig\_versjon\_151117" effectuated from 2018), which is an internal document, states tasks and competencies for the team. Among the tasks listed there, are (our translation):

- Development and running of infrastructure and digital services for teaching, research and outreach. - Co-operation on infrastructure and digital services for teaching, research and outreach.
- Work on encoded text resources using XML technologies.
- Work on semantic technologies.
- Development of web applications and free software.
- Provisioning, running and configuration/automation of servers

The staff regularly participate in scientific meetings at national and international levels where research community members are present, such as annual CLARIN conferences. They also participate in training and knowledge, sharing activities and CLARIN committee meetings. Bruvik is part of the SCCTC [3] group and remains in contact with the CLARIN DSpace development group. Gjesdal participates in the slack DSpace chat and plans to participate in the upcoming CLARIN-Dspace upgrade. Paul Meurer received the Steven Krauer Award for CLARIN Achievements in 2017.

The development team also follows and participates in activities outside of the CLARIN context. We subscribe, follow, and participate in the official DSpace mailing lists DSpace-tech and DSpace-community [5]. For other technical communities working on publishing data and science, we also host services using the technologies from the Public Knowledge Project and participate in their forums. In addition, the library also has a team focused on Open Science [6]. This team also participates in Research Data Alliance activities on the national level. In addition, the University of Bergen is a member of the European Open Science Cloud (EOSC) association [7]. The University library manages multiple other services in NREC and has the expertise to run and manage the servers.

Overall scientific coordination of the CLARINO Bergen Centre is in the hands of Prof. Koenraad De Smedt at the Department of Linguistic, Literary and Aesthetic Studies, who is a member of the CLARIN National Coordinator's Forum and makes sure that community relevance as reflected in CLARIN strategies percolate to that of the Repository.

The Repository staff provides feedback to all requests and submitters through email. In some cases, community members request personal discussions and guidance, in particular in the context of data management plans to be drafted at the start of new research projects. In cases of work-intensive data management, a contract with the depositing researcher of research group is drawn up, which may involve payment.

[1] <https://www.clarin.eu/content/participating-consortia>

[2] <https://www.uib.no/en/ub>

[3] <https://marcus.uib.no/>

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

[5] <https://dspace.lyrasis.org/join-the-community/>

[6] <https://www.uib.no/en/ub/111372/open-access-research-data>

[7] <https://eosc.eu/members/university-bergen-uib>

### Reviews

#### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

Accept

## CLARINO Bergen Centre

### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

I accept this but here I would have expected some hints on expert guidance by networking in the digital preservation communities. There are plenty of communities, that may support Clarino. I have the impression that you are networking mainly in the Clarin-context and that would be too little when you consider digital preservation issues

## 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 Repository is based on the CLARIN DSpace which is a fork of DSpace software [1]. To verify that a digital object has not been altered or corrupted, the repository uses MD5 checksums for all objects and checks these periodically.

This is done using cron-jobs using the dspace-cli [2] on different schedules to check integrity and the status of our repository. We also run a health check [3] weekly of the repository, and have scheduled recurring weekly tasks for going through the results. For more on the long-term preservation of the repository data and metadata which we check, see R10.

The Repository automatically performs regular checks on the integrity and the file formats of data. There is a list of supported and known formats whose consistency is regularly checked using existing tools (e.g., integrity testing of bzip format is done using bzip -t). A report is then sent to the editors and administrators who keep track of all used formats. If a new, more commonly used format emerges, the Repository will add it to the recommendation.

We adhere to the principles regarding standards and data formats are those formulated by CLARIN [4], but if the submitter has a good reason to use a non-open format, we accept that if properly documented. Items in the repository are normally only made web-accessible in the same format as submitted. All submitters must be authenticated through the federation of ID providers (eduGain and CLARIN Service Provider Federation). Actions regarding each Repository item by submitter and administrators are logged by the DSpace software and added to the DSpace metadata for the object. Examples of logged actions are the uploading of files, editing of metadata, approval of the submission and publishing of the submission. The logs are directly available for administrators.

