

U can't touch this Climate Data without EERIE's km-scale-cloud



Scan for the full abstract, catalog links, and reproducible access examples.
km-scale-cloud.dkrz.de
digital-earth-hackathon.github.io/hk25

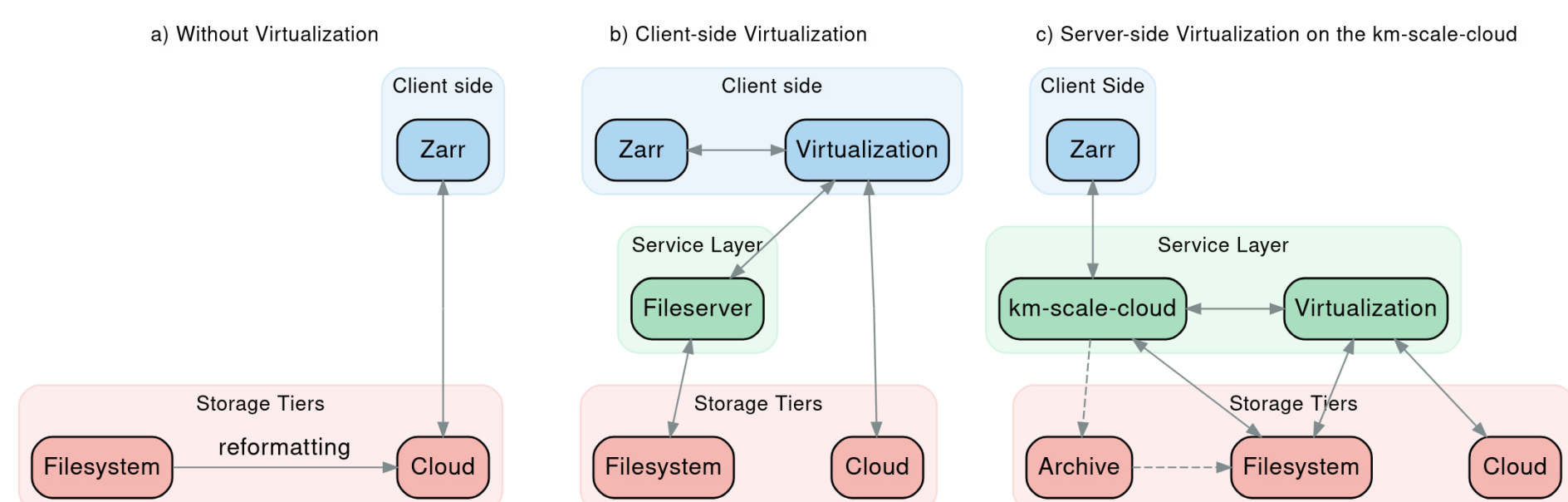


Let's Access PB-Scale ESM Output

Fabian Wachsmann, C. Wickramage, A. Fast, M. Aengenheyster, K. Modali, M. Reis, F. Ziemen, T. Rackow, K. Peters-von Gehlen

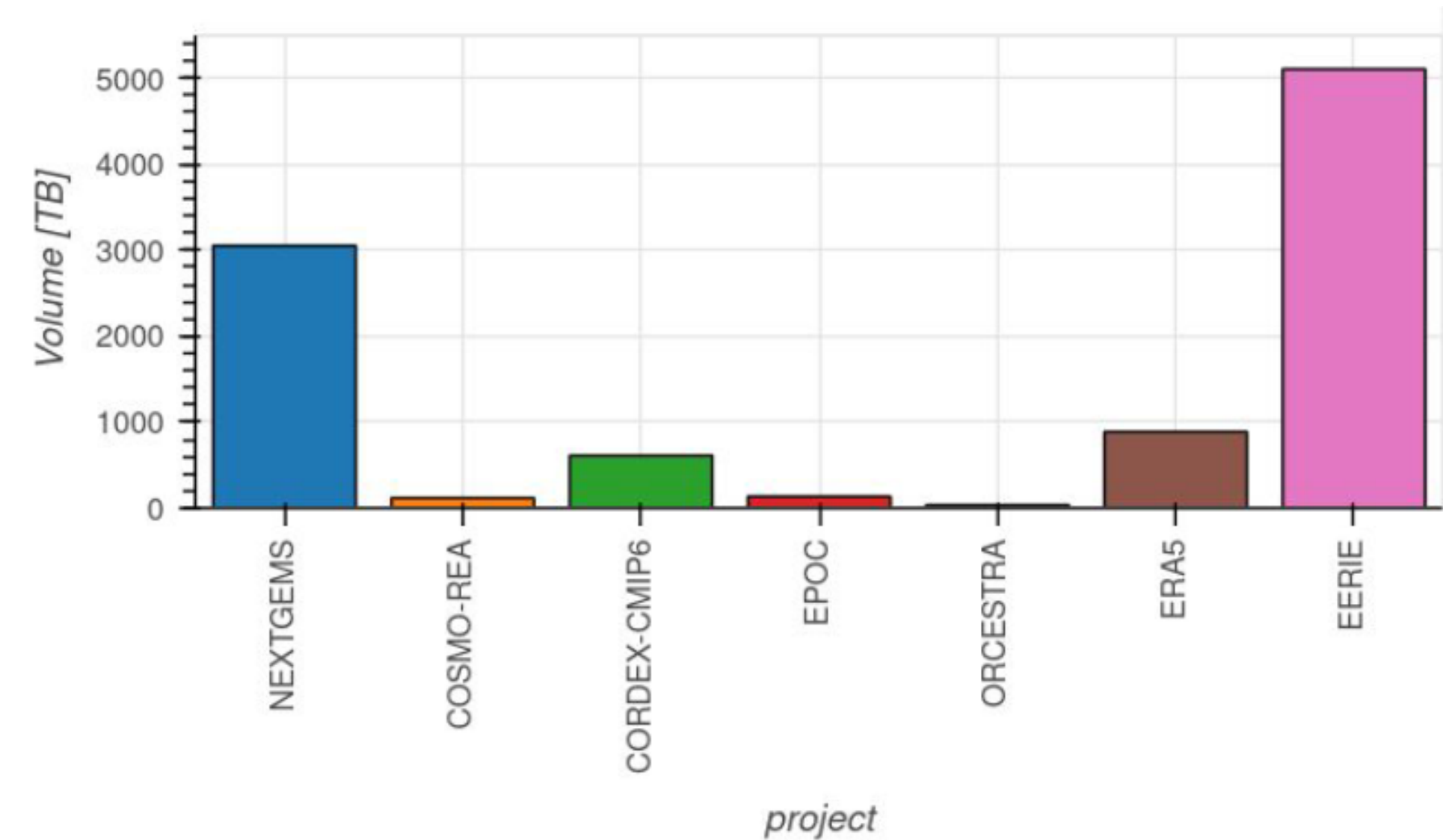
Nice to meet you

km-scale ESM simulations produce petabytes of data across heterogeneous infrastructure. The EERIE km-scale-cloud offers a **single open access layer** supporting reproducible analysis workflows for users across archive, filesystem, and cloud storage environments.



Who I am

- Uniform Zarr-over-HTTP access for Zarr, NetCDF, and GRIB sources.
- About 600 datasets available across seven ESM projects (07/2026).
- Stable transfer rates per client on the order of 100 MB/s.



Dive in a Data Lakehouse

Datasets are *"virtualized"* as Zarr (in the Kerchunk Parquet format): highly aggregated across simulation time, renamed according to CF convention and enriched with metadata.

Example: ERA5-Land daily dataset:

Dimensions: (time: 31412, cell: 542080)

Coordinates:

time	(time)	datetime64[ns]	1940-01-01T11:30:00	
array(['1940-01-01T11:30:00.000000000', '1940-01-02T11:30:00.000000000', '1940-01-03T11:30:00.000000000', ..., '2025-12-29T11:30:00.000000000', '2025-12-30T11:30:00.000000000', '2025-12-31T11:30:00.000000000'], shape=(31412,), dtype='datetime64[ns]')				
lat	(cell)	float64	dask.array<chunksiz...	
lon	(cell)	float64	dask.array<chunksiz...	

Data variables:

bld	(time, cell)	float32	dask.array<chunksiz...
cape	(time, cell)	float32	dask.array<chunksiz...

Open everything with the km-scale Tree

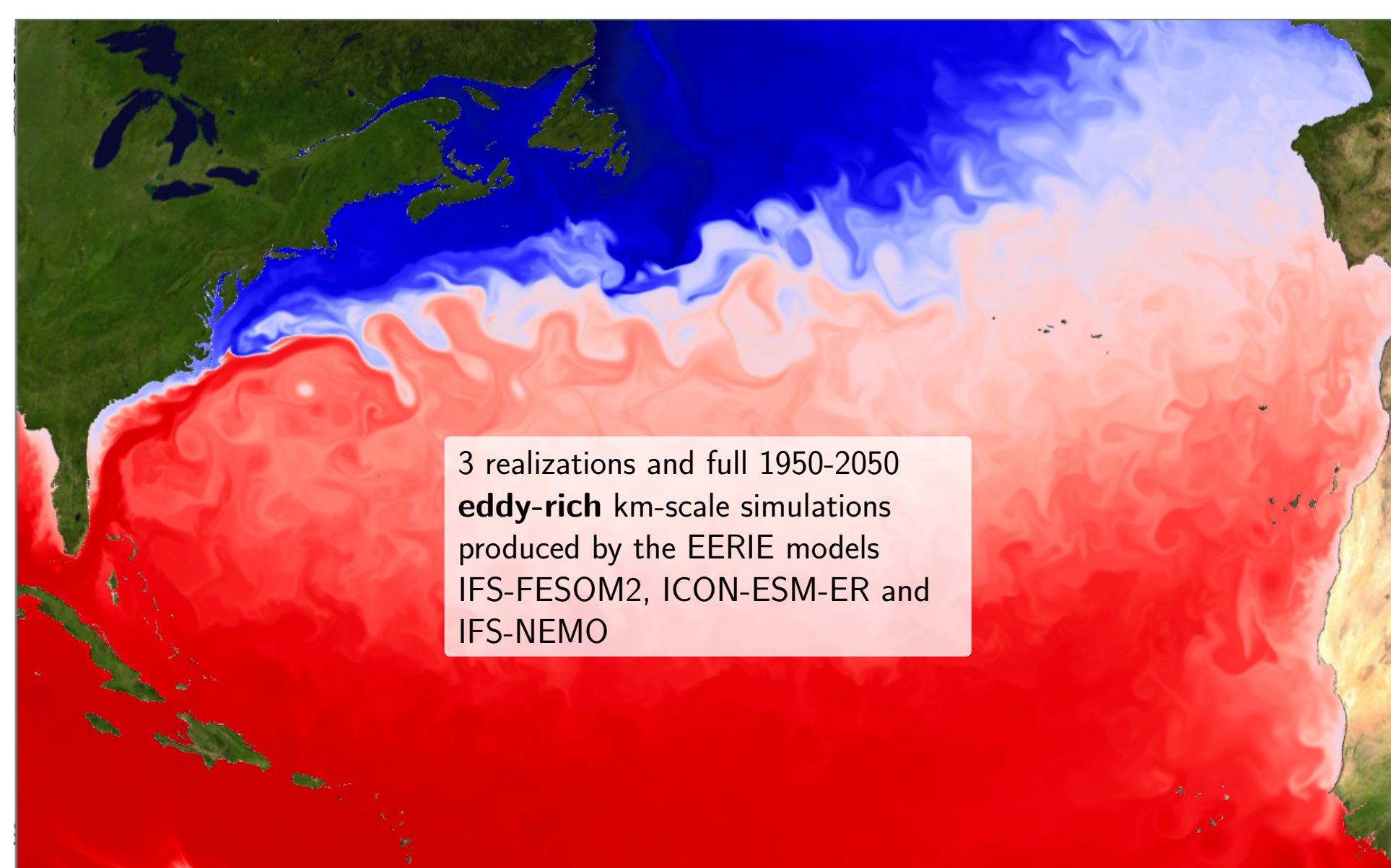
The whole 10PB from the cloud **programmatically accessible** initialized in ms with a memory footprint of <100MB as a DataTree object

```
dt=xr.open_datatree(  
    "https://km-scale-cloud.dkrz.de/datasets",  
    engine="zarr",  
    zarr_format=2,  
    chunks=None,  
    create_default_indexes=False,  
    decode_cf=False  
)
```

Can be filtered similarly to an *Intake* catalog.

```
dt.filter(lambda ds: ds if "s2024-08-10" in ds.path else None)
```

Enjoy the highlights



Some Cookies to take home

km-scale-cloud demonstrates

- Petabytes at your fingertips
- Dual format access with Zarr on top
- Cutting is easier than gluing
- The less endpoints the less problems
- Rapid application development is fueled by open Zarr access

Article: "Taming the Petabytes" on DKRZ page in the spotlight section

Report: "Overview of Cloudify" for the technical background

References

- The km-scale-cloud. <https://km-scale-cloud.dkrz.de>
- STAC: SpatioTemporal Asset Catalog Specification. <https://stacspec.org>
- Xpublish. <https://github.com/xpublish-community/xpublish>
- Gridlook. <https://github.com/d70-t/gridlook>
- VirtualiZarr. <https://github.com/zarr-developers/VirtualiZarr>

Browse Actionable Catalogs

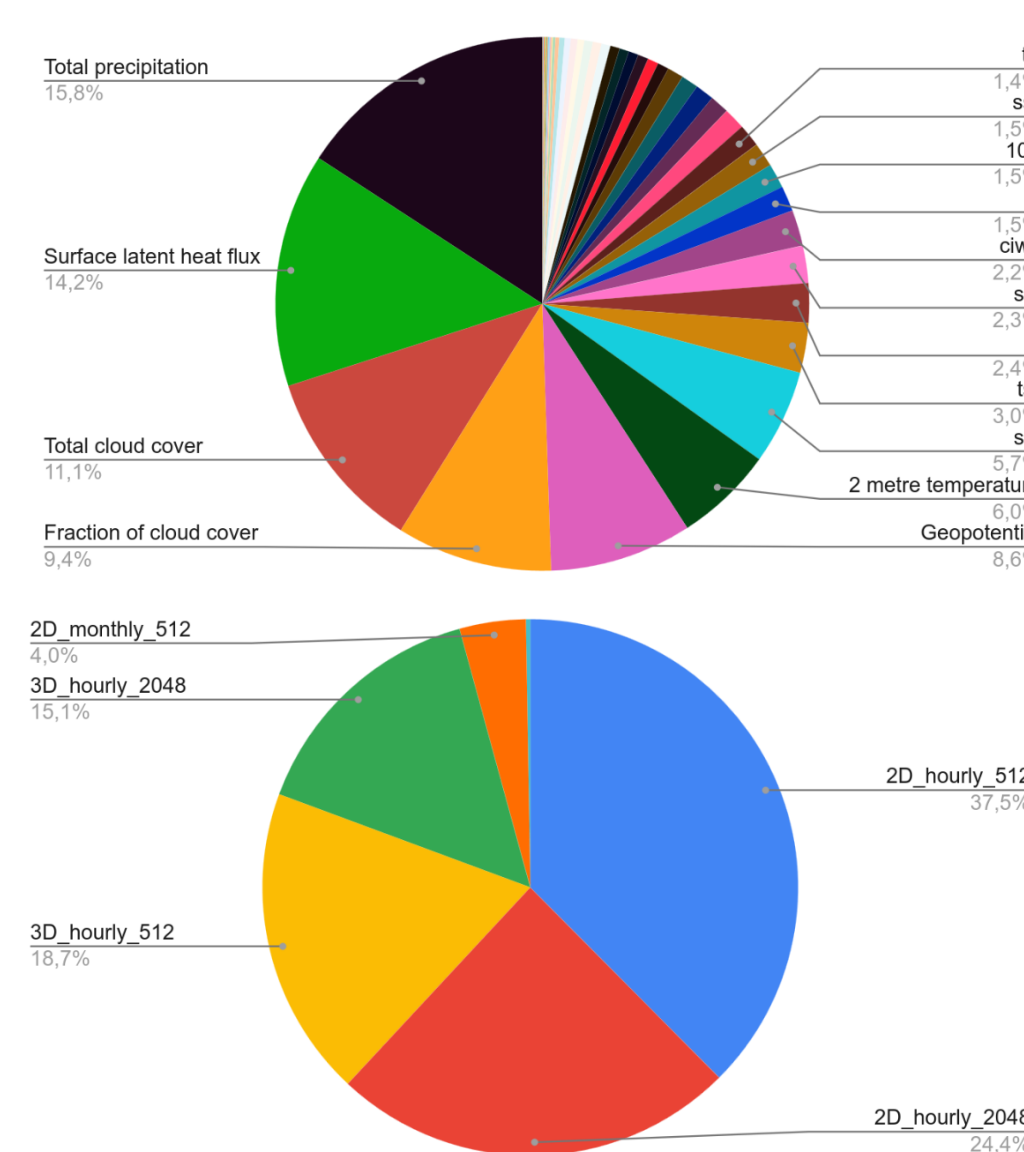
Browser interface for km-scale-cloud at DKRZ. It shows a search bar, a list of catalogs (CORDEX-CMIP6, COSMO-REA, EERIE), and a detailed view of the EERIE catalog with a map and data description.

Embed Interactive Live Views

Interactive live view interface showing a globe with a data overlay and a sidebar with controls for time, location, and data selection. A QR code is also present.

Adopt to Real Access Patterns

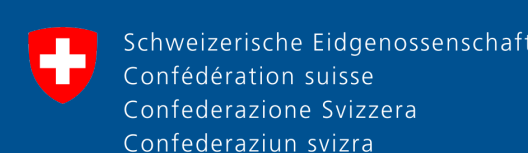
Usage evaluation of an IFS-FESOM2 simulation served for the global WCRP HK25 Hackathon enables chunk shape optimization.



This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101081383.



All UK Partners in EERIE are funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (grant numbers 10057890, 10049639, 10040510, 10040984).



ETH Zurich's contribution to EERIE is funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract #22.00366.