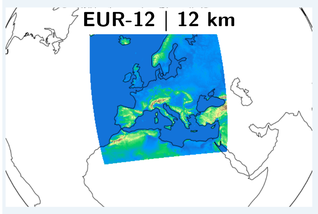
CORDEX-CMIP6 Data at DKRZ

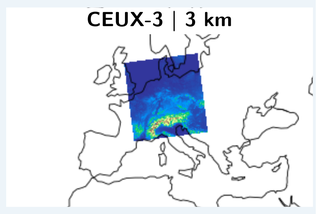
Advanced data infrastructure for discovery, analysis and long-term preservation

Martin Schupfner, F. Wachsmann, I. Anders, C. Kadow, A. Lammert-Stockschläder, S. Kindermann, K. Peters-von Gehlen
DKRZ Deutsches Klimarechenzentrum GmbH

900+ TB data volume | **150+** unique variables | **30+** EUR-12 simulations | **10** CEUX-3 simulations



1950–2150
Simulation coverage
Evaluation: 1950–2024
Historical: 1950–2014
Scenarios: 2015–2100
(subset of simulations extended to 2150)



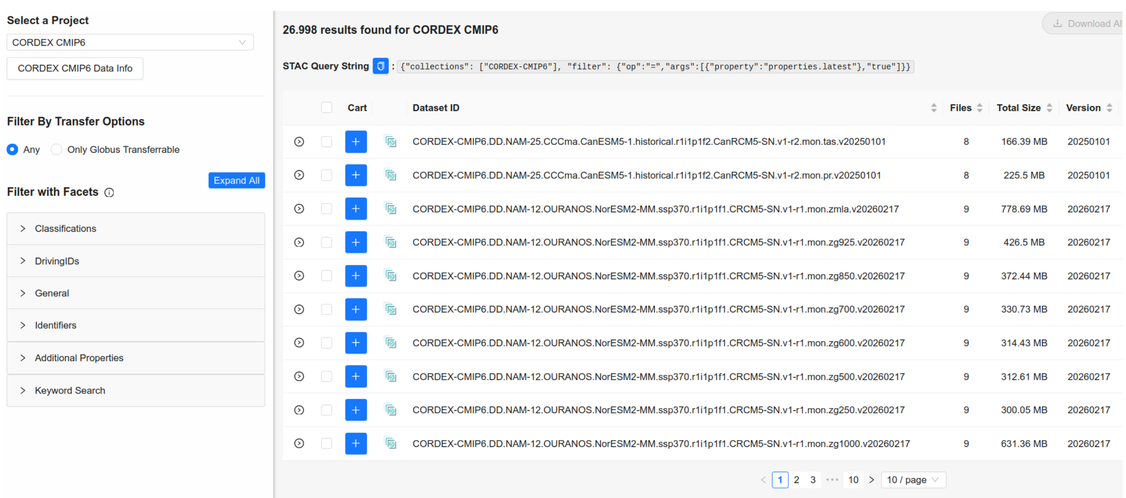
CLMcom / UDAG and partners provide ICON-CLM evaluation, historical and scenario simulations, downscaling ERA5 and CMIP6 for climate adaptation.

ESGF-NG: discover published data

Next-generation ESGF makes distributed climate data discoverable through two STAC indexes, East and West, designed to stay synchronized. In MetaGrid, users can narrow their search using filters such as domain, model, experiment, variable and frequency. The same catalogues support programmatic discovery and access through STAC, while data can be retrieved through HTTP downloads or Globus transfers where available.

ESGF East

ESGF West

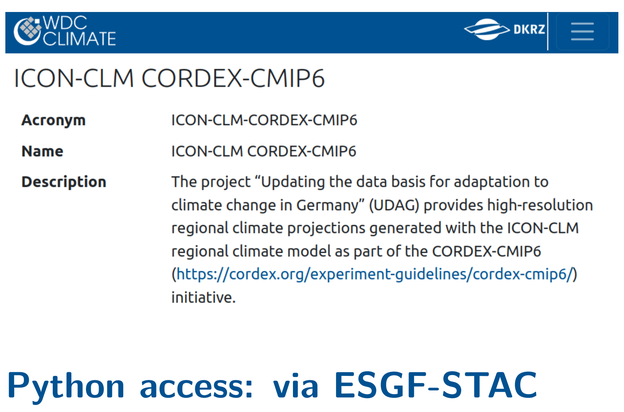


Status 2 September 2026: use East for CORDEX-CMIP6; West is not yet mirroring these records. The faceted search is currently returning inaccurate results which will be addressed.

WDCC: Preserving the data - and the information needed to use them

The World Data Center for Climate (WDCC) at DKRZ preserves climate data and their documentation beyond the project lifetime. Long-term data curation maintains the metadata needed to interpret and reuse the simulations. Preservation for more than ten years and DOI publication after final quality assurance support reproducibility, reliable citation and recognition of the data producers.

Archiving is ongoing in two complementary projects: one covers all ESGF-published data; the other preserves supplementary parameters beyond the CORDEX-CMIP6 Data Request.

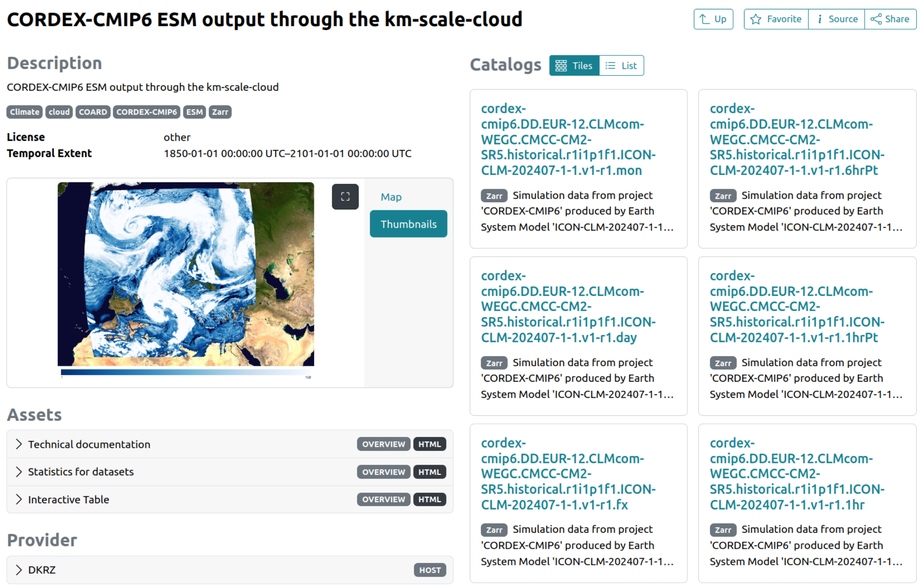


ESGF data

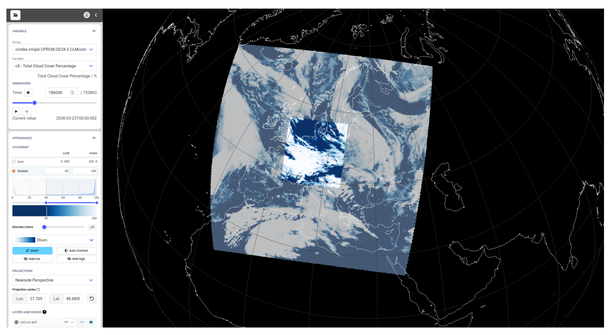
Supplementary data

Cloud: open virtual Zarr datasets

All ESGF-published data are available as virtual Zarr. Variables with the same aggregation and frequency share one Zarr store.



Gridlook: interactive live data viewer



Explore climate data on an interactive globe directly in your browser. Rotate the view, zoom into regions of interest and inspect spatial patterns. Get a quick first look at datasets from the km-scale-cloud and other sources before starting your own analysis. Share a view via a QR code, like the example shown here.

Warning: Using the QR-Code may incur data roaming charges on mobile devices.

Freva: from data discovery to analysis

Freva (Free Evaluation System Framework) connects data discovery with analysis on DKRZ's Levante HPC system. All ESGF-published CORDEX-CMIP6 data on Levante are searchable through its web interface, Python and command line. Find matching files and use them directly in your workflow, close to the compute resources. Integrated analysis plugins and recorded configurations support reproducibility and reuse of analysis workflows.

Python: find a file and open it with xarray
E.g. on Levante: `module load clint freva`. Then use `freva_client` in Python with your Freva instance configured.

```python
from freva_client import databrowser
import xarray as xr

files = databrowser(
    project="CORDEX-CMIP6",
    variable="tas",
    uniq_key="file",
    fail_on_error=True,
)
path = next(iter(files))
ds = xr.open_dataset(path)
```

This opens one matching temperature file. Refine the search to select the simulation and period you need.

Python access: via ESGF-STAC

Open monthly temperature using its full, versioned dataset ID.

```python
import requests
import xarray as xr

dataset_id = (
    "CORDEX-CMIP6.DD.EUR-12.CLMcom-ETH."
    "NorESM2-MM.historical.r1i1p1f1."
    "ICON-CLM-202407-1-3.v1-r1.mon.tas.v20260415")
url = "https://search.east.esgf.io/collections/"
r = requests.get(
    url + "CORDEX-CMIP6/items/" + dataset_id)
r.raise_for_status()
item = r.json()
files = [a["href"] for a in item["assets"].values()
         if a.get("type") == "application/netcdf"]
ds = xr.open_mfdataset(
    files, combine="by_coords", engine="h5netcdf")
```

Full notebook example
Setup, data discovery and a simple analysis example.

Python access: via cloud

Open a monthly data aggregation using its full DRS identifier.

```python
import requests
import xarray as xr
dataset_id = (
    "cordex-cmip6.DD.EUR-12.CLMcom-BTU."
    "CNRM-ESM2-1.ssp585.r1i1p1f2."
    "ICON-CLM-202407-1-1.v1-r1.mon")
url = "https://eerie.cloud.dkrz.de/datasets/"
r = requests.get(url + dataset_id + "/stac")
r.raise_for_status()
a = r.json()["assets"]["eerie-cloud"]
ds = xr.open_dataset(
    a["href"], **a["xarray:open_kwargs"],
    storage_options=a.get("xarray:storage_options"))
```

Full notebook example
Setup, catalogue exploration and a simple analysis example.

Grant/Award No: 01LP2326E

Acknowledgment: The data were produced by CLMcom, the UDAG project and associated partners.

References

Wachsmann, F., Wickramage, C., and the EERIE and DKRZ Data Management team (2025): *Taming the Petabytes: Making km-scale climate data usable.* https://www.dkrz.de/en/projects-and-partners/projects/focus/eerie/

Digital Earths Global Hackathon 2025
https://digital-earths-global-hackathon.github.io/hk25/

DKRZ km-scale-cloud: service and catalogues
https://km-scale-cloud.dkrz.de/