Once deposited and archived, the submitted data sets can be changed neither by the submitter nor by editors. As stated in the Distribution License Agreement, within the repository no alteration of the submitted data will be made. This ensures that data is authentic and it is also important for the assigned persistent identifiers, which must always refer to the same content. This point is well understood by the administrators (Kyrkjebø, Bruvik, Gjesdal) and is stressed on our internal submission checklist (see R4). It is also explicated on the About-page (see Preservation Policy) [5]

Only the administrators of the Repository have the rights to make changes; these actions are logged (see above). We consider making minor changes to the metadata in case of errors; in this case, submitters should contact the helpdesk for requesting such changes as described in the "Deposited Item Lifecycle" [6]. For non-trivial changes, including changes to data formats, we require a new submission and will assign a new PID.

[1] DSpace: <https://duraspace.org/dspace/>

[2] <https://wiki.lyrasis.org/display/DSDOC6x/Command+Line+Operations>

[3] <https://wiki.lyrasis.org/display/DSPACE/Healthcheck+reports>

[4] CLARIN standards and formats: <https://www.clarin.eu/content/standards-and-formats>

[5] <https://repo.clarino.uib.no/xmlui/page/about#preservation-policy>

[6] Deposited Item Lifecycle: <https://repo.clarino.uib.no/xmlui/page/item-lifecycle>

## Reviews

### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

## CLARINO Bergen Centre

### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

Accept

### 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 documents well-defined procedures on how to deposit various collection types [1]. The repository mainly supports the following collection types to be submitted: corpora, lexical-conceptual resources, language descriptions, software, tools, web-services, and quantitative data from language surveys and experiments. The addition of other types is possible. Types corresponding to published CLARIN metadata profiles are recommended.

The Repository relies on the group of metadata standards around CMDI (ISO-CD 24622-1) [2]. The CMDI best practices guide promotes the inclusion of sufficient information to interpret the data in the context of long-term preservation [3]. Submitters are encouraged to provide a CMDI record based on a specific profile which is suitable for the submitted data type (e.g. corpus, lexical resource, etc.). In case the submitter does not upload a CMDI file, our administrator will ask for this to be provided. The automatically generated CMDI metadata exported by the DSpace application guarantees findability, but are not as rich in content as a tailored CMDI file normally would be. Part of our service is to offer assistance in generating a richer CMDI file using our tool COMEDI, which is accessible via a button on the Repository homepage, using the same federated login as for the Repository [4].

Each submitter must be authenticated through the federation of ID providers (eduGain and CLARIN Service Provider Federation), which ensures a high level of trust. No anonymous upload of items is allowed. In case of problems, Repository editors can retrieve the exact dates and users who have deposited the resources.

Submitters are referred to our FAQ page [5] which provides the list of data formats recommended by CLARIN [5]. When the submitted items contain attached bitstreams, metadata curators manually verify whether they meet the requirements of integrity, authenticity and availability. Metadata curators are always in contact with the depositor(s) in order to obtain missing information if there is any.

If the format is unknown or not in the list of the recommended standard formats [6], it must be well documented and the documentation must be either part of the submission or the metadata must contain a link to it. The final decision on acceptance/rejection of such submissions is taken by the Repository administrators in collaboration with the CLARINO metadata committee.

If the minimum formal requirements are not met during the submission process, the interface will not allow the user to complete the submission. After submission, the item is reviewed by metadata experts that will check for the quality of the metadata. Editors are also responsible of verifying that the data submitted are in a valid, correct and accepted format. However, as stated in the Distribution License Agreement, submitters are ultimately responsible for the quality of their own data. In case the submission does not comply with our expectations, the submission is returned to the data provider for correction.

Once deposited and archived, the submitted data cannot be changed neither by the submitter nor by editors. In case of severe problems e.g. regarding data quality or licenses, data can be retracted. If a dataset needs to be changed in any way, the submitter will create a new version which will be assigned a new persistent identifier while the old dataset will normally remain in place. This will also be the case if data needs to be migrated to a different format. Any new versions will be regarded as new datasets to which the entire submission procedure applies.

