Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository CLARINO Bergen Repository complies with the guidelines version 2014-2017 set by the Data Seal of Approval Board.
The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2013 on November 3, 2015.

The Trusted Digital repository is allowed to place an image of the Data Seal of Approval logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the Data Seal of Approval website.

Yours sincerely,

The Data Seal of Approval Board

Assessment Information

Guidelines Version: 2014-2017 | July 19, 2013
Guidelines Information Booklet: DSA-booklet_2014-2017.pdf
All Guidelines Documentation: Documentation

Repository: CLARINO Bergen Repository
Seal Acquiry Date: Nov. 03, 2015

For the latest version of the awarded DSA http://assessment.datasealofapproval.org/seals/
for this repository please visit our website:

Previously Acquired Seals: None

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

T +47 55 58 00 00
E rune.kyrkjebo@uib.no
W http://uib.no/ub

Data Seal of Approval Board

W www.datasealofapproval.org   E   info@datasealofapproval.org

2

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

The CLARINO Bergen Repository is a repository for language-based resources and advanced tools available at:

clarino.uib.no

Our repository is part of the Norwegian CLARINO project which aims to implement CLARIN infrastructure in Norway. More specifically our repository is part of the Bergen Centre of the CLARINO project. A news blog about the project can be read here:

http://clarin.b.uib.no/

The University of Bergen has contractually defined the contribution of the University of Bergen Library (UBL) to the CLARINO project. Our contribution focuses on establishing and running the CLARINO Bergen Repository.

Our library's role is specified in the CLARINO description of work, where it is stated that "the Bergen University Library [...] will define and test its role as an institutionwide repository for research data. It has newly created a Section for Digital Systems and Services which will participate in CLARINO", see page 6 in the DOW:

https://clarin.b.uib.no/files/2013/08/clarino-dow2.pdf

The aim of our repostiory is thus to provide for language resources to "be maintained and made accessible by appropriate interfaces in a well-structured repository system with a long-term commitment and support for metadata harvesting and tool integration" (CLARINO DOW p. 6), and thereby help fulfill the specification of a CLARIN type B Centre for the Bergen CLARINO Centre.

Like LINDAT/CLARIN does, and using the same technology, our repository implements Persistent Identifiers, authorisation and authentication, and sharing of data and metadata. Data harvesting according to OAI-PMH is implemented.

Our repository is open for self-deposit by authorised users, by a doumented procedure:

https://repo.clarino.uib.no/xmlui/page/deposit

Submissions will be reviewed. The UBL functions as editor, together with CLARINO staff here at Bergen University.

In compliance with CLARIN specifications, the LINDAT/CLARIN DSpace implementation includes the possibility of uploading CMDI metadata files. This procedure is currently an administrator function, done by UBL and CLARINO staff. Our repository homepage links to a CMDI metadata editor, open to CLARIN, EduGain or Norwegian academic (FEIDE) authenticated users:

http://clarino.uib.no/comedi/page

Our repository software is based on the software provided by the LINDAT/CLARIN DSpace repository that was developed by The Institute of Formal and Applied Linguistics at Charles University, Prague, and which is openly shared:

https://github.com/ufal/lindat-dspace

https://lindat.mff.cuni.cz/en/

http://ufal.mff.cuni.cz/

The LINDAT/CLARIN repository has obtained the Data Seal of Approval.

The CLARINO Bergen Repository was implemented by University of Bergen Librarys own staff, together with University of Bergen IT Department staff, guided by LINDAT/CLARIN staff that visited here on a CLARIN mobility grant.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this as implemented.

The depositor is required to comply with the following:

"Data have to be provided with metadata in standard formats accepted/adopted in the respective communities, persistent identifiers (PIDs) have to be assigned, IPR issues have to be resolved and clear statements with regard to licensing and possible use of the resources are to be made. The depositor is also required to electronically sign a deposition agreement acknowledging the (s)he is the holder of rights to the data and that (s)he has the right to grant the rights contained in this licence."

https://repo.clarino.uib.no/xmlui/page/faq

The Terms of Service are stated explicitly:

"To achieve our mission statement,we set out some ground rules through the Terms of Service. By accessing or using any kind of data or services provided by the Repository, you agree to abide by the Terms contained in the above mentioned document.

Data in the CLARINO repository are made available under the licence attached to the resources. In case there is no licence, data is made freely available for access, printing and download for the purposes of non-commercial research or private study. Users **must** acknowledge in any publication, the Deposited Work using a persistent

identifier (see Citing Data), 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 explicitaly granted by the attached licence without formal permission of the copyright holders."

https://repo.clarino.uib.no/xmlui/page/about#terms-of-service

Also, the user "agrees 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."

https://repo.clarino.uib.no/xmlui/page/terms-of-service

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this as implemented.

The data producer is encouraged to use one of the recommended formats mentioned in LRT:

"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. We do not strictly require you to upload the data itself, although it is always better to do it. Still, you can make a metadata-only record, if required. We also support online license-signing for immediate availability of restricted resources. When uploading language resources, please try to use one of the recommended formats mentioned in LRT Standards."

https://repo.clarino.uib.no/xmlui/page/faq#what-submissions-do-you-accept

http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

3. The data producer provides the data together with the metadata requested by the data repository.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this as implemented.

The deposit procedure requires the data producer to enter metadata as part of the deposit procedure:

https://repo.clarino.uib.no/xmlui/page/deposit

Required fields to enter are:

- type of submission (choice)

- title

- author(s)

- publisher

- date

- contact person and institution

- description, language(s), keywords, size and media type

- distribution and license information.

Metadata can also be imported from other repositories.

Metadata quality is reported to UBL repository administrator automatically weekly by a server installed validator. Like LINDAT/CLARIN, we remove submissions with invalid or missing metadata and ask the data producer to improve the metadata before republishing the item.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

4. The data repository has an explicit mission in the area of digital archiving and promulgates it.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this as implemented.

The repository shares the mission statement of the CLARIN ERIC:

"The ultimate objective of CLARIN ERIC is to advance research in 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 shall be 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 (henceforth abbreviated HSS) research communities at large."

https://repo.clarino.uib.no/xmlui/page/about#mission-statement

The CLARINO Bergen Repository is a dedicated part of the Norwegian CLARINO and the international CLARIN infrastructures. It is hosted and maintained at the University of Bergen by the The IT-department, The Department of Linguistic, Literary and Aesthetic Studies and The University Library.

We plan to officially launch our repository the 15th of October this year in connection with a poster presentation at the annual CLARIN conference which takes place in Wroclaw, Poland.

We promote our repository in the CLARINO network and locally at Bergen University and to the Bergen CLARINO partners.

Reviewer Entry

Accept or send back to applicant for modification:

   Accept

Comments:

5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The repository is an official part of UBL and thereby a part of The University of Bergen, which is a legal entity.

The repository requires the depositor of data or tools to sign a Deposition License Agreement, which specifies that they have the right to submit the data and gives us (the repository) right to distribute the data on their behalf. This means that depositors are solely responsible for taking care of IPR issues before publishing data or tools by submitting them to us.
Should anyone have a suspicion that any of the datasets or tools in our repository violate Intelectual Property Rights, they should contact us immediately at our help desk.

https://repo.clarino.uib.no/xmlui/page/about#about-ipr

The repository states the following privacy policy:

https://repo.clarino.uib.no/xmlui/page/privacypolicy

The repository complies with the Data Protection Code of Conduct:

http://www.geant.net/uri/dataprotection-code-of-conduct/v1/Pages/default.aspx

The repository requires data users to comply with the following data citing policy, "Data Users must acknowledge and cite data sources properly in all publications and outputs. To make reference to resources deposited in our repository, use a handle as a persistent identifier instead of an URL."

https://spaced.uib.no/xmlui/page/about#citing-data-policy

https://repo.clarino.uib.no/xmlui/page/about#citing-data-policy

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

6. The data repository applies documented processes and procedures for managing data storage.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this as fully implemented.

