

Open Science in the Climate Sciences and the role of DKRZ

CHIST-ERA Project Seminar
Bucharest, 3 Apr 2019

Karsten Peters
Deutsches Klimarechenzentrum (DKRZ)



Data Sharing in the Earth System Sciences

A field of fundamental science relying on global data – data which is produced at dedicated sites and then made available to the community.

IPCC Assessment Reports since 1990: WGI Contribution



IPCC AR5 Working Group I
Climate Change 2013: The Physical Science Basis

ipcc
INTERGOVERNMENTAL PANEL ON climate change
WMO UNEP

Climate Model Data

Large-scale international **climate model intercomparison** efforts have been the **backbone** of the policy relevant **assessment reports of the IPCC*** – all of which would be impossible without Open Access to PetaBytes (PBs) worth of globally distributed research data

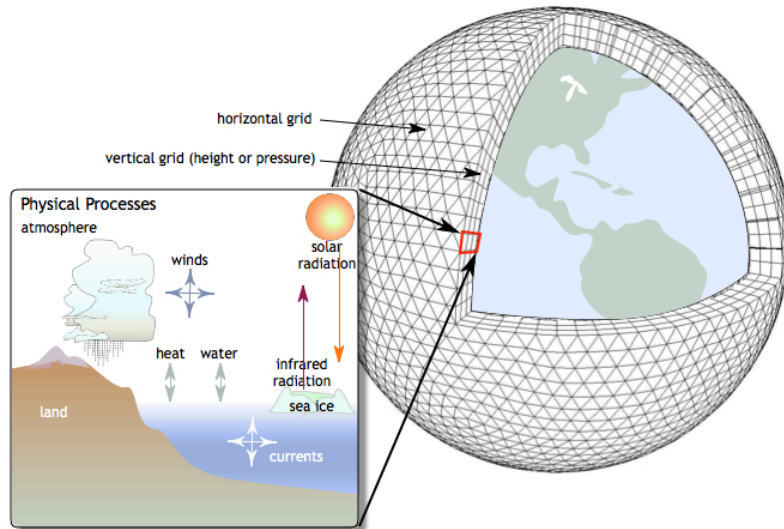
“Policy relevant data have to be openly accessible for informing the public”

*IPCC: Intergovernmental Panel on Climate Change

DKRZ – the beginnings

1978

MPI-M* and University of Hamburg establish a joint computing infrastructure to meet the demands of computational Earth System Science



<http://kiwi.atmos.colostate.edu/cmmmap/learn/modeling/whatIs2.html>

*MPI-M: Max-Planck Institute for Meteorology, Hamburg

1987

A “**window of opportunity**” came up:

- political initiative in Germany to put climate science on the national agenda
- 18Mio DM of available federal funds (on rather short notice)
- A handful of renowned scientists put together a proposal for a supercomputing centre dedicated to Earth System Sciences



Founding of the DKRZ in 1987, in order to serve the German community with adequate **compute and storage capacities and associated services**

Supporting first model intercomparisons for policy decisions

1990

First IPCC Assessment Report including calculations made at DKRZ

- An international effort illustrating the need for Open Data and **efficient ways to get them**

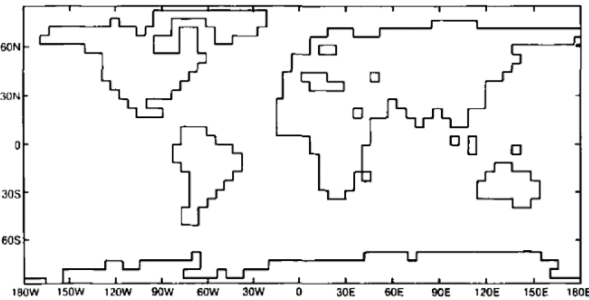
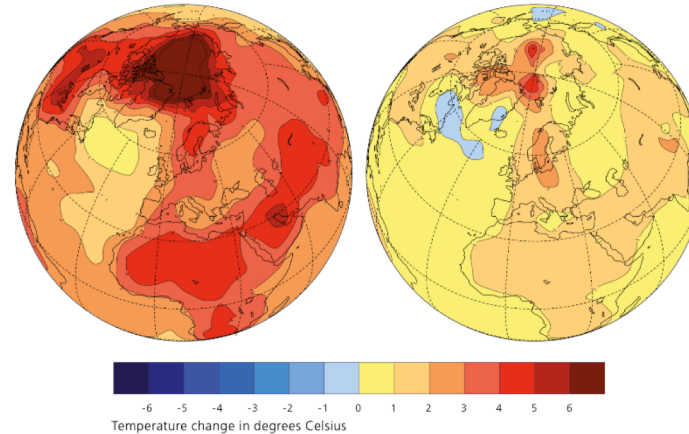