[1] How to Deposit: <https://repo.clarino.uib.no/xmlui/page/deposit>

[2] About CMDI: <https://www.clarin.eu/content/component-metadata>

CMDI schema: [https://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/profiles/clarin.eu:cr1:p\\_1349361150622/xsd](https://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/profiles/clarin.eu:cr1:p_1349361150622/xsd)

[3] Best practices guide: <https://www.clarin.eu/content/cmd-best-practices-guide>

[4] COMEDI: <https://clarino.uib.no/comedi/page>

[5] FAQ: <https://repo.clarino.uib.no/xmlui/page/faq?locale-attribute#what-submissions-do-you-accept>

[6] Standards and formats recommended by CLARIN: <https://www.clarin.eu/content/standards-and-formats>

Metadata Standards: <https://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf>

### Reviews

#### Reviewer 1:

#### Compliance level:

## CLARINO Bergen Centre

The guideline has been fully implemented in the repository - 4

### Comments:

Accept

### 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:

For our storage we have several types of storage in use. The application itself runs on ephemeral storage. Each image in NREC comes with storage which is meant to be ephemeral, i.e., the OS and the application. Secondly NREC offers creation of disk volumes [1] for persisting storage. We use volumes for the database storage and for persisting certificates and some configuration settings. However, these volumes do not come with any backup. Thirdly we use mounted network storage from the IT-department. This storage is used for database backups, the persisting parts of Dspace found on disk (Assetstore, SOLR index and logs) and AIP backups [3]. We can restore and recreate the content of the second type of storage, from the third. The third type is present for maintaining long term preservation, see the description of the third option translated to English[2]:

- Snapshots of the disk area are done hourly in the timeframe (9-17). Up to 30 versions are stored.
- There are also nightly snapshots. 14 versions are stored of these.
- All the storage areas are mirrored in Tromsø (for storage in Bergen – and vice versa) for fast recovery in case of disaster.
- There are nightly backups of all file areas – For up to 92 days you can get up to 7 versions of a file restored, which can be used for individual rollbacks. If a file is deleted, it stays in the backup for one year.

For the first step, the ingestion process, Submission Information Packages (SIPs) are received for curating and are assigned to a task pool where our curators can process them. The default ingestion process is done through our web-based interface. There is a number of pre-configured supported SIP formats (see [1]). For the second step, the archival storage, one of our curators takes charge of the submission, the metadata are updated (added, deleted, modified), and the submitted bitstreams are validated. In general, the curators ensure consistency and quality of each submission. If a curator approves an item, the Archival Information Package is available.

All backups follow standardized ways of using MD5 checksums for determining consistency. We apply two backup procedures described by Dspace [3]. First the traditional backup and restore – Database and files) which we run daily (files are placed on UiB storage. We schedule daily dumps of the database which are also stored on the UiB storage. Secondly, we use Archival Information Packages and restore [3]. We schedule AIP exports twice yearly.

The repository application source code is maintained at the UiB Git repository. The configuration of servers is set up by Ansible – a configuration management tool which is also stored in a separate git repository. NREC offers a Open Stack cloud infrastructure to run self-hosted services on. We've chosen Enterprise Linux (Rocky Linux 8) operating systems for their long-term support. Since the Repository does not store highly sensitive data the security measures are generally considered as sufficiently safe.

The above procedures are internally documented but due to their technical nature, they are not described at the same level of detail in the documentation on the Repository webpages.

[1] SIP formats: <https://wiki.duraspace.org/display/DSDOC18/Importing+and+Exporting+Content+via+Packages#Importing+and+Exporting+Content+via+Packages-Supported+Package+Formats>

[2] Storage (Norwegian only): [https://it.uib.no/Standard\\_Fillagring\\_Med\\_Tiering](https://it.uib.no/Standard_Fillagring_Med_Tiering)

[3] AIP backup <https://wiki.lyrasis.org/display/DSDOC5x/AIP+Backup+and+Restore>

### Reviews

### Reviewer 1:

### Compliance level:

The guideline has been fully implemented in the repository - 4

## CLARINO Bergen Centre

### Comments:

Accept

### 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 digital assets of the repository will be long-term preserved as part of the UiB digital assets. The cloud infrastructure is interchangeable and could be replaced by other providers. The IT department guarantees persistence of data and regular backup. The repository is covered by University of Bergen IT-department backup routines, security measures and crisis management for storage, complying to ISO/IEC 27001 (see also R9 and R16 for details).

