

CLARIN Repository Requirements

Context

CLARIN is a European Research Infrastructure for Language Resources and Technology. It is based on a distributed infrastructure with currently 66 centres that provide resources and expertise in their scientific domain. Becoming an official CLARIN centre requires a certification process every three years in which (among others topics) technical and organisational competence and the availability of a stable technical infrastructure has to be proven.

This certification is based on the following two parts, which result in the specific requirements for a CLARIN-compatible repository implementation that are stated below.

- To receive the [Core Trust Seal](#), 16 requirements focusing on the stable and **long-term availability** of a data repository must be fulfilled.
- The additional [CLARIN centre requirements](#) focus on project-specific issues, including availability of **key interfaces** like OAI-PMH.

Main Requirements

According to the aforementioned factors the following requirements are considered essential:

- Long-term support version (LTS) of the software
 - In case of a software version reaching EOL: support of simple migration to a supported (LTS) version.
- Generic OAI-PMH interface, supporting:
 - Publication of non-DCMI (oai_dc) metadata, as CLARIN centres and the [Virtual Language Observatory](#) rely on other formats (mostly [CMDI](#) based).
 - OAI-PMH sets to support selective harvesting. Selective harvesting is used in CLARIN by [WebLicht](#) for harvesting relevant web service records, and by centres that also provide resources beyond the scope of CLARIN to present only specific information to the [CLARIN OAI-PMH harvester](#) or Virtual Language Observatory, often via declaring dedicated OAI-PMH sets.
 - Option to exclude data in the repository from OAI-PMH, e.g. for experimental, confidential or obsolete data.

Optional Desirables

- Access to / modification of repository via stable API or client software in a popular programming language (e.g. Java, Python).
- Open development process.

Contact in the CLARIN project

CLARIN Fedora interest group under the auspices of the CLARIN Standing Committee for CLARIN Technical Centres ([SCCTC](#)), Coordination via mailing list: lg-fedora@lists.clarin.eu