

DKRZ¹ Long Term Archive²: Preservation and Storage Policy

Revision History

Revision	Author	Scope
August-2015	DKRZ Data management	public release
October-2018	DKRZ Data management	public release

Policy Statement

Data preservation is part of the operations and planning of DKRZ-LTA to support climate research.

Intended Audience

DKRZ LTA Managing Director, DKRZ Data Manager, DKRZ-LTA Users.

Motivation for This Policy

The main tasks of the DKRZ Long Term Archive is to choose, preserve and make available for use data and metadata from numerical climate models and related observations. Focus is on data products and not on raw output from models. Data objects are preserved in discipline acknowledged formats.

This policy is driven by customer acknowledged guidelines like “Rules of good scientific practice” from the Max Planck Society or the DFG³, by community driven standards (like OAIS reference model), Data Seal of Approval (DSA) and other accepted standards like ISO16363 (audit and certification of trustworthy digital repositories).

Guidelines

Storage duration

Data in DKRZ-LTA is stored for 10 years. After the expiration of this term DKRZ-LTA will not remove these data from its archive. However, regular data curation will be stopped, i.e., regular validation of access and integrity (see below) will terminate, unless the data owner opts for a new storage period.

Data Integrity

Upon delivery of data, DKRZ staff accepts data objects and its linked metadata, checks its completeness and compliance with community and preservation standards, helps data producers to fix errors or deviations from standards if required, runs checksums and compares them to attached checksums if possible. Only open source data formats are

¹ Deutsches Klimarechenzentrum GmbH

² DKRZ-LTA

³ Deutsche Forschungsgemeinschaft

accepted for preservation. Should individual data objects be stored in proprietary formats, those objects are not accepted for preservation unless they are archived in an open source format as well. Metadata, containing provenance information, is compiled from relevant resources and kept in a SQL database. The maintenance, including backup, of this database follows recommended practices by this ensuring metadata integrity and access. On a scheduled basis access and integrity of data objects is validated. In case of errors all necessary steps will be taken to re-enable access to the data objects in its original form.

Data Access

Data objects are made accessible by persistent links. This can be Digital Object Identifiers (DOI), any other data handles as, e.g., Persistent Identifiers (PID) or stable links to catalogue entries. The access to data objects is regulated. Only DKRZ-LTA staff is authorized to upload data into the archive. Download is possible for registered users via password authentication; restrictions can be implemented if required in consultation with the data owner.

Curation

DKRZ-LTA is continuously involved in diverse scientific and technological developments. By this DKRZ-LTA staff is in a position to monitor development of community needs and evolution of data formats. Should the need for changes in DKRZ-LTA come up necessary steps are taken. No modification of the data themselves will be performed by DKRZ-LTA or the data owner after the data is archived. Should there be any need for changes, modified data objects can be archived as new versions while keeping the old ones.

Storage Infrastructure

DKRZ-LTA uses the DKRZ storage infrastructure for its data holdings. In compliance with quality control and security specifications this DKRZ infrastructure is specifically designed to accommodate scalability, reliability and sustainability. Data in the DKRZ Long Term Data Archive is kept in a HSM system at least on two tape copies, internal and external. DKRZ staff actively monitors user requirements and technologies, by that adapting DKRZ-LTA preservation practices.

*For issues about security and for a risk assessment see the document *DKRZ-LTA Risk Assessment* which can be found, e.g., on docs.wdc-climate.de .*

For all matters concerning general protection of personal data see the DKRZ website at <https://www.dkrz.de/about-en/contact/en-datenschutzhinweise> . This not only refers to data access by web browser but also by any API and by the jblob download tool.

Contacts

In case any questions regarding this preservation and storage policy arise please contact DKRZ-LTA user support at data@dkrz.de.