For each item in the Repository, the CLARINO Bergen Centre has a Distribution License Agreement (see also R2). The default agreement is an easy-to-understand text [1] giving the CLARINO Bergen Centre the non-exclusive right to store the item and distribute it under specified conditions. This agreement does not imply transfer of ownership. The agreement does not by default specify a time period or limit. Thus the Repository may store and distribute the data for as long as it wants, and retract it if wishes to do so. The agreement gives the Repository the right to move the data to new media as necessary for long-term preservation. As mentioned in R9, we store AIP-packages on backed up volumes, as an extra measure for long term preservation. The repository also has the right to convert the data to new formats; such conversions (resulting in new items) have until now not been necessary, since the Repository promotes widely accepted open formats which tend to stand the test of time. In case of changing formats we would follow the same procedure as we do updating other items in DSpace, eg. that the old item will still be available, but also with a "this item has been superseded" note and a link to the current version. Relevant actions during long-term operation are automatically logged by the DSpace system.

In terms of community embedding, the Repository's strong link to CLARIN and its common strategy aimed at open data, open standards and formats, and open software maintained by an international group of CLARIN centres is a major factor promoting longevity of data access. The core of the plan for managing the function of long-term preservation is therefore to maintain this strong link to CLARIN and the community behind this large European research infrastructure. Participation in CLARIN committees and working groups is part of this plan.

In order to realize this data preservation plan, the UiB Library is committed to spend the necessary efforts to sustain long term access to and preservation of the data contained in the Repository (see also R3 and R5).

The Repository preservation policy is documented at [2].

[1] Distribution License Agreement (also called "Deposition License Agreement"): <https://repo.clarino.uib.no/xmlui/page/contract>

[2] <https://repo.clarino.uib.no/xmlui/page/about#preservation-policy>

### Reviews

#### Reviewer 1:

#### Compliance level:

Not Applicable - 0

### Comments:

#### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

Accept



## CLARINO Bergen Centre

### 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:

The CLARINO Bergen Repository has experts who are able to address technical data and metadata quality (see R5 and R6 for details).

Every submitted item undergoes technical assessment of data metadata quality by experts affiliated with the CLARINO Bergen Centre. Automatic tools are helping the editors to verify and validate metadata and the integrity of the submitted data [1][2]. The basic curation process foresees the possibility of returning the submission to the submitter for additional changes before the dataset is published in the repository. On some occasions the repository assists in enhanced curation, for example, conversion to a different format, enhancement of metadata, or customization of licenses.

The metadata policy ensures that sufficient information is available for end users (see R8 for details). It is, however, ultimately up to end users to assess to which extent the data suit their own goals. The Repository does not provide any guarantees regarding data content or suitability, since this is the responsibility of the rights holders. Neither does the Repository require that sufficient instructions are provided to use certain data for certain purposes.

The CLARINO Bergen Centre provides an online tool called COMEDI [3] for creating, modifying and validating CMDI-conformant metadata [2]. In this way we help users/depositors of our repository and metadata curators to ensure that resource quality is maintained. The use of COMEDI is however not required.

The Repository does not provide links to related works. The portal CLARIN Virtual Language Observatory (VLO) [4] harvests the CMDI metadata from the CLARINO Bergen Repository, together with that from other Repositories, and can provide lists of items that match faceted search criteria.

The Repository has contact information for reporting issues. Furthermore, in the VLO overview, i.e. when browsing or searching for relevant resources, and on the individual resource pages displaying further information on a specific resource, the user can report an issue or give feedback on metadata records or resources using a ticketing system. The helpdesk agents can thus forward issues directly to our Repository for correcting and/or harmonizing metadata if needed.

Data and metadata quality is also promoted through guidance to users through lectures and training events (see R6). We also encourage users to attend CLARIN training courses and other user involvement actions [5].

[1] CLARIN DSpace metadata info: <https://github.com/ufal/clarin-dspace/wiki/Metadata-info>

[2] DSpace curation system: <https://wiki.duraspace.org/display/DSDOC5x/Curation+System>

[3] COMEDI: <https://clarino.uib.no/comedi>

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

[5] CLARIN user involvement actions: <https://www.clarin.eu/content/clarin-for-researchers>

### Reviews

#### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

Accept

#### 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:

## CLARINO Bergen Centre

The guideline has been fully implemented in the repository - 4

### Response:

The Repository has a well-defined and documented workflow for the curation process and a procedural documentation for archiving of data [1]. The repository also documents deposit life cycle for creating, editing or deleting resources [2]. The workflow is managed by Rune Kyrkjebø and his colleagues at the Section for Digital Development of the UiB Library (see also R5 for more detail on the organizational infrastructure).

