



Assessment Information

[CoreTrustSeal Requirements 2017–2019](#)

Repository: CLARIN Centre at the University of Copenhagen
Website: <https://repository.clarin.dk/repository/xmlui/>
Certification Date: 27 July 2019

This repository is owned by: **University of Copenhagen**



The CLARIN Centre at the University of Copenhagen

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository, National repository system; including governmental

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of Repository

The CLARIN Centre at the University of Copenhagen [1], Denmark, hosts and manages a data repository, CLARIN-DK-UCPH Repository [2], which is financed by the University of Copenhagen, and a part of the national infrastructure collaboration DIGHUMLAB in Denmark [3].

Our archive is part of the European Research Infrastructure Consortium CLARIN ERIC [4]. In 2012 CLARIN ERIC was established and took up the mission to create and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences. Currently CLARIN provides easy and sustainable access to digital language data (in written, spoken, or multimodal form) for scholars in the social sciences and humanities, and beyond. CLARIN also offers advanced tools to discover, explore, exploit, annotate, analyse or combine such data sets, wherever they are located. This is enabled through a networked federation of centres: language data repositories, service centres and knowledge centres, with single sign-on access for all members of the academic community in all participating countries. The CLARIN Centre at the University of Copenhagen hosts and manages the CLARIN-DK-UCPH Repository which serves as one of these nodes. In addition CLARIN-DK is a knowledge centre for Danish language technology and Danish language resources.

[1] The CLARIN Centre at the University of Copenhagen: <https://info.clarin.dk/en/the-clarin-dk-infrastructure>

[2] CLARIN-DK-UCPH repository: <https://repository.clarin.dk/repository/xmlui/>

[3] DIGHUMLAB (Digitalt Humaniora Laboratorium): <https://dighumlab.org>

[4] CLARIN ERIC: <http://www.clarin.eu>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of the Repository's Designated Community.

The designated community of the repository comprises scholars in the humanities and social sciences from disciplines where language plays a central role. The community can be subdivided into data producers and data consumers where the former are encouraged to deposit research data and their corresponding research material (documentation and annotations) in the repository. The latter comprises students, linguists, philologists, historians etc.

For the data producers the centre provides data management consultation and support for preparing of information packages to be deposited in the repository. The centre suggests the use of open standards and open access to data objects. The centre ensures preservation of data according to the definition of Preservation Description Information (PDI) as documented in the OAIS reference model [1].

For the data consumers the centre offers advanced tools to explore, exploit, annotate, analyse and combine data resources, together with knowledge sharing about the Danish language and tools for text analysis. Furthermore, the centre offers workshops for the data users. This is regarded as extended support to the users of the repository.

The repository service is designated for academic and non-commercial use, and currently, the repository stores only non-sensitive research data.

[1] The OAIS reference model: <https://public.ccsds.org/pubs/650x0m2.pdf>.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

A. Content distributed as deposited, B. Basic curation – e.g. brief checking; addition of basic metadata or documentation

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

The repository curators performs a basic curation of the submissions mostly by checking and editing the metadata. The repository uses the DSPACE submission workflow, which foresees several steps before completing a submission. The depositor goes through each of the steps guided by the software. The curation workflow should ensure quality and consistency of the data and it offers the possibility to return the metadata and /or the data to the depositor for additional changes before it enters the repository and is visible and harvestable by users and other services.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Outsource Partners. If applicable, please list them.

University of Copenhagen's IT department (UCPH IT [1]) is an organisational outsource partner. The repository is hosted on server infrastructure owned by the University of Copenhagen. UCPH IT manages and supports the server infrastructure, and in addition maintains backups and surveillance [See attachment "Agreement with UCPH IT"].

The repository uses WAYF [2], an AAI infrastructure, run by the Danish e-Infrastructure Cooperation (DeIC), which connects Danish institutions as federated identity providers (IdPs), to enable authenticated access to the repository. Public resources in the repository can be accessed without using WAYF, but authentication is required for restricted resources.

The repository also uses the CLARIN Service Provider Federation (CLARIN SPF [3]) provided by CLARIN ERIC to allow login by users outside Denmark. The CLARIN SPF is the organisational vehicle by which all available CLARIN Service Providers (SPs) are connected in identity federations who have signed agreements with CLARIN SPF. CLARIN SPF provides a centralised Discovery Service that allows researchers from the other participating CLARIN countries to login via their local federated identity provider (IdP) to the repository.

[1] UCPH IT: <https://it.ku.dk/english/>

[2] WAYF: <http://wayf.dk/en>

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

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Other Relevant Information.

See the repository in the Registry of Research Data Repositories: <https://www.re3data.org/repository/r3d100010387>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The mission of CLARIN-DK is to provide easy and sustainable access for scholars in the humanities and social sciences to digital language data (in written, spoken, video or multimodal form) and to provide advanced tools for discovering, exploring, exploiting, annotating, and analysing them. CLARIN-DK also shares knowledge on Danish language technology and resources and is the Danish node in the European CLARIN-ERIC.

The objective of the CLARIN Centre at the University of Copenhagen is to fulfil the CLARIN-DK mission. The mission is also supported by the integration of the repository into the European CLARIN ERIC infrastructure whose ultimate objective is to advance research in the humanities and social sciences by giving researchers unified single sign-on access to

language resources and technology [1].

Digital archiving and easy and sustainable digital access to data resources and tools give the scholars the possibility of developing new research methods and addressing new types of research questions. This will facilitate and enhance participation in collaborative international research.

The CLARIN Centre at the University of Copenhagen supports research data management by providing researchers with a repository (the CLARIN-DK-UCPH Repository) for storage of research data, which are findable, accessible and easy to cite using persistent identifiers. The centre provides data management consultation and support in connection with depositing and reuse of research data.

