Assessment Information

CoreTrustSeal Requirements 2017–2019

Repository:	World Data Centre for Climate (WDCC)
Website:	http://www.wdc-climate.de
Certification Date:	17 December 2018
This repository is owned by:	German Climate Computing Centre (DKRZ)

Core Trustworthy Data Repository Requirements

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type.

Domain or subject-based repository

Publication repository

Library/Museum/Archives

Research project repository

Other (please describe)

Comments

The WDCC is maintained by the department Data Management at the German Climate Computing Center (DKRZ).

The services of the DKRZ LTA comprise among others:

- Data ingest, archival storage, management, administration and access operated by the CERA (Climate and Environmental Retrieval and Archiving) data and information system. The typical archiving duration is >10 years.
- A DOI (Digital Object Identifier) data publication service for the direct integration of research data in scientific publications after thorough checks by peers, qualified personnel, or the data producer.
- An editorial process which ensures the quality of metadata and research data in close collaboration with the data creators.
- Hosting of various data centres / collections, e.g. the IPCC Data Distribution Center (IPCC-DDC) and collaboration within scientific or technical projects.

Brief Description of the Repository's Designated Community.

The mission of the "Long Term Archive" (DKRZ LTA) is to provide central data preservation support for the climate modelling research community. For this service the DKRZ LTA has received the Data Seal of Approval (DSA) and acts as "DKRZ - WDC Climate" (WDCC).

Small amounts of data from other parts of Earth System sciences are included but these fields are not in focus.

Level of Curation Performed.

A. Content distributed as deposited
B. Basic curation – e.g. brief checking; addition of basic metadata or documentation
C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation

Comments

As repository, we decline to edit scientific data unless definitely instructed by the data creator. Instead, we prefer to annotate possible data accuracy issues or we ask the data creator to do the necessary edits.

Strong emphasis is put on unique identification of data and possible (re)use of data within the designated community. Further details on the concept of the data publication process and quality control can be found at the DOI publication process webpage of DataCite (datacite.org).

The different quality levels resulting from the quality assessment process correspond to the different project requirements. The WDCC developed a common quality assessment with the Quality Maturity Matrix method.

Data Submission Preparation Guide: http://cera-www.dkrz.de/docs/DataSubmissionPreparationGuide.pdf

CERA2 Metadata Submission Guide: https://cera-www.dkrz.de/docs/CERA2MetadataSubmissionGuide.pdf

DOI publication process: http://doi.wdc-climate.de

Quality Maturity Matrix (poster at EGU 2015):
https://www.dkrz.de/pdfs/poster/Hoeck_et_al_EGU2015_maturitymatrices_15apr.pdf

WDCC supported formats: http://cera-www.dkrz.de/docs/DKRZ-LTA-Formats.pdf

CF conventions: http://cfconventions.org/faq.html

WDCC preservation and storage policy: http://cera-www.dkrz.de/docs/DKRZ-LTA-PreservationAndStoragePolicy.pdf

Outsource Partners. If applicable, please list them.

We do not outsource our data related functions.

As part of multicopy redundancy we have a copy of the data at the central computing centre of the Max Planck Society at Garching near Munich, 800km from DKRZ.

Hardware maintenance is partly outsourced to the manufacturers.

Other Relevant Information.

Besides its data centre, DKRZ also constitutes an outstanding research infrastructure to compute model-based simulations of global and regional climate change and its effects. This mission refers to the High-Tech Strategy for climate protection as presented at the second Climate Research Summit in Berlin in October 2007. It is this high performance computer, where much of the data at WDCC comes from.

This work and the data produced are directly and/or indirectly used on the international political level by the International Panel on Climate Change (IPCC) as well as by other scientists worldwide. WDCC runs an IPCC Data Distribution Centre (IPCC-DDC). We hold about 5 Petabytes of data online accessible on tapes.

At DKRZ, the necessary transparency of research, re-use of data in larger research fields, and increased quality of data and metadata led to two complementary components of a data management system.