Ingest includes appraisal and selection of data by the Repository editors in accordance with the mission of the repository [3]. In certain cases, requests for the deposit of valuable but unsuitable data may be deflected to other repositories.

The Repository distinguishes between known submitters and unknown ones, where submissions from the latter will be specially validated and verified. In the day-to-day running of the repository, the University of Bergen Library relies upon linguistic advice from CLARINO partners when needed in connection with metadata or licensing questions. The necessary skills for the workflow to function is provided on a long-term basis.

The workflow includes checks for sensitive data (e.g. containing personal identifying information). In general the Repository refrains from accepting personal identifying data (see also R4), with the exception of sign language, for which participant consent has been obtained and which is released under a stringent restricted license. Sufficient levels of security are in place for handling non-sensitive data in the workflow (see also R9).

In case of migration, we would follow the guidelines for replacing the old item, where the old item retains its handle and link, but will be updated with a replaced by annotation. We have not had any data migrations so far.

If the bitstream check fails, we will update and sync our test-server with data and storage from the production server and try to replicate the failing bitstream check there. We will then attempt to restore assets which report the bitstream error in the test server and rerun the check to see if this fixes the bitstream error. If there are discrepancies, we will go back in snapshots until we have a matching checksum. If we are unable to resolve the issue with these steps, we also have AIP backups, which can be used for restoring one or more items, and can be used as a fallback strategy for restoration. We have not had any incidents on bitstream checking so far.

Submissions that do not meet all criteria, e.g. due to inconsistent formats, incomplete or invalid metadata, missing or inappropriate license, personal data issues, etc. are rejected.

[1] Deposit procedure: <https://repo.clarino.uib.no/xmlui/page/deposit>

[2] Item Lifecycle: <https://repo.clarino.uib.no/xmlui/page/item-lifecycle>

[3] Mission statement: <https://repo.clarino.uib.no/xmlui/page/about#mission-statement>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

Accept

#### 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:

The items in the Repositories are findable in several ways. The Repository is browsable and searchable with keywords and facets. Metadata on all resources in CLARINO Bergen Centre can be harvested by authorized systems through OAI-PMH endpoint. They are regularly harvested and findable through CLARIN Virtual Language Observatory (VLO) which supports keywords and faceted search [1]. They are also harvested into a national catalog maintained by the National Library of Norway and can be discovered there [2].

## CLARINO Bergen Centre

We assign persistent Identifiers (PIDs) in the form of Handles to all items by the repository, allowing for reliable citation. The Repository generates full bibliographical references on the fly, both in textual format (author/year style) and in BibTex.

Access to data is subject to registration and authentication of the user, with the exception of non-restricted, public data.

Submitters of data in a non-recommended format will be encouraged to convert the data to a recommended format. Administrators will ask for this to be done before approval of the submission. Our staff will assist with advice on conversion of data. If time and resources allow it, we might offer to assist with the data conversion itself. This is stated in our internal submission checklist and on the Deposited Item Lifecycle-page [3].

Resources are provided by the community and to the community in commonly used formats. As stated in our FAQ, "We accept any linguistic and/or NLP data and tools: corpora, treebanks, lexica, but also trained language models, parsers, taggers, MT systems, linguistic web services, etc." [4].

[1] VLO: <https://vlo.clarin.eu/>. An example of a record resulting from a search in the VLO and describing an item in the Repository is the following: [https://vlo.clarin.eu/record?4&count=3&docId=http\\_58\\_47\\_47\\_hdl.handle.net\\_47\\_11509\\_47\\_124\\_64\\_format\\_61\\_cmdi](https://vlo.clarin.eu/record?4&count=3&docId=http_58_47_47_hdl.handle.net_47_11509_47_124_64_format_61_cmdi). This record contains a handle referring to the landing page in the Repository.

[2] Catalog at the National Library of Norway:

<https://www.nb.no/sprakbanken/#ticketsfrom?lang=en&query=alle&tokens=&from=1&size=12&collection=clarino>

[3] Deposited Item Lifecycle (under Edited Item): <https://repo.clarino.uib.no/xmlui/page/item-lifecycle>

[4] FAQ: <https://repo.clarino.uib.no/xmlui/page/faq#what-submissions-do-you-accept>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

Accept

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

Compliance Level Comment: ok

### 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 CLARINO Bergen Centre has been established with data reusability as one of the main components in mind. It provides clear information about what metadata are required and how the data and metadata are disseminated [1]. Depositors can fill metadata based on the profiles for specific resource types so that appropriate information is given to support the understanding and use of the data. The Repository does, however, not provide any guarantees for suitability of the data for any purposes of reuse.

