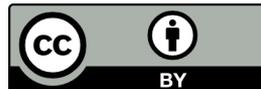


# Long-term archival and global dissemination of climate data at DKRZ

**Karsten Peters**, Stephan Kindermann, Hannes Thiemann

Deutsches Klimarechenzentrum (DKRZ), Hamburg, Germany



# What are we talking about?

Who we are:



- Computing Centre dedicated to the needs of Earth System Science (ESS)
- **HPC-system ranked #62** worldwide in terms of power
- **Storage system ranked #10** worldwide in terms of I/O
- **Sustainable funding**
- **We offer a range of data management services** specifically tailored for ESS



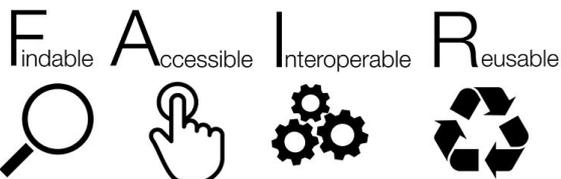
- **CoreTrust Seal certified long-term archive @DKRZ**
  - Domain-specific for ESS
  - FAIR
  - **Focused on long-term data reusability**



- Federated global data infrastructure
- Established for **PB-scale global data dissemination**
- **DKRZ is one of the core partners**



# World Data Center for Climate (LTA WDCC) (1)



## data for the Earth System Sciences

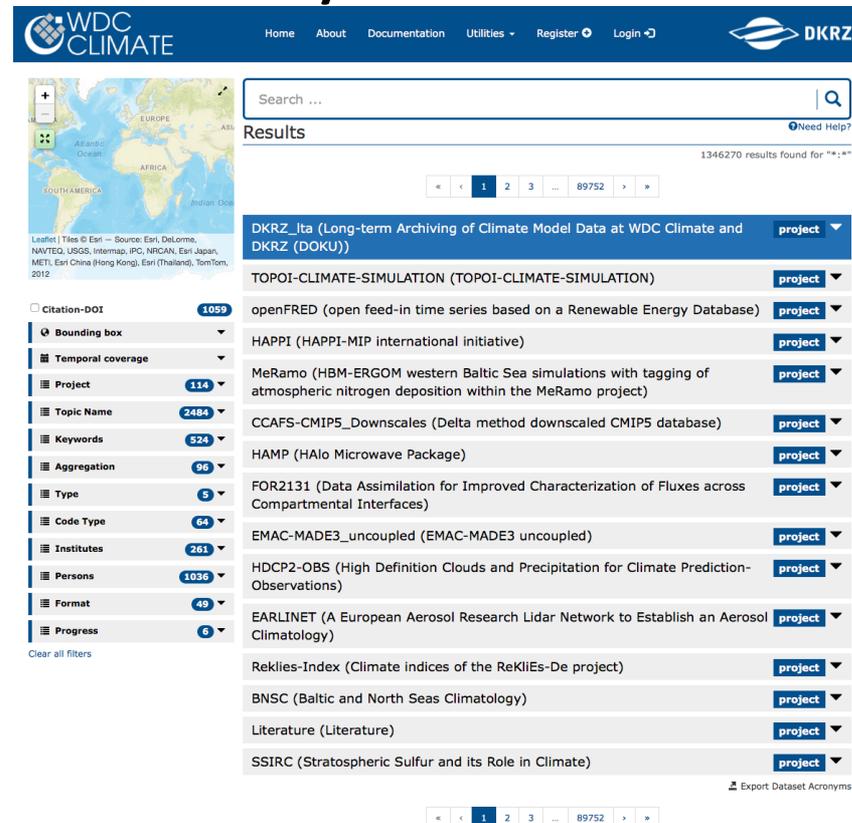
Sungya Pundir, Wikimedia Commons, CC By-SA 4.0

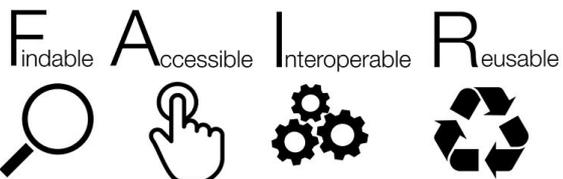
### Findable

- **DOI-assignment**
- **Indexed** in searchable resources, e.g. Google Dataset Search
- **Extensive metadata**

### Accessible

- **Open access** for most datasets
- **Data access free of charge**
- **Metadata remain accessible** in case data is deleted





# data for the Earth System Sciences

Sungya Pundir, Wikimedia Commons, CC By-SA 4.0

## Interoperable

- **domain specific open file formats**, e.g. NetCDF, GRIB, ASCII
- **domain specific conventions** for (meta)data with published vocabularies (CF-conventions)

## Reusable

- **F, A and I** are fulfilled -> **R**
- Metadata contain **information on usability, uncertainties, methods** and links to associated resources

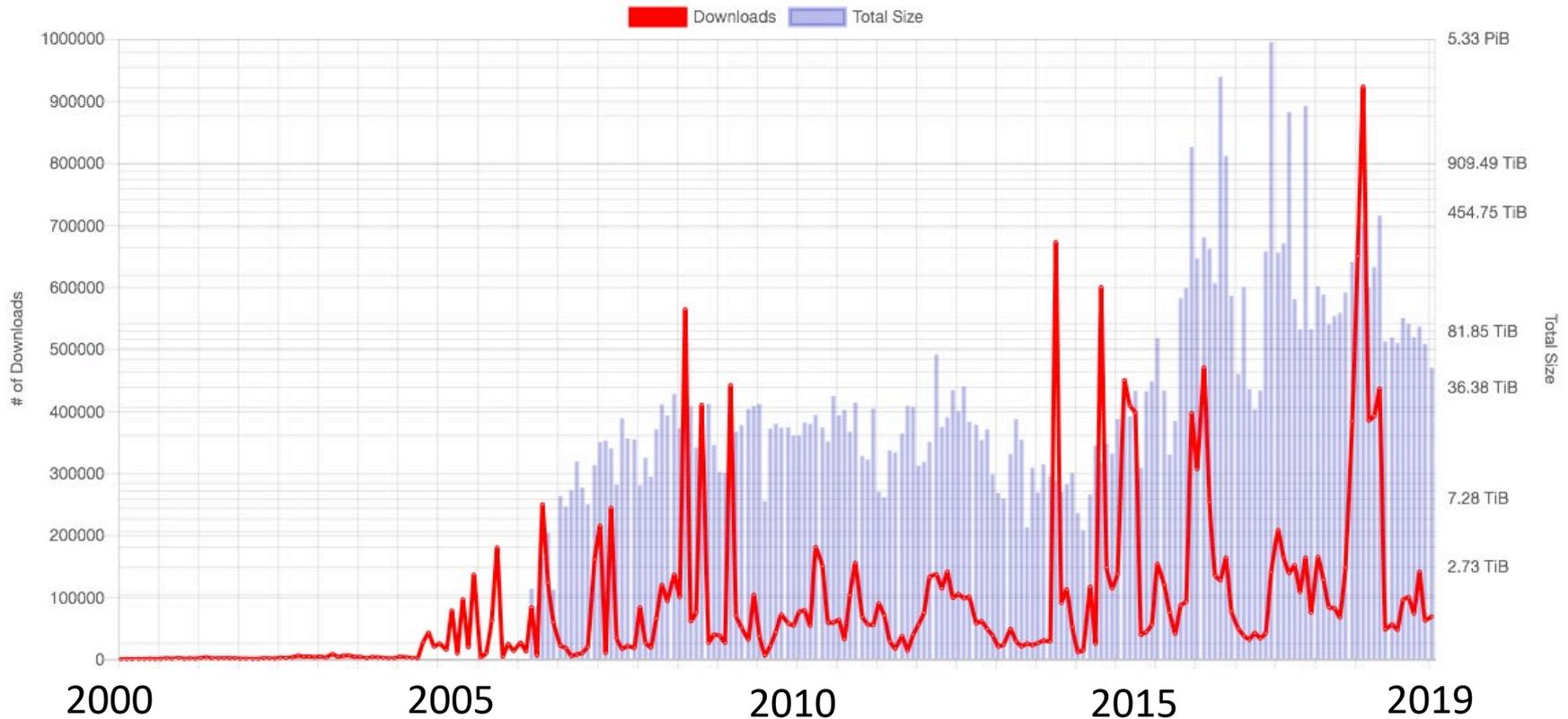
**One-on-one user support** throughout the whole process, resulting in **tailored data preservation** specific for your needs!

**Contact:**  
[data@dkrz.de](mailto:data@dkrz.de)

<https://www.dkrz.de/up/services/data-management/LTA>

# LTA WDCC data (re)use

## Download Statistics

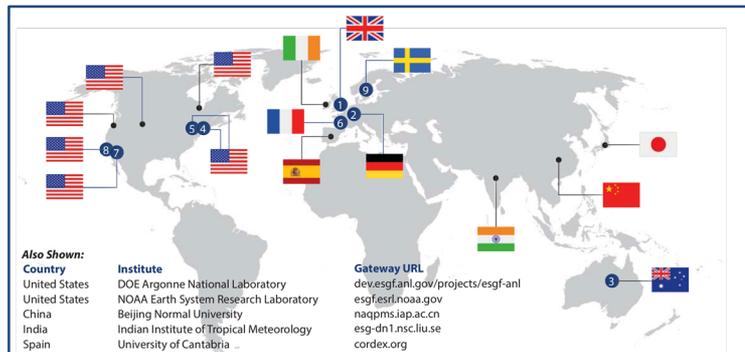


WDCC archived data @DKRZ are being actively re-used (disciplinary and interdisciplinary)

# ESGF: Global Data Dissemination (1)

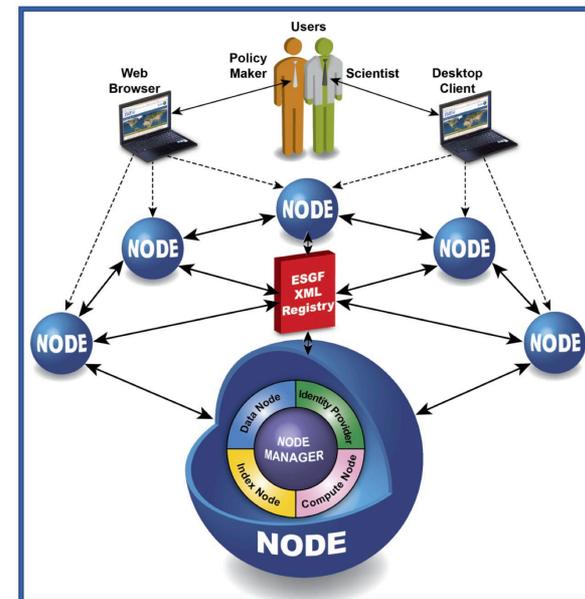
## Earth System Grid Federation (ESGF), established 2006

- infrastructure of **globally distributed data nodes** disseminate **highly standardised** large-volume ESSdata (ca. **3.5 PB for CMIP5**)
- **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**



<https://esgf.llnl.gov>

<https://esgf-data.dkrz.de/>



# ESGF: Global Data Dissemination (2)

**DKRZ publishes community relevant datasets in ESGF – enabling global low-threshold sharing of very large datasets**

<https://www.dkrz.de/up/services/data-management/esgf-services-1>

„I have (lots of) data!“



# ESGF: Global Data Dissemination (3)

## Reusability of ESGF-published data

### Findable

- **PID-allocation** possible
- **Ample metadata**

### Accessible

- **Open access** to all published datasets with user account
- **Download via wget**

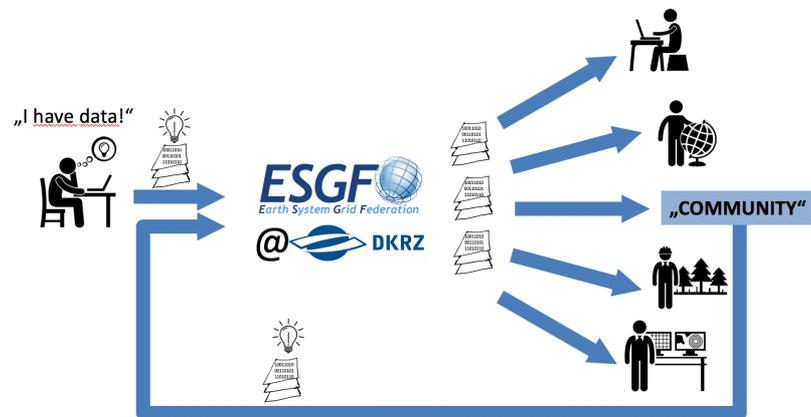
### Interoperable

- **Highly standardised file formats** and (meta)data standards (mandatory!)

### Reusable

- **F, A, I fulfilled** -> Reusable

**AND**



DKRZ-hosted ESGF-data can be **accessed and analyzed** using DKRZ HPC-environment



<https://jupyterhub.dkrz.de>

# Summary

## Long-term archival and global dissemination of Earth System Science / climate data at DKRZ



**Long-term and FAIR preservation** of Earth System Science Research data focused on long-term reusability



**Enabling global dissemination and efficient reuse** of high-impact, large-volume Earth System Science Research data

Contact: [data@dkrz.de](mailto:data@dkrz.de), [esgf@dkrz.de](mailto:esgf@dkrz.de), [peters@dkrz.de](mailto:peters@dkrz.de)