The Earth System Grid Federation (ESGF) data dissemination infrastructure (esgf.llnl.gov) is focused on the needs of users as partners in globally running projects (project & analysis phase). It includes replication tools, detailed global standards on project level, and efficient search for the data to download. ESGF is a global collaboration with distributed data nodes and large quantities of disk storage. During the community review data still can be enhanced by new versions or withdrawn. Project specific data quality assurance procedures are supported. Data published in ESGF can in principle be transferred to WDCC DKRZ long-term archive later on.

In contrast to ESGF, DKRZ's digital long-term archive (DKRZ-LTA) aims for long term data holding and data reuse requiring high generic metadata quality standards (archiving and bibliometric phase).

For a mission description cf the DKRZ web page at https://www.dkrz.de/about-en.

For information on user statistics: https://www.dkrz.de/up/systems/wdcc/statistics.

Reviewer Entry

Accept or send back to applicant for modification:
Accept

Comments:

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

Compliance Level: 4

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Implemented: This guideline has been fully implemented for the needs of our repository.

The DKRZ's mission is fixed in its shareholders' agreement which is also signed by the German Ministry of Education & Research: Computing power, data management, and guidance to users. The German Climate Computing Centre (DKRZ) has the mission to provide tools and the associated services which are needed to investigate the processes in the climate system and to ensure long term availability of the results. This also yields for the WDCC as integral part of the DKRZ that was established as World Data Centre in 2003.

As a national service provider, DKRZ operates a supercomputer centre to enable climate simulations and provides the scientific users with the technical infrastructure needed for the processing, analysis, and long term preservation of climate data. This also includes support for related application software, and advice and support in data processing issues. Finally, DKRZ also participates in national and international joint projects with the aim of improving the infrastructure for climate modelling.

The WDCC helps the data producers to comply with rules of good scientific practice as called for by the WDS Data Policy, DFG (German Research Foundation), MPG (Max Planck Society) or others. This comprises formatting, checking, and post processing of the data, as well as converting them into time series per physical variable which are easy to download for the user.

The WDCC long term archive preservation and storage policy is published and is relevant for all staff and users. The CERA data archive presently (2018-02) comprises about 1180 thousand datasets which are grouped into 101 projects with 1801 experiments.

The WDCC homepage at http://www.wdc-climate.de describes the mission of WDCC.

For a mission description of DKRZ, cf the DKRZ web page at https://www.dkrz.de/about-en/.

WDCC preservation and storage policy: http://cera-www.dkrz.de/docs/DKRZ-LTA-PreservationAndStoragePolicy.pdf

WDS Data Policy: http://www.icsu-wds.org/organization/data-policy

Rules of good scientific practice at Max Planck Society: http://www.mpg.de/197494/rulesScientificPractice.pdf

For a short history of the DKRZ as supercomputing centre, see https://www.dkrz.de/about-en/aufgaben/dkrz-geschichte/dkrz-geschichte?set_language=en&cl=en

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

II. Licenses

Compliance Level: 4

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Implemented: This guideline has been fully implemented for the needs of our repository.

The WDCC enforces the Creative Commons Licenses "Attribution" and "Attribution/Non-Commercial" as far as those conditions are not in any way modified by the data creator (see terms.wdc-climate.de). Until now, we did not get notice of any violations of these licenses by users. As the data licenses are relatively open, we until now did not get notice of any violations of them by users. In case we would, our first contact was the owner of the intellectual rights. Via access control the offending account can be blocked.

Applicable licenses regarding the use of research data are by default addressed in our terms of use and if necessary in the metadata of each data object: http://terms.wdc-climate.de.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

III. Continuity of access

Compliance Level: 4

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Implemented: This guideline has been fully implemented for the needs of our repository.

The WDCC is hosted by the Data Management department of the DKRZ which is funded by the German Government and by the four shareholders Max Planck Society, University of Hamburg, Helmholtz Centre Geesthacht, and the Alfred Wegener Institute. The link to these shareholders and to two other German earth system World Data Centres ensures ongoing access and preservation of the data holdings. For continuity of funding see R5.