The Repository metadata strategy is based on the CLARIN metadata strategy which supports the understanding and reuse of the data (see R8 for details). All metadata comply with CMDI profiles/schemas in the Component Registry [2] and all of them map to Dublin Core. The repository supports several protocols for data sharing and reusability such as OAI-PMH, OAI-ORE and offers resources to the community in different metadata formats such as CMDI, OAI\_DC, HTML, BIBTEX, ORE etc. [3].

The strong embedding of the Repository activities within the CLARIN community, including participation in the CLARIN CMDI Taskforce and use of the CMDI component registry, promotes longevity of the metadata strategy, ensuring the validity of the metadata over time and allowing for further developments, e.g. the ongoing migration from CMDI 1.1 to CMDI 1.2 which is currently carried out in CLARINO+. However the current status of CMDI 1.1 metadata and the alternative metadata profiles we offer, are sufficient for reuse and understanding.

## CLARINO Bergen Centre

The Repository is open to adoption of future standards and formats and to further development of the CLARIN DSpace system thanks to cooperation with a strong community of developers and users all over the CLARIN centres in Europe. This makes it easier to incorporate any features based on future requirements. As an example, we mention that CMDI profiles were originally not supported in DSpace out of the box but were later incorporated in the system by the CLARIN developer community. The Repository supports submission of new versions of the resources in case there is a need for enriching or reformatting submitted datasets.

Several academic courses and researcher training courses have provided instruction and demonstration of reuse of data in the Repository. Repository staff participate in giving courses and presentations developed and run within the CLARINO+ project [4].

[1] Metadata: <https://repo.clarino.uib.no/xmlui/page/metadata>

[2] CLARIN Component Registry: [https://catalog.clarin.eu/ds/ComponentRegistry?registrySpace=published&itemId=clarin.eu:cr1:p\\_1403526079380](https://catalog.clarin.eu/ds/ComponentRegistry?registrySpace=published&itemId=clarin.eu:cr1:p_1403526079380)

[3] List of metadata formats: <https://repo.clarino.uib.no/oai/request?verb=ListMetadataFormats>.

[4] For example, a two hour course 21th of May 2021 at the University of Bergen in VLO-search and CLARINO Repository search and deposit of data.

### Reviews

#### Reviewer 1:

##### Compliance level:

The repository is in the implementation phase - 3

##### Comments:

Accept

#### 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:

The basic middleware for the Repository is CLARIN DSpace [1], which is in use by several repositories in CLARIN member countries, which means that a large European research community is familiar with its purpose, features and interface.

This software is a modification of the generic DSpace [2], a full-fledged turnkey institutional repository platform. It was modified by the LINDAT/CLARIN centre [3] at the Charles University in Prague in order to comply to CLARIN specs and requirements regarding standards, formats, authentication, etc. [4]. The CLARIN DSpace software is currently supported, documented and widely used by a community of CLARIN technical centres who produce regular improved versions, maintain a wiki [5] and organize development workshops [6]. The CLARINO Bergen Centre plans to continue its participation in this further development and will regularly upgrade to improved and tested versions of this software.

There is an ongoing process of upgrading the CLARIN DSpace fork to the latest DSpace which can be found on GitHub at DSpace [7] and frontend [8]. This is in regard to the upcoming end of life of DSpace 5 [9] which is reached in 2023. We will participate in the upgrade process as soon as the development process is mature enough for trying out an installation.

At the CLARINO Bergen Repository, this software runs on NREC, on a Rocky Linux 8 instance. The server hardware is placed in a secure air-conditioned server room with restricted access. Permanent connectivity to the internet is provided in the Norwegian academic network maintained by SIKT, which also acts as the Norwegian NREN [11]

For information on backup and long-term storage see R9.

The roles and responsibilities of the IT-department relative to the other departments ("system owners") are defined in the University of Bergen statutes [10]. This obviates the need for Service Level Agreements with each service. See also R15.

## CLARINO Bergen Centre

- [1] CLARIN DSpace <https://github.com/ufal/clarin-dspace>
- [2] DSpace: <https://duraspace.org/dspace/>
- [3] LINDAT/CLARIN: <https://lindat.mff.cuni.cz/repository/xmlui/>
- [4] CLARIN standards and formats: <https://www.clarin.eu/content/standards-and-formats>
- [5] CLARIN DSpace wiki: <https://github.com/ufal/clarin-dspace/wiki>
- [6] CLARIN DSpace workshop: <https://www.clarin.eu/blog/clarin-dspace-workshop>
- [7] <https://github.com/dataquest-dev/DSpace>
- [8] <https://github.com/dataquest-dev/dspace-angular>
- [9] <https://wiki.lyrasis.org/display/DSPACE/Support+for+DSpace+5+and+6+is+ending+in+2023>
- [10] Statutes (most relevant parts are in Norwegian only): <https://regler.app.uib.no/regler/Del-4-OEkonomi-eiendom-og-IKT/4.3-Informasjons-og-kommunikasjonsteknologi/Styringssystem-for-informasjssikkerhet-og-person>
- [11] Uninett NREN is now part of SIKT: <https://sikt.no/>

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

Accept

#### 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:

UiB has documented procedures for information security and personal protection [1] consisting of goals, rules and management procedures which also govern the Repository. They are anchored in laws, regulations and management responsibilities as defined in the organization of the university, with the University Board as the highest responsible. All activity must comply with various Norwegian laws regarding personal data protection (including the GDPR), electronic communication, transparency, archiving, special protection, copyright, higher education, health research, health registers, and health personnel, and must also comply with the general UiB regulations. Confidentiality must be guaranteed in accordance with all these laws and regulations. Personal information must not be processed to a larger extent than necessary and must be removed when no longer needed for defined academic purposes. Personal data should be correct. Everyone who processes personal data must be known with relevant rules and requirements.

The framework of its IT policy is within the standard ISO/IEC 27001: 2013 [2]. This standard is used to identify all domains of security which are applicable to the university's IT systems including those of the Repository. The standard is continually operationalized by the IT-department. The operative responsibility for IT security is with the IT director.

In line with the regulations, the Repository has a documented privacy policy [3] describing its jurisdiction, the personal data processed, the purpose of processing personal data, how to access, rectify and delete personal data, data retention and the data protection code of conduct. The policy also states that no personal data are disclosed to anyone outside of the Repository team. Personal data such as name and email are used to allow user authentication and authorization, to log the signing of licenses, to contact the user if necessary, and for monitoring and technical issue solving. Personal data are also used (in aggregated form only) for statistics on usage by groups in society, as required in reports by the Research Council of Norway.

Secure protocols (mainly https) are used for all communication with the system.

For information on backup procedures, see R9.

The IT department follows an internal action plan that is executed in case of major security breaches. The IT department regularly runs crisis management exercises, for example to quickly remedy the situation should a machine crash occur. In such situations a crisis management group enters

## CLARINO Bergen Centre

into force. There are also established routines for reporting security incidences to the relevant authorities.

Through automation with Ansible and recovery of data from the IT-department, we can restore systems and data in case of problems. The effectiveness was proven in 2021 during a power outage which disrupted our disk storage.

Downtime is monitored and documented at CLARIN using the Icinga service, where also scheduled downtime can be registered [4].

[1] UiB security policy (in Norwegian):

<https://regler.app.uib.no/regler/Del-4-OEkonomi-eiendom-og-IKT/4.3-Informasjons-og-kommunikasjonsteknologi/Styringssystem-for-informasjonsikkerhet-SIP-Del-1>

[2 ] UiB security policy, executive part (in Norwegian). The ISO/IEC 27001: 2013 standard is used. See:

<https://regler.app.uib.no/regler/Del-4-OEkonomi-eiendom-og-IKT/4.3-Informasjons-og-kommunikasjonsteknologi/Styringssystem-for-informasjonsikkerhet-og-person>

[3] Privacy policy: <https://repo.clarino.uib.no/xmlui/page/privacypolicy>

[4] <https://monitoring.clarin.eu/monitoring/service/show?host=CLARINO%20Bergen&service=OAI-PMH-31> (requires CLARIN login).

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

Accept

#### 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:

##### 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

## CLARINO Bergen Centre

**Comments:**

A very good reapplication. One of the better I have seen.