

Rook

Data Reduction Service for the Copernicus Climate
Data Store

Carsten Ehbrecht, DKRZ

Kick-off Meeting C3S2_380, 11 February 2022



Rook

Remote Operations On Klimadaten





(The K is not a typo)

A data reduction service

<https://rook-wps.readthedocs.io/en/latest/>

Climate Data Store

CMIP6 data in CDS is provided using Rook.



Home Search Datasets Applications Toolbox FAQ Live

CMIP6 climate projections

Overview

Download data

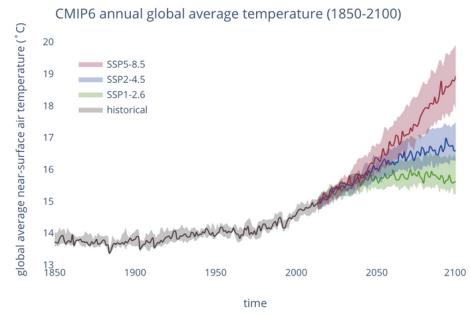
Documentation

This catalogue entry provides daily and monthly global climate projections data from a large number of experiments, models and time periods computed in the framework of the sixth phase of the Coupled Model Intercomparison Project (CMIP6).

CMIP6 data underpins the Intergovernmental Panel on Climate Change 6th Assessment Report. The use of these data is mostly aimed at:

- addressing outstanding scientific questions that arose as part of the IPCC reporting process;
- improving the understanding of the climate system;
- providing estimates of future climate change and related uncertainties;

CMIP6 annual global average temperature (1850-2100)



Climate Data Store - Download data

CMIP6 climate projections

Overview

Download data

Documentation

Temporal resolution

☐ Monthly

☒ Daily

☐ Fixed (no temporal resolution)

Experiment ?

☒ Historical

☐ SSP1-1.9

☐ SSP1-2.6

☐ SSP4-3.4

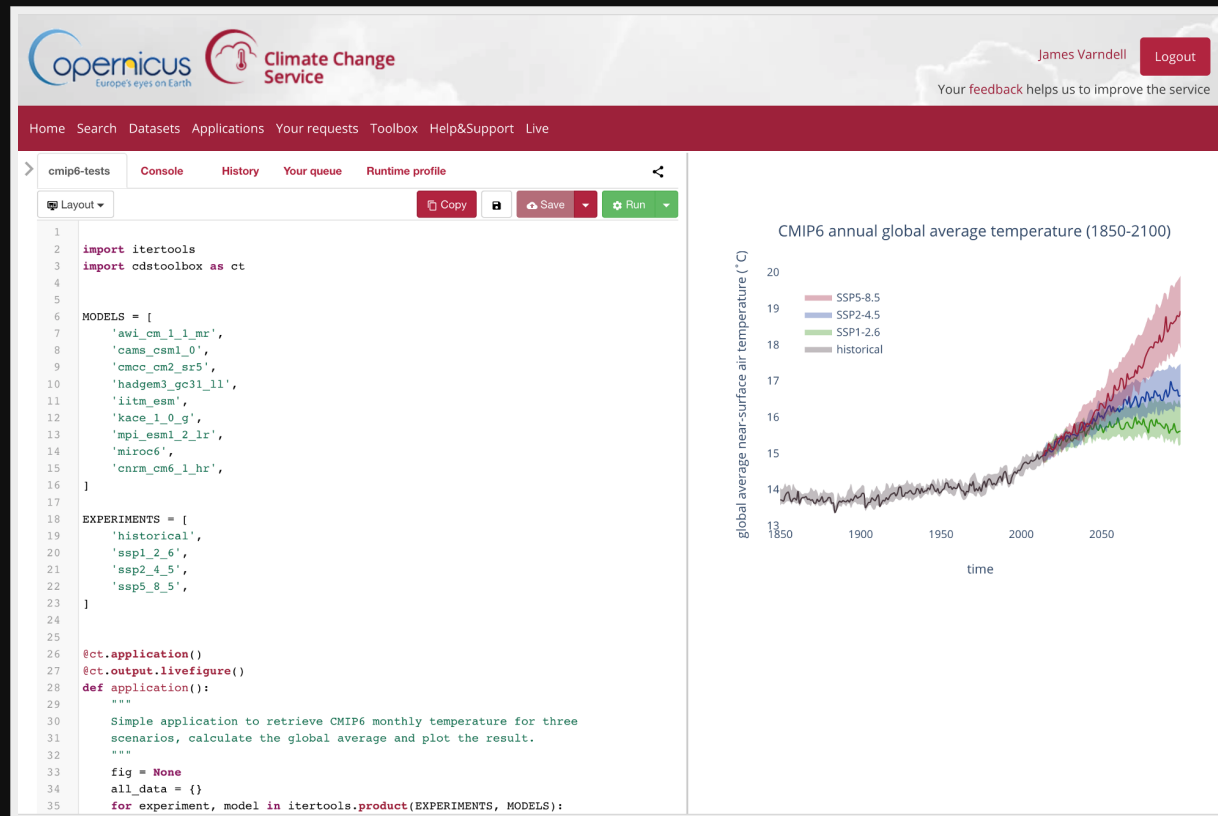
☐ SSP5-3.4OS

☐ SSP4-6.0

☐ SSP3-7.0

☐ SSP5-8.5

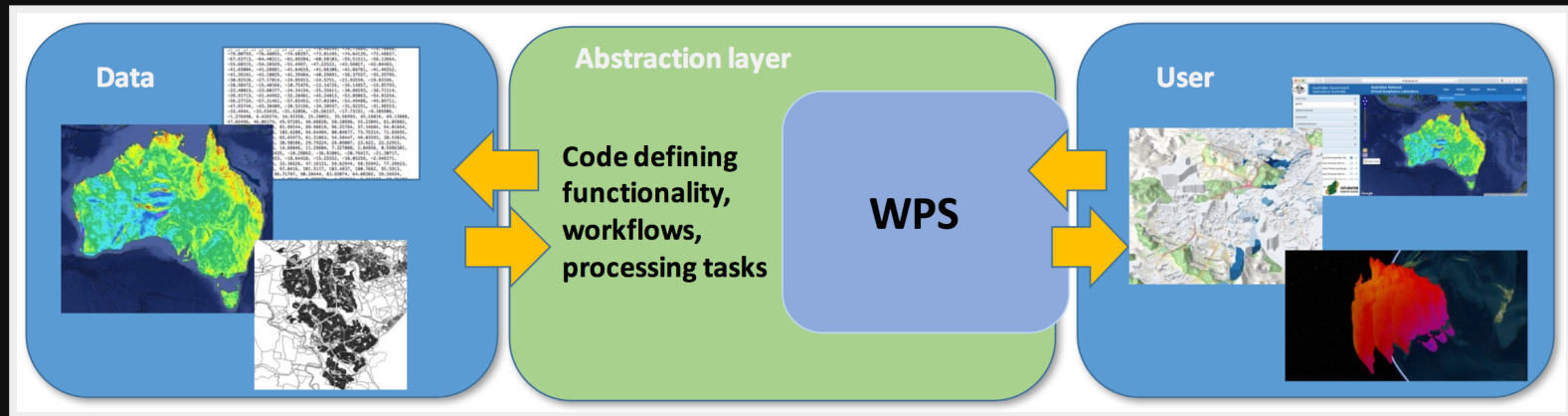
Climate Data Store - Toolbox



Climate Data Store - Rook

- Climate data, CMIP6, is accessed remotely
- Whole files are downloaded via data nodes
- Using Rook: download only a subset of the data
- Example: Temperature, 1990, Africa

Web Processing Service



Call a function remotely

Rook - WPS

- An OGC Web Processing Service
- Using PyWPS - GeoPython
- Providing climate data operators as a service
- Used for data reduction: Temperature, 1990, Africa

Rook - Operators

- Subsetting - time, area, level
- Subsetting - time components:
 - `year:2016,2017|month:jan,feb,mar`
- Averaging - over dimensions (time, ...)
- Regridding (still a pain!)
- ... can be extended

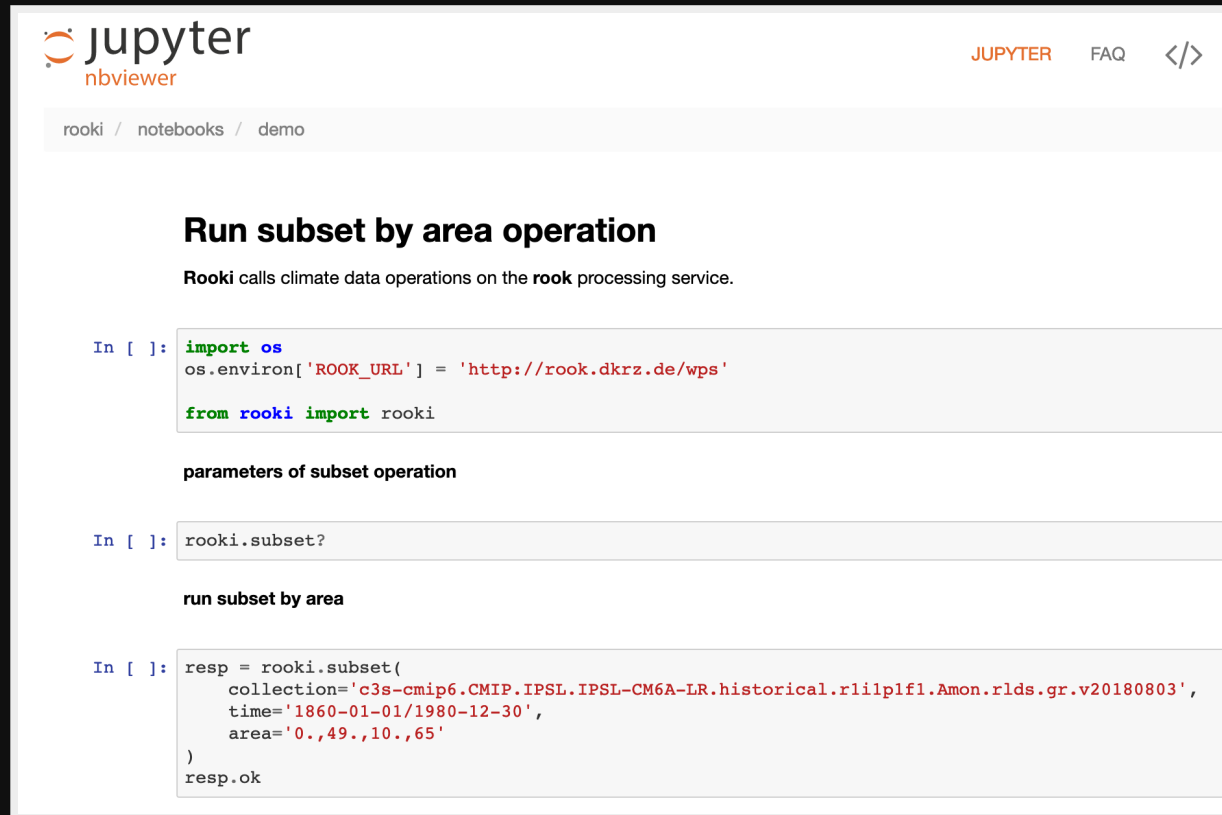
Rook - Clisops

- The Python library implementing these operators
- Using xarray - low level library
- Joint effort together with Ouranos, Canada
<https://clisops.readthedocs.io/en/latest/>

Rooki

- Python WPS client - interactive or as library
- Using OWSLib - GeoPython
- Joint effort with **Ouranos, Canada**
- <https://rooki.readthedocs.io/en/latest/>

Rooki - Notebook



The screenshot shows a Jupyter Notebook interface with the following elements:

- Header:** The Jupyter logo and "nbviewer" text are on the left. "JUPYTER", "FAQ", and a code icon are on the right.
- Breadcrumbs:** "rooki / notebooks / demo" are displayed below the header.
- Section Title:** "Run subset by area operation" is centered in bold.
- Description:** "Rooki calls climate data operations on the rook processing service." is centered below the title.
- Code Cell 1:** Contains Python code to set the environment and import the rooki module.

```
In [ ]: import os
os.environ['ROOK_URL'] = 'http://rook.dkrz.de/wps'

from rooki import rooki
```
- Text:** "parameters of subset operation" is centered below the first code cell.
- Code Cell 2:** Contains a single line of code to check the subset method.

```
In [ ]: rooki.subset?
```
- Text:** "run subset by area" is centered below the second code cell.
- Code Cell 3:** Contains Python code to perform a subset operation with specific parameters.

```
In [ ]: resp = rooki.subset(
    collection='c3s-cmip6.CMIP.IPSL.IPSL-CM6A-LR.historical.r1i1p1f1.Amon.rlds.gr.v20180803',
    time='1860-01-01/1980-12-30',
    area='0.,49.,10.,65'
)
resp.ok
```

Online Notebooks

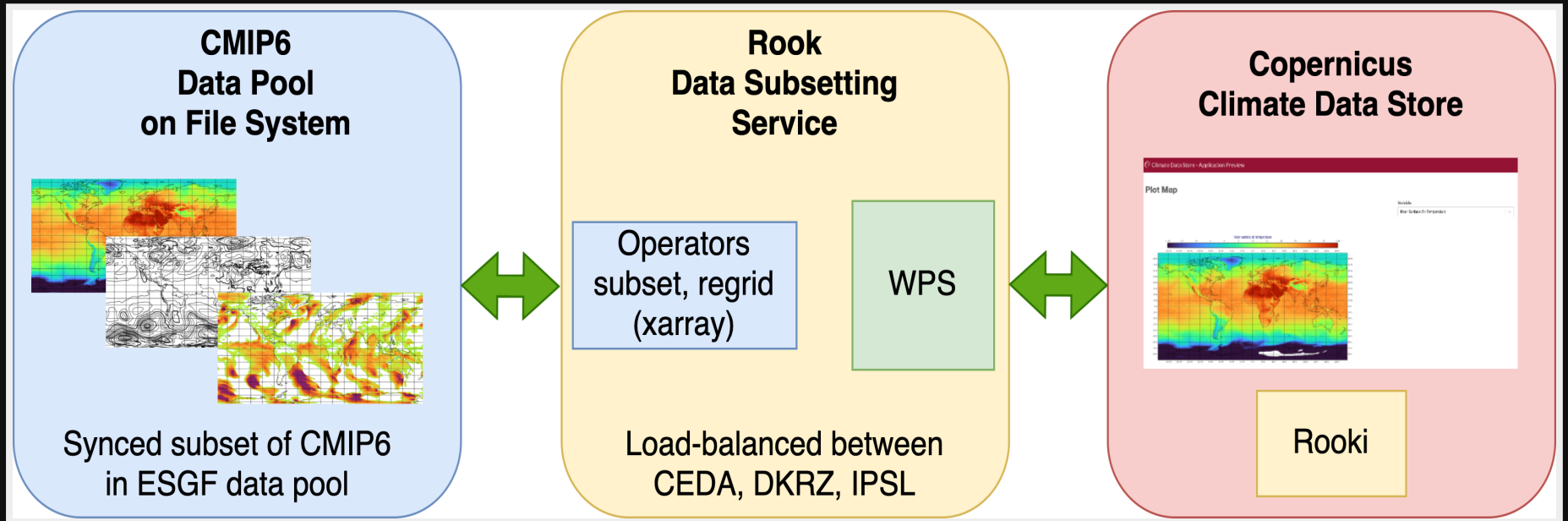
Deployment - Birdhouse Tools

- Rook generated from a Cookiecutter template
- Ansible playbook to roll out on cluster with Slurm scheduler
- Joint effort with **Ouranos, Canada**

Availability

- Data pool is replicated to three sites
- Load-balanced access to Rook WPS and data nodes

All together

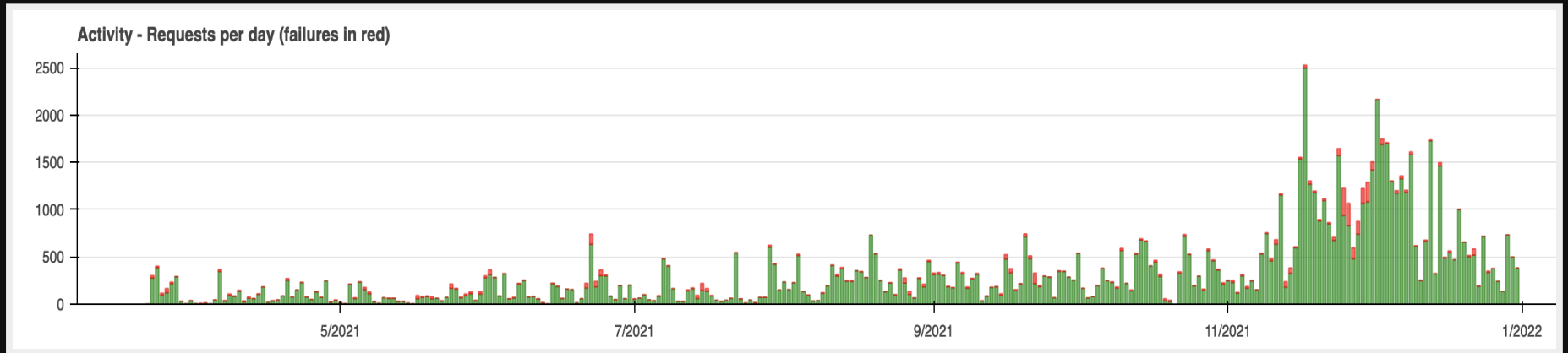


Status - in Production

- Deployed at CEDA, DKRZ and IPSL
- Used for CMIP6
- Subset (time, area, level) operator
- Whole CMIP6 files are downloaded from data nodes

Status - Activity in 2021

200 requests per day ... up to 2500



Online dashboard

Status: Data

- CMIP6: used since March 2021
- CMIP6-decadal: ready for integration with CDS
- CORDEX: ready for integration with CDS
- CMIP5: ready for integration with CDS

Status: Operators

- Subset: used since March 2021
- Subset with time components: available
- Average: ready for integration with CDS
- Regridding: initial version will be available soon

Next in C3S2_380

- Integrate data in CDS
 - CMIP6-decadal, CORDEX, CMIP5
- Updates of existing data pool
- Averaging and Regridding
- New *Concat* operator for CMIP6-decadal
 - `avg(concat(sub(ens1), sub(ens2)))`
- Improve deployment and monitoring
- ???

Issues

- CEDA is not part of C3S2_380
 - CEDA is still online ... backup mode
- With only two sides, the service is less stable
- CEDA is the main driver of the developments

Projects

- Copernicus C3S: <https://climate.copernicus.eu/>
- Roocs: <https://roocs.github.io/>
- Birdhouse: <http://bird-house.github.io/>
- GeoPython: <https://geopython.github.io/>

Thanks

Questions?