Furthermore, established arrangements and precautions, such as redundant off-site data storage at Garching / Munich, make it possible in the event of a crisis, to re-enable access to the data within reasonable time.

The disclaimer of WDCC, as written in the "Depositor Agreement", states that WDCC is not liable for the loss or damage of data, although every care is taken to preserve the objects.

Depositor Agreement: http://cera-www.dkrz.de/docs/DKRZ-LTA-DepositorAgreement.pdf

WDCC preservation and storage policy: http://cera-www.dkrz.de/docs/DKRZ-LTA-PreservationAndStoragePolicy.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

IV. Confidentiality/Ethics

Compliance Level: 4

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Implemented: This guideline has been fully implemented for the needs of our repository.

The WDCC does not host sensitive personal data nor any other delicate or confidential data.

WDCC supports the data producers to comply with rules of good scientific practice as called for by the WDS Data Policy, DFG (German Research Foundation), MPG (Max Planck Society) and others.

The "Submission Preparation Guide" (http://cera-www.dkrz.de/docs/SubmissionPreparationGuide.pdf) contains a checklist for depositing, which helps the data providers to verify and organise their materials, to carefully consider copyrights, download permissions, licenses, etc.

It also contains links to further documents, an important one being a link to the "Depositor agreement" which lays out the legal basis for the deposit.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

V. Organizational infrastructure

Compliance Level: 4

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Implemented: This guideline has been fully implemented for the needs of our repository.

The WDCC is hosted by the Data Management department of the DKRZ which is funded by its shareholders and the German Ministry for Research and Technology. In October 2017 in addition, Helmholtz Community, Max Planck Society and the City of Hamburg singed a mutual contract for an ongoing funding of necessary upgrades of DKRZ's supercomputing technology every five years. So the DKRZ as host of WDCC and hence the WDCC as integral part of it has robust funding commitments.

DKRZ has a staff of 80 employees, of which more than 25 work in the Data Management department, 10 of them on permanent posts. About half of the data management staff is concerned with Long Term Archiving. They regularly take part in training courses on, e.g., programming or web security. By career, the majority of them come from information science and physics. They take part in and make contributions to international conferences; some are members of steering groups of projects like ESGF, RDA, ENES and CMIP.

Skills of employees: The team has expertise in the areas of data management, data archiving & preservation, data formats, controlled vocabularies, quality assessment, data usage, data publication and registration. LTA staff strives to improve its knowledge in these areas by close cooperation in scientific and technical projects, training and participation in a technology watch team. Members of WDCC actively participate in working and interest groups of RDA and WDS.

In addition, DKRZ holds courses on various topics. These courses are partly in collaboration with the Research Data Alliance (RDA) or with international projects like EUDAT.

The work and its quality are regularly monitored by a Scientific Steering Board and a dedicated User Group (see R6).

Reviewer Entry

Accept or send back to applicant for modification:
Accept

Comments:

VI. Expert guidance

Compliance Level: 4

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).

Implemented: This guideline has been fully implemented.

The work and its quality are regularly supervised and judged by a Scientific Steering Committee. This comprises members from DKRZ's shareholders and from important institutes of the German Earth System Science community, sometimes joined by a representative of the Ministry.

In addition, a dedicated User Group supervises DKRZ's work from the German users' view. User queries are collected and evaluated.

Both groups give advice and guidance via reports and personal communication. They meet two times a year. These meetings are joined by DKRZ staff members.

For the Scientific Steering Committee, see https://www.dkrz.de/about-en/Organisation/wla.

For the DKRZ User Group, see (in German) https://www.dkrz.de/about-en/organization/nutzervertretung?set_language=de&cl=de.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

Compliance Level: 3

R7. The repository guarantees the integrity and authenticity of the data.

Checksums of the files are verified. Mostly they are calculated already at the producers' site; they should be stored by the producer as metadata and can be checked at the data users' site after download.

The quality assurance of the data and metadata is part of the publication process and documented in the CERA database. Changes to data objects are not allowed, new versions of data objects are possible but they are versioned and treated as independent objects and equipped with mutual pointers. Old versions are kept in the archive. This also applies to the corresponding metadata entries.

1. Data is delivered to us from clients who are known to us personally or are designated by someone whom we know personally.

2. Data from projects with their own data infrastructure (like ESGF/CMIP5) in which WDCC participates are long-term archived according to specific workflows. Those are agreed on by the project and WDCC. The authenticity underlies the project regulations.

Archiving of data is only allowed if metadata is provided. The qualification of metadata follows the standard as described in https://www.dkrz.de/up/services/data-distribution/data-publication/quality-assurance-of-data. If needed, links to external resources (data and/or metadata) are kept as well. Provenance data is to be included. All links are kept in the Relational Database CERA.

Deliberate changes to the metadata or to the data structure are documented in a dedicated table of the metadata base.

Guidelines regarding integrity of data objects and metadata are covered in the WDCC Preservation and Storage policy.

The review of the metadata is part of the DOI data publication process. After DOI assignment, the complete data is accessible by the registered DOI.

Future plans comprise the extension of cross referencing to other publications and to other publishers.

WDCC Preservation and Storage Policy: http://cera-www.dkrz.de/docs/DKRZ-LTA-PreservationAndStoragePolicy.pdf

Submission Preparation Guide: http://cera-www.dkrz.de/docs/SubmissionPreparationGuide.pdf

Concept of the DKRZ DOI data publication process: https://www.dkrz.de/daten-en/data-services/Datapublication/konzept-datacite-tib-metadatakernel

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

VIII. Appraisal

Compliance Level: 4

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Implemented: This guideline has been fully implemented for the needs of our repository.

The “Submission Preparation Guide” contains a checklist for depositing, which helps the data providers to verify and organise their materials. This guide also contains links to further documents, an important one being a link to the “Depositor Agreement” which lays out the legal basis for the deposit.

Different levels of quality assessment ensure the quality of data and metadata and the consistency between those two. Strong emphasis is put on unique identification of data and possible (re)use of data within the designated community. Further details on the concept of the data publication process and quality control can be found at the DOI publication process webpage.

The different quality levels of quality assessment correspond to the different project requirements. The WDCC developed a common quality assessment with the Quality Maturity Matrix method.

Independently of specific project requirements, the WDCC asks for a bundle of information that should be deposited. Required are the core metadata of CERA-2. Information about methodology, quality checks and validation against independent data should be deposited at the latest in the DOI publication process step. A standard set of checks is applied to all submissions; further project specific checks can be defined and implemented.

At DKRZ as a thematic data centre the number of formats is limited.

In general, WDCC accepts long-lasting open source formats for preservation. WDCC supports the data producers in the conversion of data into an accepted format e.g. NetCDF/CF and WMO GRIB. Both formats are open standards. As “Open Source” and coming with a huge number of available open source analysis tools these formats will not become obsolete within the promised storage duration of 10 years.

NetCDF/CF brings in the CF convention which is self-contained. So for example the CF Standard Names attempt to be as general and well-defined as possible, so the reader does not have to access outside sources to understand the terms.’

For other formats the WDCC asks the data producers to provide a description of the model including a section “how to read the data”. If the data is binary we ask to convert the data in parallel into an ASCII version (txt).

However, should a very unlikely format obsolescence for the supported formats in the promised storage duration occur, WDCC will provide either an on-the-fly format conversion into an actual and accepted format or will do the format conversion offline. This is possible although likely time consuming. The second would be typically done within the framework of media conversion to avoid unnecessary media mounts.

Formats are checked by the repository before data is archived. NetCDF/CF checkers are available for the data producers on the internet.

Data Submission Preparation Guide: http://cera-www.dkrz.de/docs/SubmissionPreparationGuide.pdf

Depositor Agreement: http://cera-www.dkrz.de/docs/DKRZ-LTA-DepositorAgreement.pdf

DOI publication process: https://www.dkrz.de/up/de-services/de-data-distribution/datenpublikation/koncept-datacite-tib-metadatakernel

Quality Maturity Matrix (poster at EGU 2015): https://www.dkrz.de/pdfs/poster/Hoeck_et_al_EGU2015_maturitymatrices_15apr.pdf

Core metadata CERA2: https://www.dkrz.de/daten-en/cera/data_model/tables-attributes-1

WDCC preservation and storage policy: http://cera-www.dkrz.de/docs/DKRZ-LTA-PreservationAndStoragePolicy.pdf

WDCC format policy: http://cera-www.dkrz.de/docs/DKRZ-LTA-Formats.pdf

NetCDF: http://www.unidata.ucar.edu/software/netcdf

CF convention: http://www.cfconventions.org

WMO GRIB format: http://www.wmo.int/pages/prog/www/WDM/Guides/Guide-binary-2.html

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

IX. Documented storage procedures

Compliance Level: 4

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Implemented: This guideline has been fully implemented for the needs of our repository.

The WDCC long term archive preservation and storage policy is published and is relevant for all WDCC staff and users.

All processes on the data and/or metadata are documented in a dedicated CERA metadata table with timestamp and responsible person.

DKRZ locally keeps the data. In addition, a copy of the entire database is held at the Max Planck Society's central data repository at Garching near Munich. Media are changed after five years.

WDCC preservation and storage policy: http://cera-www.dkrz.de/docs/DKRZ-LTA-PreservationAndStoragePolicy.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

X. Preservation plan

Compliance Level: 4

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Implemented: This guideline has been fully implemented for the needs of our repository.

The archive is run by the DKRZ which is a Limited Liability Company. The organizational structure of DKRZ is described at https://www.dkrz.de/about-en/Organisation. Special contracts with data producers are not required if they are a member of one of our shareholders. For other data producers, regulations are laid out in the Depositor Agreement or individual contracts based on WDS regulations apply. The terms of use of the archive are published and apply to all clients of WDCC.

WDCC guarantees storage for ten years. After 10 years DKRZ will not remove these data from its archive. However, regular data curation will be stopped, i.e., regular validation of access and integrity will terminate, unless, after being contacted by WDCC, the data owner opts for a new storage period.

Data with disclosure risks are not handled. Few submissions are under embargo to allow data producers the first use of data.

DKRZ structure: https://www.dkrz.de/about-en/Organisation

Terms of Use: http://terms.wdc-climate.de/

For the rights, see the Depositor Agreement (http://cera-www.dkrz.de/docs/DKRZ-LTA-DepositorAgreement.pdf).

WDCC preservation and storage policy: http://cera-www.dkrz.de/docs/DKRZ-LTA-PreservationAndStoragePolicy.pdf (new Version of Oct-2018).

Reviewer Entry

Accept or send back to applicant for modification:
Accept

Comments:

XI. Data quality

Compliance Level: 4

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Different levels of quality assessment ensure the quality of data and metadata and the consistency between those two. Strong emphasis is put on unique identification of data and possible (re)use of data within the designated community. So rich metadata is expected from the data producer. Further details on the concept of the data publication process and quality control can be found at the DOI publication process webpage.

Independently of specific project requirements, the WDCC asks for a bundle of information that should be deposited. Required are the core metadata of CERA-2, if applicable. Information about methodology, quality checks and/or validation against independent data should be deposited at the latest in the DOI publication process step. A standard set of checks is applied to all submissions; further project specific checks can be defined and implemented.

References of type “is documented in” and/or “is cited by” are provisioned and can be stored in the database.

DOI publication process:
https://www.dkrz.de/up/de-services/de-data-distribution/datenpublikation
https://www.dkrz.de/up/de-services/de-data-distribution/datenpublikation/qualitaetskontrolle-fur-primaerdaten

Core metadata of the CERA2 database: https://www.dkrz.de/daten-en/cera/data_model/tables-attributes-1

Reviewer Entry

Accept or send back to applicant for modification:
Accept

Comments:

XII. Workflows

Compliance Level: 4

R12. Archiving takes place according to defined workflows from ingest to dissemination.

The DKRZ LTA workflows across the data life cycle are published.

Selection process and type of data: The WDCC is a thematically oriented archive. The focus is on output from numerical climate models. The archival process is integrated into and adapted to the workflow of the scientists. Archiving could take place at varying phases of the workflow: either at a very early stage in order to make data reusable and properly citable or at a very late stage where the main motivation is the application of the Rules of Good Scientific Practice. All storage work is done in close coordination with the data provider.

A continuous adjustment of this integration is established.

WDCC maintains an extensive Graphical User Interface (GUI) to enable the users to find and retrieve data via rich metadata.

Approach towards data that do not fall within the mission: The WDCC is member of KomFor. KomFor supplies information where data can be submitted. Data repositories with different scope are members of KomFor.

Its services are connected with existing ones provided by:

- Re3data
- The World Data System

In cooperation with partners from these networks we support data producers to find the most appropriate repository.

The repository team maintains a dialogue with the researchers who want to deposit data. The information about the “Terms of Use” and the archiving workflow is part of this negotiation.

Decision-making process for archival data transformation:

Data transformation within the archive only takes place in close cooperation with data producers (as far as they are available) or in cooperation with the relevant scientific community.

Supporting Links and Documentation

- DKRZ LTA Workflows across data life cycle:
  https://www.dkrz.de/up/services/data-management
  https://www.dkrz.de/up/services/data-management/LTA
  https://www.dkrz.de/up/services/data-management/LTA/archiving-concept

https://www.dkrz.de/pdfs/poster/SC11_P1_DataLifeCycle_klein.pdf
https://www.dkrz.de/pdfs/poster/DataServices_ENG.pdf

- WDCC preservation and storage policy: http://cera-www.dkrz.de/docs/DKRZ-LTA-PreservationAndStoragePolicy.pdf
- Submission Preparation Guide: http://cera-www.dkrz.de/docs/SubmissionPreparationGuide.pdf
- Rules of Good Scientific Practice: http://www.mpg.de/197494/rulesScientificPractice.pdf
- List of recent projects: https://www.dkrz.de/daten-en/wdcc/projects_cooperations
- KomFor: http://www.komfor.net/
- KomFor submission: http://www.komfor.net/submit-data.html
- Re3data: http://service.re3data.org/search/
- WDS members: http://www.icsu-wds.org/community/membership
- Terms of Use: http://terms.wdc-climate.de
- Coupled Model Intercomparison Project, Phase 6 (CMIP6): http://pcmdi.llnl.gov/CMIP6/
- Data Distribution Centre of the Intergovernmental Panel on Climate Change (IPCC-DDC): http://www.ipcc-data.org/
- RDA: https://rd-alliance.org/

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

XIII. Data discovery and identification

Compliance Level: 4

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

All data can be searched and downloaded free of charge through one of our web interfaces http://cera-www.dkrz.de/. Browse and Search are available. Metadata is also offered via OAI-PMH. WDCC offers a DataCite DOI data publication for long-term archived data as well as ePIC PIDs. A proposed citation text is offered to users. Published data is granted permanent access by the assigned unique identifier DOI, data and citation metadata remain unchanged.

As a regular member of WDS and thereby accepting the regulations WDCC takes over responsibilities according to these rules. WDCC is listed at re3data. All DOI landing pages show the DOI and a citation hint. In 2017, we implemented the machine readable landing pages which has been described in Nature-Data by a paper of Martin Fenner (DataCite) et al.

WDCC maintains an extensive Graphical User Interface (GUI) to enable the users to find and retrieve data via rich metadata.

OAI-PMH access: http://c3grid1.dkrz.de:8080/oai/oaisearch.do

DOI publications in the CERA catalogue: https://cera-www.dkrz.de/WDCC/ui/cerasearch/q?query=*%3A*&ref_type_name_s=Citation-DOI&page=0&rows=15

WDS regulations: https://www.icsu-wds.org/files/wds-certification-summary-11-june-2012.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

XIV. Data reuse

Compliance Level: 4

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Implemented: This guideline has been fully implemented for the needs of our repository.

See answers to R0 Comments, R8, and various others.

Independently of specific project requirements, the WDCC asks for a bundle of information that is deposited together with the data. Required are the core metadata of CERA2. Information about methodology, quality checks and validation against independent data should be deposited at the latest at the DOI publication process step. A standard set of checks is applied to all submissions but further project specific checks can be defined and implemented.

At present, DKRZ has no migration plans for data or metadata.

Core metadata CERA2: https://www.dkrz.de/daten-en/cera/data_model/tables-attributes-1

Reviewer Entry

Accept or send back to applicant for modification:
Accept

Comments:

TECHNOLOGY

XV. Technical infrastructure

Compliance Level: 4

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Implemented: This guideline has been fully implemented for the needs of our repository.

The WDCC at DKRZ follows the guidelines given in the OAIS reference model across the whole workflow. There is an exception with regard to preservation planning where plans are adjusted to administrative, infrastructure (e.g. CERA), and project requirements. Besides other standards, we use OAI-PMH to have our data harvested.

Hardware and software are in line with the necessities of the in house supercomputer and its data output. They are continuously reviewed and every couple of years updated or renewed.

The WDCC is integrated part of the DKRZ and uses common hard- and software. For a detailed description of the hardware like the HPSS of IBM see the DKRZ page at https://www.dkrz.de/systems. The operating system of all machines is UNIX or a derivative. At DKRZ and WDCC, community software is largely in use, such as netcdf.lib, ESGF (see above), data editing / visualisation / evaluation software. Although we do not handle real time data, bandwidths are comfortable and are continuously monitored.

Mapping of WDCC components to OAI and implementation: https://www.dkrz.de/up/services/data-management/LTA/oais_dkrz_ltamapping.pdf/view

CERA Climate and Environmental Retrieval and Archive (CERA): https://www.dkrz.de/up/systems/cera

Published workflows are, e.g.,

Quadt, F., Düsterhus, A., Höck, H., Lautenschlager, M., Hense, A.V., Hense, A.N. & Dames, M. (2012). Atarrabi - A Workflow System for the Publication of Environmental Data. Data Science Journal, 11, 89-109, http://dx.doi.org/10.2481/dsj.012-027

and

Stockhause, M., Höck, H., Toussaint, F., Lautenschlager, M. (2012). Quality assessment concept of the World Data Center for Climate and its application to CMIP5 data. Geosci. Model Dev., 5, 1023-1032, 2012, http://dx.doi.org/10.5194/gmd-5-1023-2012

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

XVI. Security

Compliance Level: 4

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Implemented: This guideline has been fully implemented for the needs of our repository.

The keystones of security and safety at WDCC are:

Physically: only few persons of trained staff are given admittance to the storage and service hardware. This is well controlled for security reasons as for safety reasons by an admittance system.

Logically: users and staff mainly only have read access to the data. Only few persons of trained staff are given write access for data and metadata ingest. Write and download access is monitored.

Technically: an Uninterruptable Power Supply (UPS) is in place. As fire protection, WDCC has a low oxygen atmosphere in the computer hall.

Disaster recovery: for recovery after disaster, WDCC has offline copies of the data: As part of multicopy redundancy we have a copy at the central computing centre of the Max Planck Society at Garching near Munich, 800km from DKRZ.

As most data are openly distributed, there is little risk for WDCC to be spied out. However, we do have a security officer, as a high performance computing centre usually is an interesting target for hackers.

For further details, see the Risk Assessment available from https://cera-www.dkrz.de/docs/DKRZ-LTA-RiskAssessment.pdf.

For details on plans in case of discontinuity of funding see R3 and R5.

Reviewer Entry

Accept or send back to applicant for modification:
Accept

Comments:

APPLICANT FEEDBACK

Comments/feedback

Thanks for the good cooperation and useful hints.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: