

Preservation Policy

Purpose of this document

This document specifies the preservation policy at the CLARIN-D Center Leipzig. In case of questions please contact: clarin@informatik.uni-leipzig.de

Data Storage and Backup

Data is stored on a RAID system and all contents are regularly copied to separate hardware. The deterioration of storage media is monitored via Nagios probes (e.g. S.M.A.R.T. - Self-Monitoring, Analysis and Reporting Technology – data). Access to the archive system is limited to a small group of people. While read-access is possible from external systems, write access to data storage systems running on the archive machine is limited to the machine itself (e.g. no write access to MySQL database from external sources).

Data Preservation

Depository are encouraged to use standardized formats (UTF-8, documented XML formats, ...) when submitting their data for deposition. In case custom / proprietary formats are used, an exhaustive documentation has to be provided. The CLARIN-D center Leipzig will check once a year:

- whether the data that was stored is unchanged (e.g. checksums)
- whether an update of the deposited data due to the obsolescence of the used format is necessary
- whether an update of the available metadata is necessary

In case an update is necessary, the depositor is contacted and asked to provide an updated version. In some cases the center may also decide to do the update on its own. In this case the original depositor will be informed (in case the person/institution is still available).

Backups are held on hardware that is situated on a location that is separated from the live system. In the future the CLARIN-D center Leipzig aims at establishing co-operations with external data centers (e.g. other CLARIN-D centers) in order to distribute (encrypted) backups to locations that are geographically separated from the hosting site in Leipzig.

Software Stack Preservation

The center utilizes widely used open source software stacks (MySQL, Tomcat, Fedora Repository) that are installed on virtual machines in order to facilitate all repository services (data storage, OAI-PMH, ...). This maximizes the probability of long term support (updates, security fixes) for the tools being used and improves the ability to run installations of these software stacks independent from the underlying hardware and/or operating system. The CLARIN-D center Leipzig will check at least once a year:

- whether major updates of software components are available and necessary
- whether security fixes are available
- whether software components are still actively updated / developed or abandoned and switching to alternatives should be considered

- whether providing access to the data/metadata stored in the repository should be made available via additional interfaces (or updated versions of existing ones)