DKRZ[1] Long Term Archive[2]: supported formats within data preservation

Revision History

Revision	Author	Scope
July 2016	DKRZ Datamanagement	public release

Policy Statement

Selection of data formats is crucial for data preservation which 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 DFG[3], 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 kept for 10 years. After the expiration of this term DKRZ-LTA will not actively remove this data from its archive although the obligation to preserve it no longer exists.

Supported data formats

Only open source data formats are 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.
As of today this is the list open supported data formats:

Network Common Data Format (NetCDF)

NetCDF is a set of interfaces for array-oriented data access and a freely distributed

[1] Deutsches Klimarechenzentrum GmbH
[2] DKRZ-LTA
[3] Deutsche Foschungsgemeinschaft

collection of data access libraries for C, Fortran, C++, Java, and other languages. The netCDF libraries support a machine-independent format for representing scientific data. Together, the interfaces, libraries, and format support the creation, access, and sharing of scientific data.
More information on NetCDF can be found at http://www.unidata.ucar.edu/netcdf
DKRZ-LTA strongly recommends to use the NetCDF CF metadata convention. More information on CF can be found on http://cfconventions.org . The CF conventions define metadata that provide a definitive description of what the data in each variable represents, and the spatial and temporal properties of the data.

GRIB

GRIdded Binary (GRIB) is a bit-oriented data exchange and storage format approved by the World Meteorological Organization (WMO). More information can be found on http://www.wmo.int/pages/prog/www/WDM/Guides/Guide-binary-2.html.

GeoTIFF

"GeoTIFF" refers to TIFF files which have geographic (or cartographic) data embedded as tags within the TIFF file. The geographic data can then be used to position the image in the correct location and geometry on the screen of a geographic information display. More information can be found at http://trac.osgeo.org/geotiff/.

Contacts

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