Data storage is done in a DSpace repository application which is part of the authorised University of Bergen IT infrastructure for computing and data storage.

Our CLARINO database is backed up daily by the central IT department. The resource folder containing the deposited items is backed up daily. We also have a daily snapshot taken of the installation itself, going back at least 2 weeks.

The repository application sourcecode is maintained in the University of Bergen git-repo, which is backed up.

The University of Bergen runs the server software RedHat which has a long term support and is strongly security updated.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. The data repository has a plan for long-term preservation of its digital assets.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this to be in practice fully implemented, although we seem for the time being to be in lack of a public policy statement for our university that explicitly adresses long-term data usability and future obsolence of file formats.

The digital assets of the repository will be long-term preserved as part of the UBL/University of Bergen digital assets.

An additional option exists as a plan is being made for long term backup copies stored by the National Library of Norway acting as a national CLARIN center type A.

http://www.clarin.eu/sites/default/files/CE-2012-0037-centre-types-v07_0.pdf

The University of Bergen RedHat server software has long term support. The IT department guarantees persistence in operating system and file systems, which also ensures the possibility for migration of data.

The University of Bergen highest level information system security policy states that the framework for University of Bergen policy in this field is the standard NS-ISO/IEC 17799: Norwegian text: "Som rammeverk for sikkerhetspolitikken er standarden NS-ISO/IEC 17799 benyttet og gjelder som utfyllende bestemmelser så langt formuleringene passer på forholdene ved UiB".

http://regler.app.uib.no/regler/Del-4-OEkonomi-eiendom-og-IKT/4.3-Informasjons-og-kommunikasjonsteknologi/Overordnet-IKT-sikker

The standard is presented here: http://www.iso.org/iso/catalogue_detail?csnumber=39612

The information system security policy also states how the responisbility is taken by university units and the IT-department to secure that data systems and the information in the systems can be retrieved at all times, Norwegian text: "Driftsenhetene i samråd med systemeiere har ansvaret for å sikre at systemet og informasjon i systemet kan gjenopprettes."

http://regler.app.uib.no/regler/Del-4-OEkonomi-eiendom-og-IKT/4.3-Informasjons-og-kommunikasjonsteknologi/Overordnet-IKT-sikker

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

A public policy statement that explicitly addresses long-term data usability and future obsolence of file formats would be a good addition.

8. Archiving takes place according to explicit work flows across the data life cycle.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We are in the implementation phase of the workflow for the CLARINO Bergen Repository.

The repository is run by the UBL in cooperation with linguistic staff at CLARINO/University of Bergen. Like LINDAT/CLARIN, we distinguish 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 UBL 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 repository has a curation process and a procedural documentation for archiving of data.

https://repo.clarino.uib.no/xmlui/page/deposit

The repository also documents life cycle questions like editing / modifying or deleting data:

https://repo.clarino.uib.no/xmlui/page/item-lifecycle

There will probably not be many data submitted that are not part of the mission of CLARINO. But if/when that happens, the UBL has other repositories available, and the CLARINO partners will be able to decide together with us where to direct such submissions.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this to be implemented.

The author of the work remains the proprietor. UBL assumes responsibility for access and availability of the repository in accordance with CLARINO.

Data producers have to state a license as part of the submission procedure. If there is need for a license that is not provided in our procedure, a new license can be designed and put in place by repository administrators.

The repository is covered by University of Bergen IT-department backup routines and crisis management (see also 7 and 9). The risk of data being lost due to minor or major crises is very low.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. The data repository enables the users to discover and use the data and refer to them in a persistent way.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this to be implemented.

The repository is browsable and searchable. Data will also be findable through CLARIN Virtual Language Observatory. References can be persistently made to PIDs provided by the repository. The repository also generates full bibliographical references on the fly.

It is a strength of CLARIN/CLARINO that data is provided by the community in currently used formats. As stated, "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."

https://repo.clarino.uib.no/xmlui/page/faq#what-submissions-do-you-accept

The repository is stable.

PIDs are implemented. We run a handle server which comes with the DSpace software.

OAI harvesting is permissible and is implemented.

There are advanced search facilities.

https://repo.clarino.uib.no/xmlui/discover?advance

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

11. The data repository ensures the integrity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this to be implemented.

MD5 checksums are utilised by the DSpace software that the repository runs on. The software also monitors data and metadata integrity.

A versioning system is not implemented. Submissions can be withdrawn. PIDs will still exist for the original submission and metadata indicate that it is replaced by a new version.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this to be implemented. Before admission of a digital object, the repository expects relevant information about provenance and relations between data sets to be explicitly stated.

Our item lifecycle information makes data producers aware of our strategy for data changes:

https://repo.clarino.uib.no/xmlui/page/item-lifecycle

Provenance metadata are stored in log messages for every change.

Data providers can not change the submitted files or metadata without contacting repository administrators. We do not encourage deletions or editions of submitted material, but are open to minor edits (cfr. item lifecycle page).

Depositors are only persons that are identified and authorised by the federated log-on providers.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

Identical to LINDAT/CLARIN our repository is based on the DSpace technical infrastructure and committed to the CMDI component metadata model (ISO-CD 24622-1):

http://registry.duraspace.org/about

http://www.iso.org/iso/catalogue_detail.htm?csnumber=37336

DSpace is fully implemented, and we consider that our repository explicitly supports the tasks and functions described in OAIS. The details are the same as for the LINDAT/CLARIN repository of which ours is a clone (cfr. LINDAT/CLARIN DSA Self Assessment document - https://assessment.datasealofapproval.org/assessment_92/seal/pdf/):

1) Ingestion - DSpace receives Submission Information Packages for curating.The default way is through the web based interface.

2) Archival storage - Repository administrators updates metadata and validates the submission. When the administrator has approved the item, then the Archival Information Package becomes available.

3) The Data Management function is executed during the creation of the metadata (descriptive, administrative, structural).

4) Preservation planning: The repository is monitored and backed up. See 6) and 9).

5) Administration: Data producers must be authorised and authenticated before submitting data. The repository is open to all submissions which meet our standards. A contract is signed during the ingestion process electronicall by accepting the Terms Of Service.

6) All metadata are publicly available. Some submissions require authenticated access, which is granted to academic users of the whole CLARIN community. DSpace allows for searching, locating and description of the information stored.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

14. The data consumer complies with access regulations set by the data repository.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this to be implemented.

Access regulations are based on federated log-on systems, by which consumers are authenticated. In case of stricter licences for certain data sets, a contract is provided in the form of click-through acceptance of licenses. Acceptance of licenses is logged by DSpace and available for administrator inspection.

Each submission is clearly marked with its license.

We do not monitor actively how consumers use the downloaded data. But if such issues should arise, we can provide the download details including time and date, identity of consumer and the documented acceptance of the licence.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this to be implemented.

The data consumer pledges to conform to and agree with the general codes of conduct that are accepted by the data consumer when she is granted access within a national federated log-on system for the academic sector.

Data providers are required explicitly to ensure that IPR and personal rights are respected in their data.

The license of each item is clearly stated.

The ethical terms of service are also clearly stated by the repository:

https://repo.clarino.uib.no/xmlui/page/terms-of-service

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

16. The data consumer respects the applicable licences of the data repository regarding the use of the data.

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

We consider this to be implemented.

The data consumer is expected to respect the general codes of conduct, as stated above. The data consumer is further expected to explicitly acknowledge any stricter license that may be applicable to certain data.

We base access regulations on international standarda as much as possible, and provide for the use of CC licences.

Signed licenses are stored and can be retrieved in case of any irregularity.

It is possible for us to close the repository for users guilty of misuse.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: