

World Data Center for Climate (WDCC) – Preservation and Storage Policy –

version v2.1; 04 October 2023

How to cite:

World Data Center for Climate (WDCC) (2023, October 04), Preservation and Storage Policy, Version v2.1, World Data Center for Climate (WDCC) at the German Computing Center (Deutsches Klimarechenzentrum DKRZ).

https://www.wdc-climate.de/docs/WDCC_PreservationAndStoragePolicy.pdf (last accessed: YYYY-MM-DD).

This document describes the policy for the long-term preservation of digital objects in the World Data Center for Climate (WDCC), hosted by the German Climate Computing Center (Deutsches Klimarechenzentrum, DKRZ), Hamburg, Germany.

Preservation Purpose and Guiding Principles

The main purpose of the WDCC is to actively preserve data relevant for climate science in order to make them available for long-term reuse. The preserved data foremost include output from numerical climate models whereby the focus is on post-processed data and not on raw model output.

The WDCC policy is guided by the FAIR principles and accepted standards like ISO16363 (audit and certification of trustworthy digital repositories) and OAIS (ISO14721). Furthermore, customer-acknowledged guidelines like “Rules of good scientific practice” from the Max Planck Society and the German Research Foundation (DFG), are taken into consideration.

Storage Duration

WDCC ensures the accessibility and curation of the archived data and metadata for at least 10 years. Beyond this period, WDCC ensures further accessibility and does not remove data from its archive. However, data curation upon these data will be reduced, except for data that are still regularly downloaded.

Supported File Formats

The WDCC only supports open source formats for preservation that are well established in the climate science community. The WDCC strongly recommends the use of:

- **Network Common Data Format (NetCDF)**

NetCDF is a binary, non-proprietary file format that is self-describing. This means that the metadata describing the data can be stored in the NetCDF file itself. More information on NetCDF can be found at <http://www.unidata.ucar.edu/netcdf>

WDCC strongly recommends usage of the Climate and Forecast (CF) metadata convention to standardize the NetCDF data. More information on CF can be found on <http://cfconventions.org>.

- **GRIdded Binary (GRIB)**

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>.

In special cases, the WDCC also accepts proprietary format for preservation. It is, however, mandatory that the data objects are additionally provided and archived in an open source format.

Data Integrity

Upon delivery of data, WDCC 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 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 are 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 via persistent URLs to the catalog entries. Additionally, WDCC assigned Digital Object Identifiers (DOI) for data collections and Persistent Identifiers (PID) for datasets. The access to data objects is regulated. Only WDCC staff are 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

The WDCC consistently ensures active preservation by adapting metadata to updates of the Community Standards and incorporating technical advancements (e.g. linked data, PIDs for CF names), with the overarching goal of enhancing the FAIRness of WDCC data publications.

During the archiving process the curation includes checks of data and metadata, while the long-term curation comprises the maintenance and updates of the metadata, like reference lists, author contacts, errata information.

Once archived in the WDCC, the data itself will not be modified. However, the data provider has the option to supply modified data objects, which will subsequently be archived as new versions while retaining the old ones.

In case that a data provider withdraws its data after the publication, the WDCC curators will disable the access to the data, but not delete data. The landing page of the data object will still be preserved, but the WDCC curator will mark the metadata objects as withdrawn.

In the extremely unlikely event of data damage in the archive that cannot be recovered, the landing page of the data object will also be preserved, but the WDCC curator will mark the metadata objects as inaccessible.

Storage Infrastructure

WDCC uses the DKRZ storage infrastructure for its data holdings. In compliance with quality control and security specifications, the DKRZ infrastructure is specifically designed to address scalability, reliability and sustainability. Data in the WDCC is kept in a hierarchical storage management (HSM) system at least on two tape copies, internal and external. DKRZ staff actively monitors user requirements and technologies, by adapting WDCC preservation practices. For issues about security and for a risk assessment see the document WDCC 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 WDCC user support at data@wdc-climate.de.