Figure 3.4: The model land sea mask for a typical climate model (T21, ECHAM, after Cubasch et al, 1989)

1992

Earth Summit in Rio de Janeiro

- Simulation data from Hamburg support formulation of the United Nations Framework Convention on Climate Change (UNFCCC)



Reduced greenhouse gas emissions (right) lead to significantly less warming compared to „business as usual“ (left) (presented in Rio de Janeiro, 1992)

1995

Second IPCC Assessment Report, again with important Hamburg contributions

- available infrastructure allows proof of man-made climate change

Establishment of climate data archive and services

Climate and Environmental Retrieval and Archive (CERA) @DKRZ

- established at DKRZ in cooperation with Oracle in **1995, update in 1997, unchanged since**
 - data search and access via the web
 - **1 TB of data in 1998, 5 PB in March 2019**
- **DKRZ begins to take the role of Open Science facilitator in the Earth System Sciences**

IPCC Data Distribution Center (IPCC DDC) @DKRZ

- 1996: IPCC recommends the establishment of a climate data center to assist future reports
- 1997: establishment of the virtual Data Distribution Centre (DDC) by CRU* (UK) and DKRZ
 - Global collection and dissemination of climate model simulation data (<http://ipcc.wdc-climate.de>)

Robot tape archive at DKRZ



*CRU: Climate Research Unit, now the UK-role in IPCC DDC is taken over by CEDA (Climate and Environmental Data Archive)

Establishment of climate data archive and services



World Data Center for Climate (WDCC)* @DKRZ

- 2003: approval as **domain-specific** World Data Center
- **Certified as trusted long-term archive** in 2014 (Data Seal of Approval) and 2018 (Core Trust Seal)

Approved repositories are certified according to the WDS[†] data sharing principles (a.o.):

- Data, metadata, products and information should be fully and openly shared
- Research, education and public domain data made available with minimum time delay and, at best, free of charge
- Sensitive or restricted data should be appropriately labelled and made available on least restrictive basis



*wdc-climate.de

**WDS: World Data System, <https://www.icsu-wds.org>

Enabling data publication as citeable resource

From data archiving and dissemination to data publication

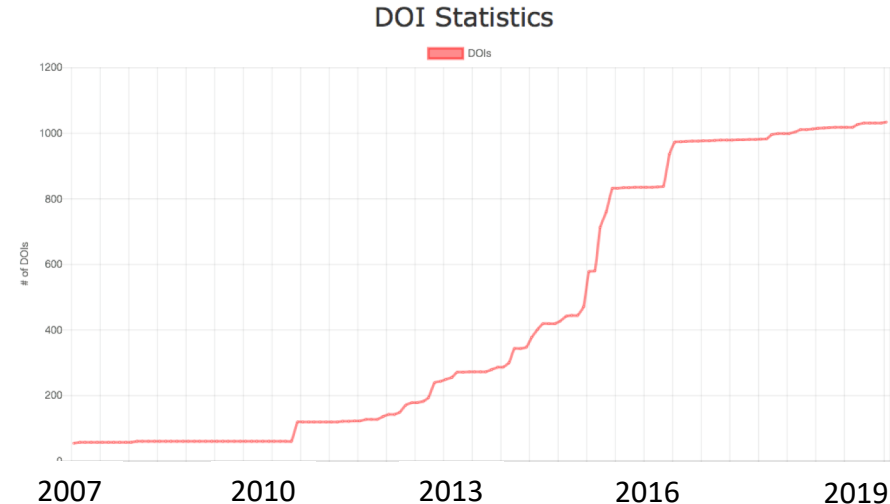
DKRZ active in research activities* enabling the developments of mechanisms to publish scientific datasets as citeable entities

- 2004: **first DOI ever** to be assigned to a dataset was for one in WDCC:

- http://dx.doi.org/10.1594/WDCC/EH4_OPYC_SRES_A2

*2003-2008: **DFG-funded project STD-DOI**

- **development of first-ever methods** to publish digital data in the internet and make them citeable; result: DataCite was founded

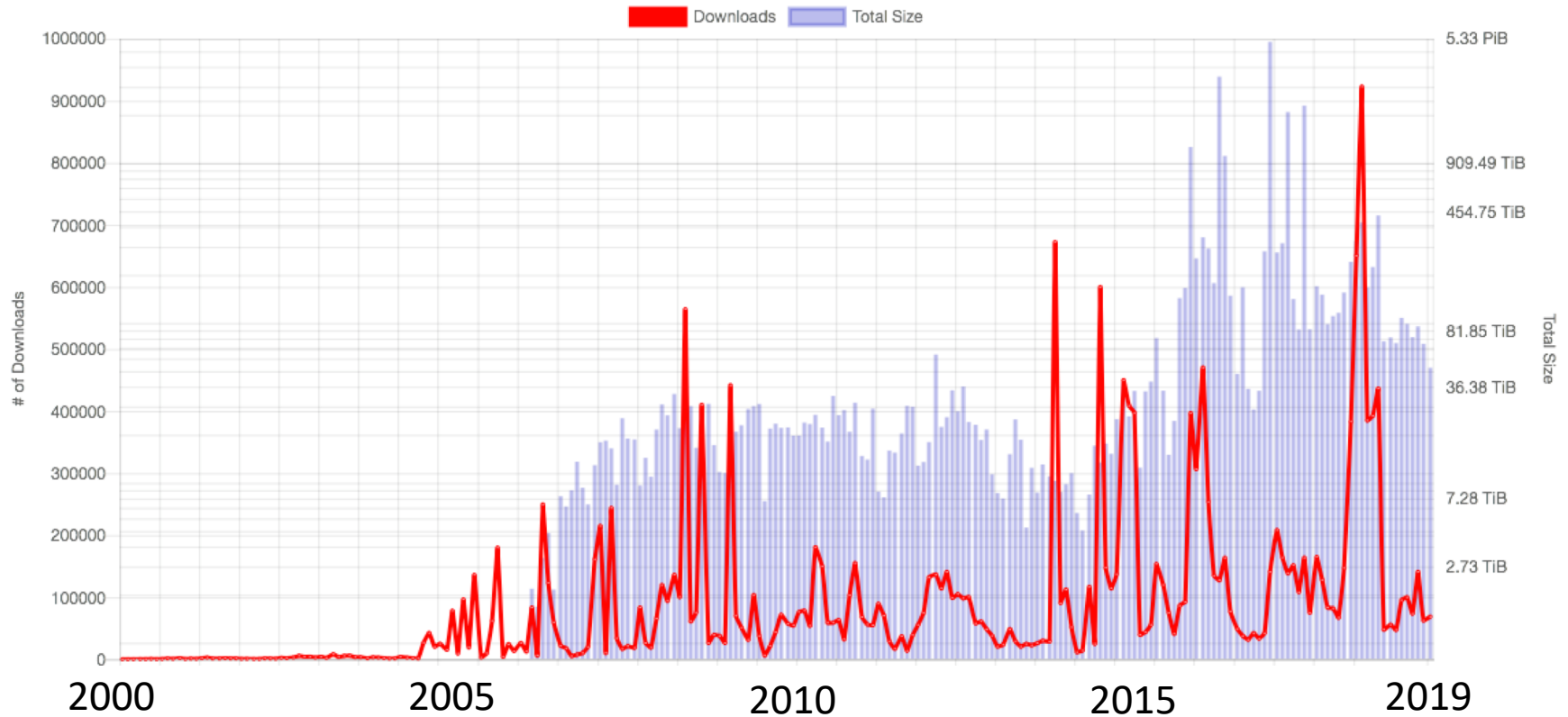


- WDCC has published **1030 DOIs assigned to PBs worth of data**
- DOIs are assigned after meticulous quality control → data are fit for re-use
 - **Compliance with the FAIR** data guiding principles**

**[FAIR](#): Findable, Accessible, Interoperable, Reusable

Re-use of WDCC archived data

Download Statistics



WDCC archived data @DKRZ are being actively re-used (disciplinary and interdisciplinary)

Global dissemination of large-volume datasets

IPCC's Assessment Report rely on the analysis of climate model output generated in globally coordinated intercomparison projects – **the CMIPs**



CMIP3, basis for the 4th Assessment Report (2007)

Approach: collect data in a centralised location in the US and disseminate from there

Data Collection: scientist shipped their data on physical drives across the world, 35 TB in total

Problems:

- Time delays between supply and availability
- Lack of data protocol and data standards
- Erroneous version control and backups
- Central location was single point of failure

The data was later replicated to the IPCC DDC at DKRZ and long-term archived in WDCC to enable re-use

- **Enabling redundancy through independent systems in a global context**

Global dissemination of large-volume datasets

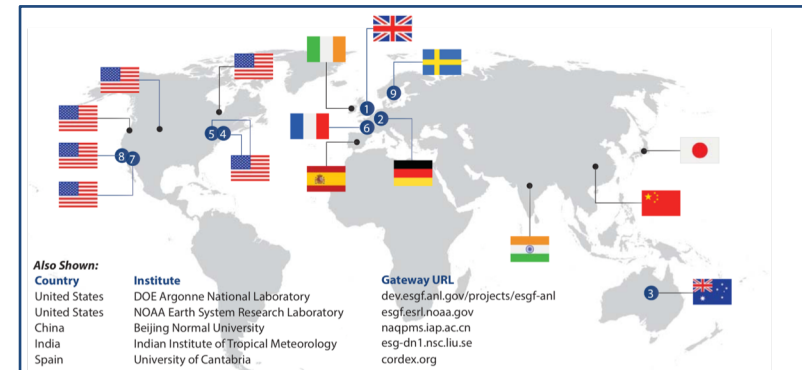
IPCC's Assessment Report rely on the analysis of climate model output generated in globally coordinated intercomparison projects – **the CMIPs**



CMIP5, basis for the 5th Assessment Report (2013) and ongoing

Earth System Grid Federation (ESGF), established 2006

- infrastructure of **globally distributed** data nodes for dissemination of **highly standardised** large-volume Earth System Science data (**ca. 3.5 PB for CMIP5**)
- data nodes back-up each other's data
- data globally accessible through unified interface
- **DKRZ is founding member and one of the core data nodes**
 - DKRZ publishes community-relevant datasets and provides support along the way
 - only ESGF data node linked to a long-term archive



Sustainability of the DKRZ-funding

DKRZ is a non-profit and non-commercial limited company with four shareholders

Shareholder agreement of 1987 secures **cover of running costs** and has **no expiration date**.

Shareholders are

- Max Planck Society via MPI for Meteorology (55%)
- City of Hamburg via Hamburg University (27%)
- Helmholtz Zentrum Geesthacht (HZG) (9%)
- Alfred-Wegener-Institut (AWI) (9%)

However, the agreement **depends on the availability of a state-of-the-art supercomputing system**.

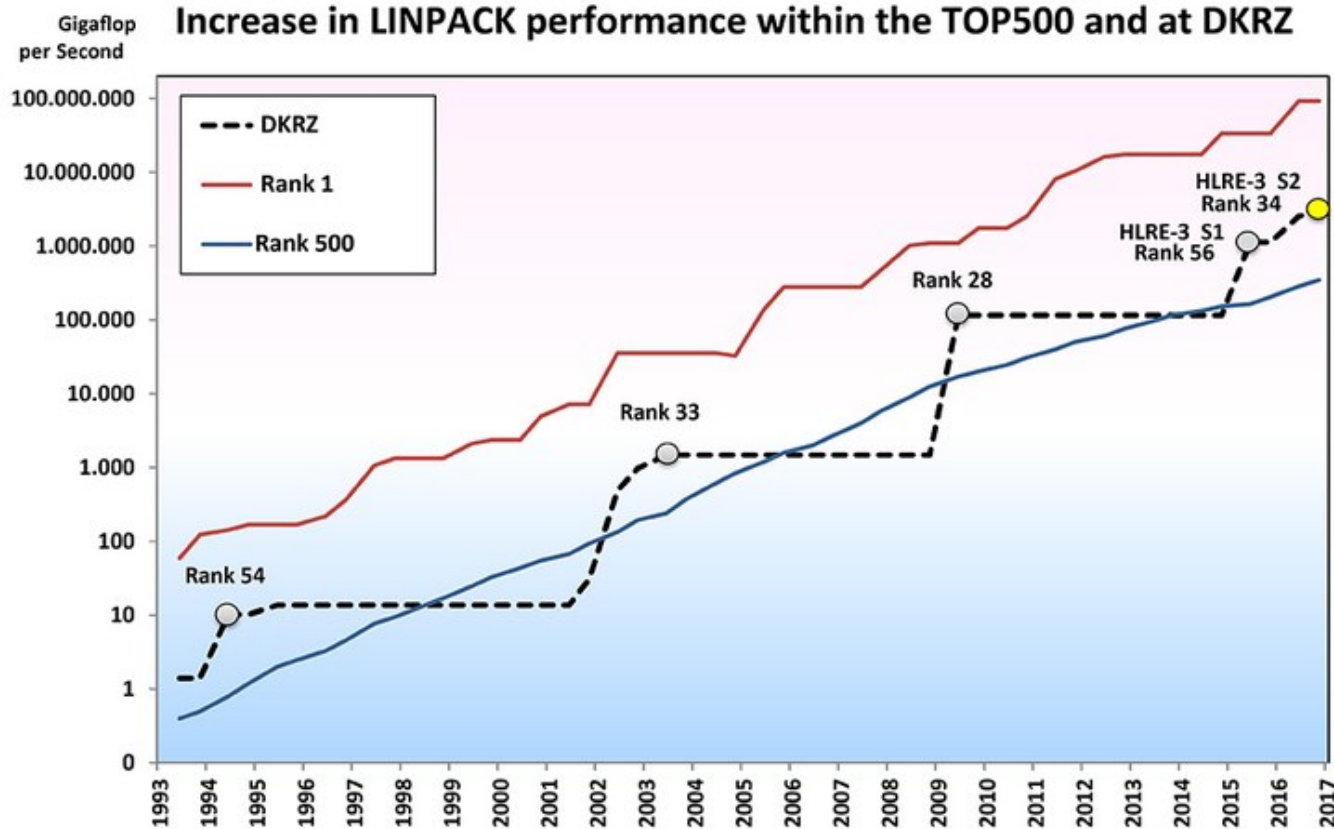
Funding had to always be negotiated for every new machine - the **current machine is the last one funded by the Federal Ministry of Education and Research**

24th October 2017



An agreement to fund future compute systems (ca. 45 Mio EUR per new system) on a regular basis is signed by the shareholders.

Sustainability of the DKRZ-funding



Availability of state-of-the-art HPC-systems at DKRZ throughout the last 30 years indicates the **clear political commitment** for investing in top-tier compute services for top-tier research

Thought experiment – where would we be without DKRZ?

Open Science promotion and facilitation in the field of Earth System Science and beyond **would be less enabled**

Important **developments for data publishing** (DataCite DOIs), accessibility of **long-term archived datasets** and **data standardization** would not have been achieved at the time

The presence of anthropogenic climate change would certainly be less prominent on the political agenda

The German Earth System Science community would definitely not be at the international forefront as it is now



Summary

Through the development of dedicated **data services enabling open and reliable access to research data**, the contributions of DKRZ have shaped Earth System Science as it is today and are also applicable to other disciplines.

Infrastructure-wise, **DKRZ is renowned for facilitating global and open dissemination** of Earth System Science datasets through two independent infrastructures (**WDCC and ESGF**), thus **enabling effective scientific work of an entire research community**. Strong links to **international partner institutions** are essential.

DKRZ has built and is continuously building **in-house expertise and infrastructural capacities** to cope with the demands of data-driven research. This makes DKRZ a **reliable partner in large-scale disciplinary and cross-disciplinary Open Science** inspired efforts, e.g. [EUDAT](#), [European Open Science Cloud](#) and [CMIP6 services](#).



**Thank you very much
for your attention!**