Data life cycle management

DKRZ - DATA INTENSIVE CLIMATE SCIENCE





GUI based catalogue and data access: The World Data Center Climate (WDCC) offers catalogue and

transparent data access by a number of community adapted data portals, which implement community

specific views for the available data services. Examples are data portals for research projects like

ENSEMBLES, for communities like ENES or international programs like IPCC. Additionally the WDCC is

GUI based

catalogue and data

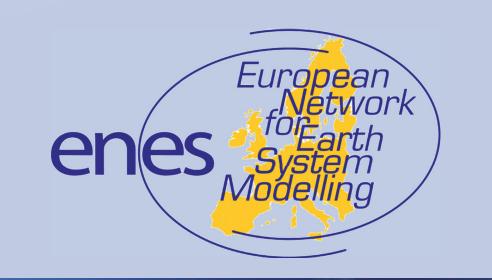
access

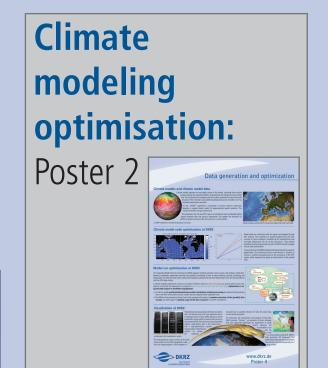
API based direct

data access

data federations

integrated in international data federations via standardized metadata and data interfaces.





DATA SERVICES

climate modeling HLRE II optimisation

consortial runs (community agreed

DATA LIFE CYCLE

Creation

virtual research

environment

long -term archiving

Archiving

data documentation

long-term archiving

scientific data publication

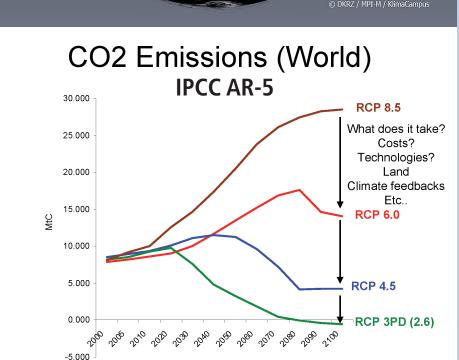
bit-stream preservation

data curation

model calculations)

Dissemination





Data life cycle management

and related services at DKRZ:

the inner circle illustrates the basic

data life cycle elements; the outer

circle illustrastes the related data

services at DKRZ. For every service

at DKRZ anexample is given. Further

details are shown on the referenced

data processing

visualisation

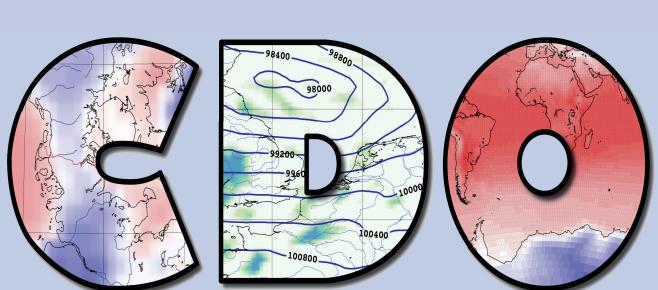
data sharing

quality control

posters.

Evaluation

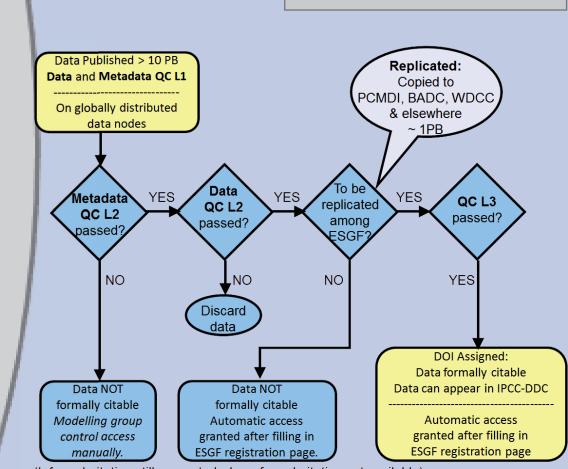
Consortial runs and visualization: DKRZ supports the execution of consortial runs — climate community agreed large model calculations — by an integrated compile and execution environment. Up to 1/3 of HLREII ("Blizzard") is reserved for consortial runs which are of interest for a larger consortium of climate scientists. Present activities are the German climate modeling contribution to the next IPCC assessment report (IPCC-AR5) and the STORM project which calculates a high resolution, ocean eddy resolving climate model for more than 300 years.



Data processing: The climate data operator

used in the international climate research community.

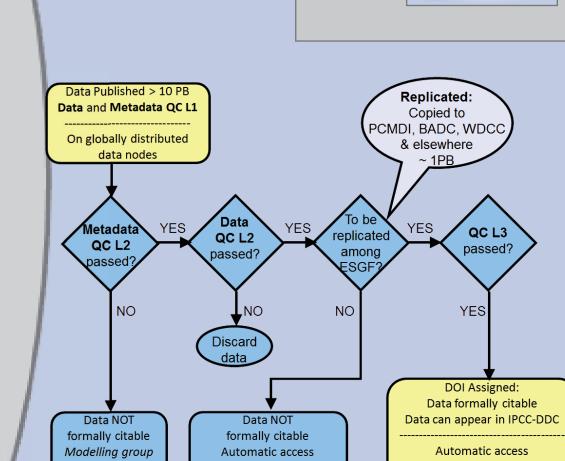




Quality control: Climate model data becomes the basis for far reaching climate change related evaluation and decision processes. This increases the importance of formalized data and data documentation quality control processes. DKRZ coordinates a formally defined three level quality assurance process for the next IPCC-AR5. The third level is directly connected with scientific data

(CDO, https://code.zmaw.de/projects/cdo) toolbox developed by the Max-Planck Institute

for Meteorology, Hamburg, provide the basis for data (post-) processing services at DKRZ. The CDO toolbox is also widely



publication and "freezing" of PetaBytes of model data.

Archive: Container: M **Blobs** DB Layer • When What **CERA** • How Where • Who

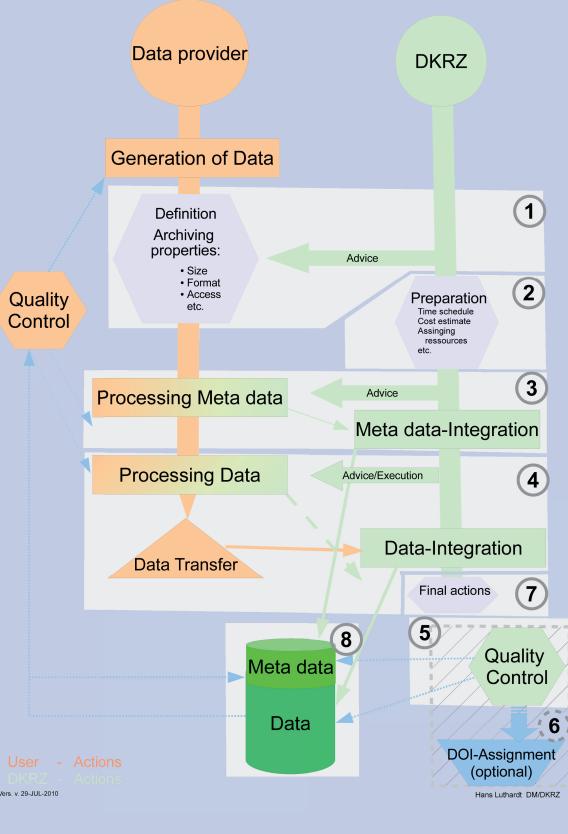
API based direct data access: WDCC offers data access services for data stored in files as well as data stored in a specific DKRZ container format. The storage location (disc or tape) is transparent to the user, only data access times are different.

Data federations: Poster 4

Reference on on the volume of space-time Any publication related to the data together with the Block describes data topic Metadata Entry Status is is the central CERA Block Spatial Reference Status information like the entry's title data quality, processing type and relation to other entrie the project the data belong to a summary of the entry a list of general keywords related to data Distribution reation and review dates of the metadata Distribution information and institutes like distributor cluding access restrictions investigator, and owner of data format, and fees, if Additionally: Modules and Local Extensions Module DATA ACCESS (physical storage)

Data documentation: WDCC uses the CERA-2 metadata model to describe its data entities. CERA-2 has been developed as modular description system which allows for easy adaption to new description requirements by adding new modules without affecting the existing ones. CERA-2 is in operation for more than 10 years now and left practically unchanged since 2000. Long Term Archiving at DKRZ

Long-term archiving: DKRZ offers long term archiving services. Storage period of at least 10 years are supported along with web-based data access. Long-term archiving is requested by the founding agencies in the context of "Rules for Good Scientific Practice" and a prerequisite to enable future research activities. This service thus includes the collection and maintenance of data documentation (metadata).



Scientific data publication: Parallel to publications

in scientific literature the scientific data publication makes data available for use

in scientific articles. After finalizing the scientific data publication process research data have an accurate citation reference, a persistent identifier has been assigned (data are accessible independent from storage location) and the research data are no longer matter of change (scientific results are provable and reproducible). The graphical, web-based application "Atarrabi" has been developed in order to support the scientific data publication process.



Bit stream preservation: Poster 3

min: 304,1 max: 324,2

Warning: 3896 values are null.

OND min; 0 max; 0

Data curation: WDCC at DKRZ offers a metadata visualization tool to provide a web-based, graphical representation of the metadata of a database entity. The metadata visualization tool is used for quality assurance and to check the metadata.



www.dkrz.de Poster 1