The CLARIN Centre at the University of Copenhagen is supported and financed by the Faculty of Humanities and the Department of Nordic Studies and Linguistics at the University of Copenhagen until end of 2022. The mission is also stated in [2].

The repository supports the strategy of the University of Copenhagen, Talent and collaboration [3], which applies for 2018-2023. The strategy states that the university will: (3.1) "Improve academic staff's opportunities to exploit the potential of digitisation and big data across the University and with our collaboration partners" and (3.3) "Increase our contribution to open science, including open access, open data and citizen science".

The centre promulgates the mission of CLARIN-DK through publications, conference attendance, organization of courses and workshops, e.g. in collaboration with other Danish Universities and the Danish Royal Library through the national digital humanities collaboration, DIGHUMLAB.

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

[2] CLARIN-DK Mission: <https://info.clarin.dk/en/the-clarin-dk-infrastructure/overview/data-management/>

[3] UCPH strategy: <https://about.ku.dk/strategy2023/collaboration/>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

CLARIN-DK repository distinguishes between two types of licenses:

Distribution license

For every deposit, the repository enters into a standard contract with the submitter, the so-called "Distribution License Agreement" [1], in which the repository describes the rights and duties and the submitter acknowledges that they have the right to submit the data and give the repository the permission to distribute the data on their behalf. The depositor (data producer) is responsible for the data's adherence to relevant legal requirements.

End-users license

The users who download data from the repository are bound by the conditions of the end-user license assigned to the data item. Restricted data requires all users be authenticated and electronically accept the end-user license in order to access the data. A list of available end-user licenses in our repository can be found at [2]. For submitters, there is an additional possibility of requesting custom end-user licenses to be assigned to digital objects during the submission workflow.

The repository does not store research data that is classified as sensitive personal data (in Danish 'personfølsomme data'). Video and sound recordings, for instance, are only allowed to be stored if the depositor has secured permission from the persons involved to distribute the data. In the submission workflow, a repository curator checks that the depositor has the required consents for all video and sound recordings submitted for deposit.

The Terms of Service (TOS [3]) for the repository are governed by Danish laws. These include the University of Copenhagen's general policy on information security [4].

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

[2] End-user licenses: <https://repository.clarin.dk/repository/xmlui/page/licenses>

[3] TOS: <https://repository.clarin.dk/repository/xmlui/page/terms-of-service>

[4] Information security at UCPH: <https://informationssikkerhed.ku.dk/english/is-policy/>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

The repository takes medium-term responsibility for the data, which currently covers the next four years until the end of 2022. In this period we take sole responsibility for the data deposited.

Until the end of 2022 the Faculty of Humanities at University of Copenhagen has agreed to fund the CLARIN ERIC membership and the CLARIN-DK centre. It is currently difficult for the Faculty to have a longer perspective. The documents with the funding statement until 2022 are in Danish and are not available publicly (see a short summary at [1]). As long as funding is available, the CLARIN-DK centre will ensure preservation and access.

In the case of loss of adequate funding of the repository, an agreement on the preservation plan will be provided by either University of Copenhagen or by another member of CLARIN ERIC. CLARIN ERIC is currently widely considered a strong and supportive infrastructure and it offers a number of possibilities for hosting deposited data. If University of Copenhagen fails to prolong the funding from 2023 onwards, the repository will negotiate with the other repositories in CLARIN about

taking over the responsibility for the deposited data. Since PID systems is used in the CLARIN repositories, moving resources from one CLARIN repository centre to another repository using a PID system is possible without affecting the validity of references already included in publications etc., and will therefore not affect the access to the resources.

If moving the resources to a repository with a PID system fails, negotiations about handing over of the data to a generic data repository such as the Danish State Archive will be considered.

The repository staff will continuously advocate for prolongation and seek additional funding of the repository to support and prolong the availability of the repository.

[1] Funding: <https://info.clarin.dk/en/about/>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The depositor (data producer) is responsible for the data's adherence to relevant ethical norms and standards in the discipline in question.

During the submission workflow (ingest) the depositor is required to accept the "Distribution License Agreement", which states that the depositor, as owner or copyright holder, is responsible for the data's compliance with legal and ethical requirements.

The submission workflow requests that depositors provide all metadata necessary to interpret the data during ingest. Metadata for the data is provided aligned with the requirements of the CMDI metadata schema [1].

All users of material from the CLARIN-DK-UCPH Repository must agree to the conditions stated in the end-user license associated with the data, which can include statements relating to the infringement of copyright, intellectual property and redistribution of content.

Furthermore, all public research in Denmark must conform to the Danish Code of Conduct for Research Integrity [2]. In addition, we have a privacy policy for the centre [3], instructions for citations [4] as well as disclaimers for the use of the data in the repository's Term of Service (TOS [5]).

As part of the University of Copenhagen, the CLARIN Centre at the University of Copenhagen is also bound by the privacy policy of University of Copenhagen [6] as well as the Information Security Policy [7] and Cookies and privacy policy for UCPH web sites [8], which protects against the disclosure of sensitive information.

The repository does not accept research data with any disclosure risk. The repository does not need to deal with codes of conduct specifically pertaining to protection of human subjects; since the deposited data are not allowed to include sensitive data that expose the integrity of human subjects without consent are given. Users have the option to ask the centre's help desk for information to clarify if the research data has disclosure risk.

If the requirements are not complied with, then legal consequences may apply, according to national and international laws.

[1] CMDI schema:

https://catalog.clarin.eu/ds/ComponentRegistry?registrySpace=published&itemId=clarin.eu:cr1:p_1403526079380

[2] Danish Code of Conduct for Research Integrity:

<http://ufm.dk/en/publications/2014/the-danish-code-of-conduct-for-research-integrity>

[3] Private Policy: <https://info.clarin.dk/en/the-clarin-dk-infrastructure/overview/privacypolicy/>

[4] Citation: <https://repository.clarin.dk/repository/xmlui/page/cite>

[5] TOC: <https://info.clarin.dk/en/the-clarin-dk-infrastructure/overview/general-terms>

[6] Private Policy at UCPH: <https://informationssikkerhed.ku.dk/english/protection-of-information-privacy/privacy-policy/>

[7] Information Security Policy at UCPH: <https://informationssikkerhed.ku.dk/english/is-policy/>

[8] Cookies and privacy policy for UCPH: <https://about.ku.dk/cookies-and-privacy-policy/>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

V. Organizational infrastructure

R5. 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:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The CLARIN Centre at the University of Copenhagen is hosted at the Faculty of Humanities, which falls under the governance of the University of Copenhagen, the largest university in Denmark.

The CLARIN Centre at the University of Copenhagen was financed by the Ministry of Higher Education and Science through the national infrastructure collaboration DIGHUMLAB [1] until 2017. For the period 2018-2022 (incl.), the Faculties of Humanities and the Department of Nordic Studies and Linguistics at the University of Copenhagen finance the centre with a budget of 90.000 Euro, providing for approx. 1.1 full-time equivalent (FTE).

The staff have different roles:

- National Coordinator (0.1 FTE)
- Repository Manager (0.2 FTE)
- Repository Administrators and Developers (0.3 FTE)

- Repository Curators (0.1 FTE)
- Guidance and end-user services (0.4 FTE)

The University of Copenhagen finances the servers and the network administration of the centre [See attachment “Agreement with UCPH IT”], and the faculty pays the yearly National CLARIN membership fee.

The staff at the CLARIN Centre at the University of Copenhagen are affiliated with the Department of Nordic Studies and Linguistics [2], and are listed at [3]. All staff involved with the centre hold permanent positions and have all been working for the centre since 2011. The staff is experienced in the area of infrastructure, web services, digital tools, collection and annotation of resources, and are active participants in both national and international forums that aim to establish best practices for infrastructures for digital archiving. The staff members regularly participate in training and professional development activities organised and supported by CLARIN ERIC.

The staff are members of CLARIN ERIC bodies (Standing Committee for CLARIN Technical Centres, Assessment Committee for CLARIN Technical Centres, CLARIN Taskforce for Authentication and Identification, and CLARIN Legal Issues Committee). A staff member represents the University of Copenhagen in the Danish National Forum for Data Management.

[1] DIGHUMLAB: <http://dighumlab.dk>

[2] NorS: <https://nors.ku.dk/english>

[3] Staff: <https://info.clarin.dk/en/about/contact>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

In-house guidance

Besides the expertise of repository staff, the repository has access to researchers at the department who are experts in different data formats, e.g. TEI, and formats for data of the type video, audio and software. Furthermore, the repository manager is a member of the national data management group, securing access to technical experts from other research fields.

International advisers

More importantly, by being part of the CLARIN ERIC federation of technical centres, CLARIN-DK is in constant contact with experts in all the CLARIN ERIC member countries, in particular with those working at B and K centres [1]. Especially Standing Committee for CLARIN Technical Centres (SCCTC) [2] and the CLARIN task forces [3] are facilitating technical knowledge sharing. Finally, CLARIN ERIC has an international Scientific Advisory Board [4], which also gives feedback, communicated to the national centres when relevant.

Designated community

The repository gets feedback from its designated community mainly through the help desk. The help desk receives requests via email or through face-to-face meetings. Such requests may result in improved guidance, or that new data types or formats should be taken into account.

The CLARIN Knowledge centre, DANSK, (DANish helpdeSK) [5] is a cooperation between two well-established institutions, the Department of Nordic Studies and Linguistics and the Danish Language Council. Together these institutions comprise a great part of the expertise on Danish language, language resources and language technology.

Users

The repository experts also meet the users at workshops and seminars organised throughout the year (normally 3-4 per year). Furthermore, the DIGHUMLAB Strategic Board provides input from community users. Finally, being a part of CLARIN ERIC enables the repository to receive input from other members countries, not least through the CLARIN Knowledge Sharing Infrastructure (CLARIN KSI) and the CLARIN Annual Conference.

[1] CLARIN centres: <https://www.clarin.eu/content/clarin-centres>

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

[3] CLARIN Task Force: <https://www.clarin.eu/node/3573>

[4] CLARIN Scientific Advisory Board: <https://www.clarin.eu/governance/scientific-advisory-board>

[5] CLARIN Knowledge centre, DANSK: <https://info.clarin.dk/en/k-center/>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Integrity

The repository uses MD5 checksums to verify the data integrity of all digital objects. The checksum is attached as metadata to the created bitstream (digital object), together with other file information (i.e. size, format, filename), during the submission. A checksum checker [1] is run periodically to verify that all digital objects validate against their recorded checksum. Integrity checks are run when submitted bitstreams are in the format of supported archive types i.e. zip, tar, gzip.

The submission workflow ensures that the Submission Information Package (SIP) is validated in compliance with established policies. All submissions follow the lifecycle policy [2].

Our policy concerning metadata changes is to allow for changes in the case of misspellings, minimal corrections or adding further information. Any changes to metadata are recorded in provenance information. A timestamp is recorded in the following internal metadata field, `dc.date.accessioned`, to hold the last updated timestamp. Log files are stored including appropriate messages regarding changes made to data and/or metadata.

As stated in the Distribution License Agreement, data is not permitted to be modified in the Archive Information Package (AIP). However, for major modifications to the AIP the depositor is requested to submit a new version of that AIP and it will be stated in the metadata that it has been replaced by a newer version. In case of withdrawal requests, these will be evaluated on case-by-case basis. Furthermore, we reserve the right to keep the metadata of published submissions available if there is no legal requirement to delete the metadata.

New versions of AIPs can be submitted via the submission workflow. Once the new version of the AIP is accepted by a repository curator, it will obtain a new persistent identifier (PID), and a dependency relation reference ('replaces') is added to the metadata record pointing to the previous version. The previous version will still be maintained and not withdrawn from the repository. It will keep its original PID and a dependency relation reference ('isReplacedBy') will be added to the metadata record, with the PID of the new version. If an AIP has multiple versions, a notice is displayed to the user when accessing the landing page. On the landing page a list of the available versions of the AIP is available for the user to load an alternate version, i.e. a superseding version.

The repository employs a number of standards:

OAI-PMH (v2) protocol standard for metadata harvesting; the repository makes metadata formats available via the repository's OAI-PMH endpoint.

The OLAC and Dublin Core DCMI are common metadata standards used in digital archiving. DCMI is used as the basis for the internal metadata registry.

Metadata are also made available according to the CMDI 1.2 [3] metadata specification, which can be harvested via the repository's OAI-PMH endpoint. CLARIN ERIC harvests our CMDI metadata to a central registry, which can be seen at CLARIN VLO [4]. Results can be filtered by choosing Collection: CLARIN-DK-UCPH Repository [5].

CMDI Profile schemas are based on the W3C XML Schema standard.

CMDI metadata schemas refer to the CLARIN Concept Registry [6] (previously ISOcat), a concept scheme model based on SKOS (Simple Knowledge Organization System).

METS (Metadata Encoding and Transmission Standard) is used as the DSpace AIP Backup and Restore format [7] for the backup of the repository.

REST API - the repository makes available the DSpace REST API [8], which provides a programmatic interface to the repository.

The repository database and HTML web pages use the UTF-8 encoding standard, for correct character encoding.

The repository uses ISO language codes (ISO-639-2) in metadata.

The repository allows for a variety of data formats for the submitted digital objects, and a weekly report is sent to administrators to keep track of all used data formats in the repository.

Authenticity

The repository stores bitstreams (digital objects) but does not process or alter them in any way. If the SIP has not yet been accepted and published, the depositor can ask for an update or modification of the bitstreams.

All AIPs in the repository contain metadata. The repository curator can update metadata, if either the depositor or repository curator sees a need for it [9].

All modifications to AIPs are recorded in provenance information as statements in the metadata record of the internal metadata field, dc.description.provenance, of the AIP. The statement contains a description of the change or action taken on the AIP by an authorised user, and includes the checksums and file information of all bitstreams (digital objects).

Depositors can review the repository item lifecycle for an overview of the strategy [10].

The repository does not compare versions of the same file. Each version is regarded as an independent resource in its own right.

The repository relies on a federated identity service via national federations i.e. WAYF-DK for authorisation. Depositors are required to have a valid user ID at their local federated identity provider (IdP). Once authenticated by the IdP, if a local repository user account does not already exist, a new user account is created in the repository containing some provided user attributes i.e. name and email address, which may be used in any provenance information in the metadata record of a submission. All submissions are registered under a user account in the repository.

[1] Checksum checker: <https://wiki.duraspace.org/display/DSDOC5x/Validating+Checksums+of+Bitstreams>

[2] Lifecycle policy: <https://repository.clarin.dk/repository/xmlui/page/item-lifecycle>

[3] CMDI 1.2 metadata specification: <https://www.clarin.eu/cmd1.2>

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

[5] Filtering in VLO: <https://vlo.clarin.eu/?0&fq=collection:CLARIN-DK-UCPH+Repository&fqType=collection:or>

[6] CLARIN Concept Registry: <https://concepts.clarin.eu/ccr/>

[7] AIP formats: <https://wiki.duraspace.org/display/DSDOC5x/DSpace+AIP+Format>

[8] DSpace REST API: <https://wiki.duraspace.org/display/DSDOC5x/REST+API>

[9] Metadata update: <https://repository.clarin.dk/repository/xmlui/page/faq#how-to-update>

[10] Lifecycle: <https://repository.clarin.dk/repository/xmlui/page/item-lifecycle>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

The repository offers to host language-based, research data from humanities and social sciences. Four types of data can be submitted; corpora, language descriptions, lexical-conceptual resources and tools.

Data has to be shareable according to the Distribution License Agreement and available end-user licenses. Data is accepted for archiving when the data is evaluated as relevant for research and size of the data is below the limit of free amount of archiving in the repository. If the size of the data exceeds the repository limit for free storage of data (data packages of a size exceeding 100 GB will be assessed) some will be offered storage for a yearly negotiated fee, others will be offered free storage or rejected. If possible, a rejection will include an advice about other options for deposit [1].

Data has to follow the requirements of the repository [1] [2]. The depositor is expected to deliver data living up to the academic standards [3].

Deposits of data is only permitted if the required metadata is provided. Data has to be accompanied by relevant metadata to enable users to assess the provenance, suitability and quality of the data.

The metadata policy page contains details about the required metadata requested during the submission workflow [4].

The submission workflow requests that depositor provide all metadata necessary to interpret the data prior to acceptance and publication in the repository.

The metadata provided with data shall be valid with respect to the CMDI metadata schema:

LINDAT_CLARIN [5] (Full XML Schema is downloadable [6])

If the metadata provided is insufficient according to the repository, the repository curator will send back the submission to the depositor for required updates.

Data is stored in its original file formats in the repository. The repository currently performs a check on all uploads using the default list of preferred formats published at [7]. The repository will publish recommendations for preferred formats during 2019 prioritizing open formats. Unsupported formats will be flagged as 'Unknown' in the metadata. The repository curator performs a curation task to obtain a summary of all formats of the uploaded bitstreams (digital objects) to check if they correspond to the preferred formats.

If the deposit includes bitstreams in non-preferred formats, the repository curator may still allow them, if a suitable description is provided, if conversion to preferred formats cannot easily be carried out, and if the bitstreams pass further verification checks against the file content type. The deposited bitstream will then show a value of 'Unknown' as the format described in the bitstream metadata.

[1] Accepted submissions: <https://repository.clarin.dk/repository/xmlui/page/faq#what-submissions-do-we-accept>

[2] Quality control: <https://repository.clarin.dk/repository/xmlui/page/faq#quality-control-checks>

[3] Code of Conduct: <http://ufm.dk/en/publications/2014/the-danish-code-of-conduct-for-research-integrity> for researchers

[4] About metadata <https://repository.clarin.dk/repository/xmlui/page/metadata>

[5] CMDI schema in the Component Registry:

https://catalog.clarin.eu/ds/ComponentRegistry#/?itemId=clarin.eu%3Acr1%3Ap_1403526079380®istrySpace=public

[6] The CMDI schema:

https://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/1.x/profiles/clarin.eu:cr1:p_1403526079380/xsd

[7] Default list of preferred formats: <https://wiki.duraspace.org/display/DSDOC5x/Metadata+and+Bitstream+Format+Registries#MetadataandBitstreamFormatRegistries-DefaultBitstreamFormatRegistry>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

The repository uses a customisation of DSpace, CLARIN-DSpace [1], as the base repository software. DSpace includes a defined submission workflow, which is supported by the repository's interface and follows OAIS guidelines concerning the ingestion process, and repository curators can process the submissions.

Data ingestion can be made with Submission Information Packages (SIP) packages. However, the default method for data submission is via the submission workflow interface. Archive Information Package (AIPs) of the deposited resources are available once the submission is approved by the repository curator.

An overview of the DSpace Storage Layer is documented [2]. In addition, an internal wiki includes documentation regarding the installation, configuration and application management of the repository including storage management.

A limited group of staff members with server permissions have access to the repository server and data partitions. Surveillance services have been setup and enabled on the repository server by the University of Copenhagen's IT department. See R15 and R16.

An initial strategy for backup storage is implemented. Data storage is preserved by backing up the DSpace AssetStore together with copies of the DSpace database, which are copied separately to a NFS backup partition. A full site backup is created by generating archival copies in Archive Information Packages (AIPs) using the DSpace AIP Backup and Restore function. Backups are run on a regular basis, which create a backup-restore point of the repository. See R16 for procedures regarding backup and recovery.

Regarding the long-term preservation of repository data, backup copies of the repository are saved to an external NFS partition managed by the University of Copenhagen IT department. In addition, multiple copies are created in nightly snapshots of the repository and external NFS partitions as mentioned in R15.

The repository's relevant file system, including project source code, installation files, and database are backed up and managed using the backup2l [3] utility, as suggested by CLARIN-DSpace software [4].

The University of Copenhagen IT department can restore the repository server and NFS storage partition to a previous snapshot on request, in the case failure or corruption of the repository. An agreement exists regarding the services the University of Copenhagen IT department delivers [See attachment "Agreement with UCPH IT"]. Backup-restore points based on archival backup copies will be available to restore repository data to a previous state.

Standard practice of MD5 checksums are used to check for consistency across the repository's AIP packages saved to the backup partition. In addition, rsync logs for copies of the bitstreams help to check for consistency of the DSpace AssetStore backup.

The University of Copenhagen IT department is responsible for storage media handling and monitoring of server infrastructure and data partitions.

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

[2] DSpace Storage Layer: <https://wiki.duraspace.org/display/DSDOC5x/Storage+Layer>

[3] Backup21 utility: <http://backup2l.sourceforge.net/>

[4] CLARIN DSpace: <https://github.com/ufal/clarin-dspace/wiki/Backup>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

Funding: The CLARIN-DK centre will ensure preservation and access to the data in this repository as long as sufficient funding is available. Until the end of 2022 Faculty of Humanities at University of Copenhagen has agreed to fund the CLARIN-DK centre and efforts will be made to extend the funding period. If a situation of insufficient funding for continuity of the repository occurs, an agreement on preservation of and access to data will be made with either University of Copenhagen or with another member of CLARIN ERIC. Otherwise, the data will be preserved by turning it over to the National Archive of Denmark [1], which takes care of preservation of Danish research data. Currently, the size of research data stored in the repository is less than 100GB and transfer of data to another repository is also in respect to size a do-able task.

The legal aspects of the process of relocating data to another institution is addressed by the deposition license, Deposition License Agreement [2]. This agreement between the depositor and the repository gives all permissions required to meet responsibilities for preservation of the data. The transfer of custody of the data provided and permission to share the data on the depositor's behalf is covered by the deposition license too. The repository has the rights to copy, transform, and store the items, as well as to provide access to them as long as it complies with the deposition license.

Technology: Our repository uses a repository system widely used at CLARIN data centres (the CLARIN DSpace repository system with well-documented modifications). It is currently running on virtual machines at the IT department of University of Copenhagen (UCPH-IT), which has the capacity to keep them up and running, also while potential negotiations to take over the hosting of either the repository or all the data in the repository are completed. In case where the data are moved to another DSpace repository within CLARIN or another repository using a PID system, the PID's will be kept unchanged and the transfer will not affect the validity of PID references already included in publications etc., and will therefore have minimal impact on user experience. Updates and upgrades of the repository system software and server software will be carried out continuously. In general, developments in technology are followed and changes relevant to the repository system are continuously discussed in the Standing Committee for CLARIN Technical Centres, SCCTC [3].

Data storage and backup: UCPH-IT is responsible for storage media handling and monitoring of server infrastructure and data partitions. Backup-restore points based on archival backup copies will be available to restore repository data to a

previous state. Furthermore, the Archive Information Packages (AIPs) are stored on an NFS storage partition. Standard practice of checksums is used to check for consistency across the repository's AIP packages saved to the backup partition.

File formats: During the submission workflow of data, the use of open and well-documented data formats are encouraged allowing for easy conversion into other formats. Data in proprietary formats will only be accepted when conversion cannot easily be done. Weekly, a report of currently used data formats is sent to the repository responsible, who in this way can monitor the use of data formats.

Information security: The repository system logs any changes to data resources, and any changes in metadata. Furthermore, the changes can only be performed by a small group of people, that are clearly defined and belong to the CLARIN-DK group. Automatic checks for data formats, links, and activity are run on a weekly basis and inspected by the repository responsible. Furthermore, the servers are under surveillance including monitoring the configured network services and disk usage, and intrusion prevention software is in use on the repository server to prevent security incidents from occurring. The servers are scanned at least four times per year and evaluated for security issues by the Danish Computer Security Incident Response Team (DKCERT) [4]. The repository administrators respond to any reported security-related issues or incidents as quickly as possible. These initiatives focus on keeping the data and metadata safe and secure.

Suggestions from users about changes in functionality are appreciated and encouraged, e.g. users can send mails to info@clarin.dk, and the staff in the helpdesk and knowledge sharing service are ready to enter into dialogue with users about their wishes for new or improved functionality, better help pages, metadata guidance or other changes in user requirements.

As mentioned above, the policy for securing preservation and access is well defined, and this is documented in the current preservation policy [5], which will be subject to change yearly based on future experience with depositors and technology development.

[1] The Danish National Archives <https://www.sa.dk/en/>

[2] Deposition License Agreement: <https://repository.clarin.dk/repository/xmlui/page/contract>

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

[4] DKCERT: <https://www.cert.dk/en>

[5] Preservation policy: <https://repository.clarin.dk/repository/xmlui/page/about#preservation-policy>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XI. Data quality

R11. 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:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The repository focuses on sharing research data, and holds digital objects from different humanistic research communities. The depositors are authenticated through WAYF [1] or through other national federations via CLARIN Service Provider Federation (CLARIN SPF [2]).

In the deposit phase of the submission workflow, the depositor (data producer) is guided through a set of metadata of which several fields (e.g. Type, Title and Description) are mandatory, and used as search facets for browsing the repository (e.g. Type). The entered metadata are automatically validated and mapped to a CMDI metadata schema [3] and furthermore linked to the pan-European CLARIN VLO search interface [4]. To enable sharing in the international and European research community, English is the recommended language for metadata.

The metadata, the data format and the data soundness are checked and approved by a repository curator before the resources are published in the repository's collection. In case submitted metadata is inadequate, or documentation is missing, the depositor must rewrite the metadata or submit the missing information using the submission workflow. The metadata allows for referring to external publications, and the persistent identifiers (PIDs) together with other information such as authors, title and publisher allows for citing of the resource itself e.g. BibTex format generated automatically by the repository:

@misc{20.500.12115/8,

title = {The Danish Parliament Corpus 2009 - 2017, v1},

author = {Hansen, Dorte Haltrup},


```
url = {http://hdl.handle.net/20.500.12115/8},
note = {{CLARIN}-{DK}-{UCPH} Centre Repository},
copyright = {Creative Commons - Attribution 4.0 International ({CC} {BY} 4.0)},
year = {2018} }
```

In the submission workflow, each created bitstream (digital object) includes technical metadata with a name, a description, file size and format to give the end-user better overview of the digital objects. The repository gives the curator access to a set of curation tools, which assist in checking the content and metadata quality in submissions. These are:

- Summarise bitstream formats
- Check for required metadata
- Check that metadata URLs are working
- Check that Items handle is working
- Update description URI
- Check metadata for quality and consistency
- Gather additional information from bitstream
- Check deposit license

The repository expects the depositors to provide data living up to the researchers' academic standards and Code of Conduct [5] implying that data is of a quality usable for research. The curator checks the data quality by opening samples of the uploaded data files.

The repository help desk is available with help and information about the metadata descriptions and use of data in the repository to both depositors and end-users. The repository does not label data with quality scores, but end-users are welcome to comment on the quality of data by contacting the help desk or directly to the repository staff directly.

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

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

[3] CMDI schema:

https://catalog.clarin.eu/ds/ComponentRegistry?registrySpace=published&itemId=clarin.eu:cr1:p_1403526079380

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

[5] Code of Conduct: <http://ufm.dk/en/publications/2014/the-danish-code-of-conduct-for-research-integrity>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

A submission workflow is offered and internally configured in the CLARIN-DK-UCPH Repository, providing the depositor (data producer) with a platform for the submission of metadata and upload of the digital objects [1].

After the depositor creates a submission, the submission workflow allows the repository curators to verify and validate the quality and consistency of the submission, with the possibility of returning the submission to the depositor for changes [2].

Finally, ensuing the acceptance of the submission by a repository curator, the submission receives a new persistent identifier (PID) and becomes visible and retrievable via the repository web interface and harvestable to the CLARIN Virtual Language Observatory (CLARIN VLO [3]) at the next scheduled OAI-PMH harvesting.

The CLARIN-DK Centre policy is to have two persons associated with each role of the repository: a primary and a secondary curator, a primary and a secondary repository developer, and a primary and a secondary repository manager. In this way, the continuity of the repository in case of staff changes is secured.

Curation rules on how to proceed and deal with deposit submissions, are defined: firstly, the curator checks the content and validity of the metadata and then the validity of the content data by checking whether the data formats are acceptable, whether the files are properly described and whether there is a proper license. The repository curator will review if there are traces of sensitive data by reading the metadata provided with submitted digital objects. In regards to submitted audio and video data, the curator will further check whether consent has been obtained i.e. via consent forms. The repository does not store research data that is classified as sensitive, including personal data, without consent forms. The repository only accepts language-based data falling under the following four data types: corpus, languageDescription, lexicalConceptualResource or toolService.

Concerning the acceptance of very large data collections or data outside the scope of the supported data types, a negotiation will occur between the repository manager and the depositor [4].

The workflow for archiving data objects to the CLARIN-DK-UCPH Repository is the following [5]:

1. Depositors submit their data to repository using the defined submission workflow [1].
2. The curator receives a notification via e-mail and performs an assessment of the submission; if necessary, the curator will contact the depositor and request re-submission. The curator works closely with the depositor in the case of insufficient data or metadata quality, and/or licensing issues.
3. The submission information package (SIP) is approved, published and receives a new persistent identifier (PID).
4. The digital objects are retrievable via the repository interface [6] and metadata is available for harvesting via the CLARIN Virtual Language Observatory [3] through OAI-PMH harvesting.

The workflow is valid for all the four data types that the repository accepts.

[1] Submission platform: <https://repository.clarin.dk/repository/xmlui/page/deposit>

[2] Lifecycle: <https://repository.clarin.dk/repository/xmlui/page/item-lifecycle>

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

[4] Accepted submissions: <https://repository.clarin.dk/repository/xmlui/page/faq#what-submissions-do-we-accept>

[5] Submission and withdrawal workflows: <https://repository.clarin.dk/repository/xmlui/page/workflows>

[6] Repository interface: <https://repository.clarin.dk/repository/xmlui/>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The repository can be searched using the DSpace provided search interface, which is powered by an Apache Solr-based index. The repository can also be browsed or filtered by facets i.e. subject or date issued. The search interface is publicly accessible at: [1]

To enhance discoverability of content, the repository supports metadata harvesting via OAI-PMH protocol using DSpace's OAI v2 server. The CLARIN Virtual Language Observatory (CLARIN VLO) harvests data from our OAI endpoint into the CLARIN VLO, where language resources may be discovered using a facet browser [2]. The metadata mapping from the repository to the VLO search facets uses the CLARIN defined CMDI metadata framework [3].

Data citations are implemented according to the recommendations of the Research Data Alliance's Data Citation Working Group [4]. End-users are asked to acknowledge and cite data sources properly in all publications and outputs [5]. See a citation example in Section 11 Data quality.

Each submission in the repository is given a unique persistent identifier (PID) when published (approved by a repository curator) in the collection. The repository issues PIDs via a local Handle.net server running on the repository's server with a global PID prefix, 20.500.12115, registered by the Handle.Net Registry (HNR) [6] and is renewed periodically every 5 years.

[1] Browsing CLARIN-DK-UCPH Repository: <https://repository.clarin.dk/repository/xmlui/discover?advance>

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

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

[4] Research Data Alliance's Data Citation Working Group: <https://rd-alliance.org/groups/data-citation-wg.html>

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

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

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XIV. Data reuse

R14. 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:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The CLARIN-DK-UCPH Repository requires metadata to be provided (mandatory and recommended metadata) describing the submitted data and the additional details regarding authorship of the data. The repository currently allows for four standard types of data: corpus, languageDescription, lexicalConceptualResource or toolService. Required metadata (Title, Date, Creator, Publisher, Contact person, Description, Language, Subject keywords, Media Type) complies with Dublin Core (DCMI) and together with additional custom metadata fields, combine to fulfil the requirements of the CMDI metadata schema [1]. A submission cannot be accepted until required metadata fields are completed [2].

The repository accepts formats that ensure that all digital objects are clearly understandable and usable by research communities in the Designated Community. Therefore, the repository recommend and support common formats used in research communities that are suitable for long-term preservation [3].

The repository uses DSpace's default bitstream (digital object) registry [4] for supported data formats. In the case of newer or commonly requested formats, we can update the registry to implement support for these, when approved by the repository manager.

The repository ensures that all resources have descriptive metadata, and a help desk is available to provide timely assistance to repository users, if they have problems understanding metadata requirements and/or licensing. Repository curators will periodically review the data in the repository, and determine if any changes to the metadata should apply during a process of normalisation or correcting errors. The repository supports re-submission of data by the depositor

(data producer) as a new version.

[1] CMDI schema: https://catalog.clarin.eu/ds/ComponentRegistry/#/?registrySpace=published&itemId=clarin.eu:cr1:p_1403526079380&_k=z11up6

[2] Metadata: <https://repository.clarin.dk/repository/xmlui/page/metadata>

[3] Recommended formats: <https://repository.clarin.dk/repository/xmlui/page/formats>

[4] DSpace's bitstream formats: <https://wiki.duraspace.org/display/DSDOC5x/Metadata+and+Bitstream+Format+Registries#MetadataandBitstreamFormatRegistries-DefaultBitstreamFormatRegistry>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

TECHNOLOGY

XV. Technical infrastructure

R15. 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:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The CLARIN-DK-UCPH Repository follows the standards that are relevant to CLARIN ERIC [1] and the research community together with the relevant international standards i.e. W3C and HTML for web applications, Creative Commons (CC) for licenses, Handle.net for persistent identifiers, Unicode for encoding text, XML and other open standards for data formats and data transformation, CMDI, Dublin Core and OLAC for metadata, and the ISO OAIS Reference Model for archives. The University of Copenhagen as the hosting organisation of the repository adheres to the information security standard, DS/ISO/IEC 27001. The repository will as part of the implementation plan, review yearly the most current international and community standards.

The core repository software is a customised fork of DSpace (version 5.x), CLARIN-DSpace [2]. The CLARIN-DSpace project is maintained by the Centre for Language Research Infrastructure in the Czech Republic [3] and supported by CLARIN community. Additional customisations to this project are stored in an independent Git repository [4] managed by the repository staff. Other customisations to the software inventory including configurations and other files used in the project build are stored in another Git repository, hosted on an internal Gitlab server [5] at the University of Copenhagen's IT department (UCPH IT). All repository digital objects are stored on external NFS partitions maintained by UCPH IT.

Installation, setup and maintenance documentation for the repository is available to the repository staff members in an internal wiki.

Community-supported software used on the repository server include Shibboleth Service Provider, PostgreSQL, Nginx, OpenSSL, Git, and Apache support-projects including Apache Solr, Apache Maven, Apache Ant. The repository server runs on community-supported, open-source, Debian-based Linux distribution, Ubuntu (LTS version), which is one of the operating systems supported by UCPH IT.

Regarding the network performance, the connectivity is sufficient as the repository is hosted on a Virtual Machine infrastructure managed by the University of Copenhagen's IT department. The repository may experience lower performance due to backup processes at early hours (CET time) due to CPU or memory usage.

Data management: Licensing standards are applied on all deposits i.e. the depositor has to choose the appropriate license from the list of available licenses supported by the repository [6]. To assist the depositor's license selection, a public license selector widget [7] is available in the interface of the submission workflow.

Preservation Planning: It is important that the research data in the repository is preserved long term. The repository backs up all Archival Information Packages (AIPs) to an external NFS partition, as part of the preservation plan. The metadata are also made available in other supported formats via the OAI-PMH endpoint to be compliant to other metadata standards, i.e. OpenAIRE [8], using crosswalks with the configured DSpace OAI server [9]. The DSpace OAI server uses the XOAI Java toolkit [10] to support the metadata harvesting according to community metadata standards.

Administration: Contract agreements in the form of a standard Distribution License Agreement [11] with the depositor (data producer), are agreed to when completing the license step in the submission workflow.

Access: The bitstreams (digital objects) or whole Archival Information Units (AIU) are delivered to consumers, who are authorised

to access the data. Metadata are available under the public CC0 license.

The persistent identifier (PID) is the unique identifier used for making the research data citable. PIDs are created using a local Handle.net server [12]. Metadata records are harvestable and the PIDs are used by metadata registries, e.g. CLARIN VLO [13].

As a Service Provider (SP), the repository is a member of the CLARIN Service Provider Federation (CLARIN SPF) [14], which connects SPs with national identity federations i.e. WAYF [15], to enable Single Sign-On (SSO) login via a federated identity based on SAML 2.0 [16].

The repository plans for infrastructure development in accordance with the development in CLARIN ERIC. Each year, the National Coordinator for CLARIN-DK together with the repository manager reviews the current resource types and the file formats available. On this basis, they decide if updates are needed for data types, recommended file formats or guidance to users. These decisions form the basis for the implementation plan for the repository for the current year. As part of the implementation plan, technical changes or updates to the repository are also considered to ensure that the implementation is up-to-date and major updates are scheduled. The repository manager continuously follows up on any security changes that are requested by UCPH IT or specified in the reports from the Danish Computer Security Incident Response Team (DKCERT) [17]. The repository backend has been updated in 2018; therefore, there are no near-future plans for other major backend infrastructure developments. There are plans to include additional web services, for example, the Korp service [18] that currently is accessible as a standalone service.

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

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

[3] LINDAT/CLARIN: <https://lindat.mff.cuni.cz/en>

[4] CLARIN-DK DSpace in GitHub: <https://github.com/kuhumcst/clarin-dspace>

[5] Internal Gitlab server: <https://git.sc.ku.dk/kuhumcst/clarindk-dspace-overlays>

[6] Available licenses: <https://repository.clarin.dk/repository/xmlui/page/licenses>

[7] Public license selector widget: <https://github.com/ufal/public-license-selector>

[8] OpenAIRE: https://guidelines.openaire.eu/en/latest/data/application_profile.html

[9] DSpace OAI server: <https://wiki.duraspace.org/display/DSDOC5x/OAI+2.0+Server>

[10] XOAI Java toolkit: <https://github.com/DSpace/xoai>

[11] Distribution License Agreement: <https://repository.clarin.dk/repository/xmlui/page/contract>

[12] Handle.net server: <http://handle.net>

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

[14] CLARIN Service Provider Federation: <https://www.clarin.eu/content/service-provider-federation>

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

[16] SAML: <http://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-tech-overview-2.0-cd-02.html>

[17] DKCERT: <https://cert.dk/en>

[18] Korp service: <https://alf.hum.ku.dk/korp/>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Backup and recovery services are provided by the University of Copenhagen's IT department (UCPH IT). Repository data is protected and secured by storing all repository data files on an external NFS partition and a backup copy on a separate backup NFS partition, which is hosted on a server with a different physical location from the repository's server.

Our backup strategy allows us to recover from the loss of repository data, or the loss of an entire server. Snapshots of the servers and NFS partitions are created daily. Only changes to the data in the infrastructure that occurred after the chosen backup-restore point will be lost. Furthermore, an entire repository and collection archive export is stored and regularly updated on the backup NFS partition as Archival Information Packages (AIP), which can facilitate restoring the entire repository or individual items, in the case of corruption or loss of repository data.

Software configuration and customisation files are stored in GIT repositories hosted internally by the UCPH IT and these can be used to easily restore the repository build, to provide swift recovery, in the case of the loss of the entire server. The servers are under surveillance using Zabbix [1], an open-source monitoring tool, from UCPH IT, including monitoring

the configured network services and disk usage. The repository server uses a valid SSL certificate.

The servers are scanned at least four times per year and evaluated for security issues by the Danish Computer Security Incident Response Team (DKCERT) [2].

The repository administrators respond to any reported security-related issues or incidents as quickly as possible. The repository staff receives weekly a sanity check report including any error reports via e-mail. Intrusion prevention software, fail2ban [3] is installed and configured on the repository server to prevent security incidents from occurring.

[1] Zabbix: <https://www.zabbix.com>

[2] DKCERT: <https://www.cert.dk/en>

[3] fail2ban: <https://www.fail2ban.org>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: