Analyzing CMIP6 data with Xarray in Google Cloud

Tim Hermans, Python for lunch 22-01-2026

With thanks to Julius Busecke!

Utrecht University

CMIP6 - Coupled Model Intercomparison Project Phase 6

Overview:

The WCRP Working Group on Coupled Modelling (WGCM) oversees the Coupled Model Intercomparison Project, which is now in its 6th phase. Background information about CMIP and its phases can be found on WGCM website as well as on the PCMDI-hosted pages. An introductory overview of CMIP6 is also provided by the WGCM.

Practical information for those interested in participating in CMIP6 is provided in three guides, tailored to different groups:

1. Modelers carrying out CMIP6 simulations,
2. Data managers responsible for data node operations, and
3. Data users analyzing and making use of CMIP6 model output

Model output Access:

- First see the Data Users Guide
- **Summary table** of currently available data
- The complete archive of CMIP6 output is accessible from any one of the following portals:
  - USA, PCMDI/LLNL (California) - https://esgf-node.llnl.gov/projects/cmip6/
  - France, IPSL - https://esgf-node.ipsl.upmc.fr/projects/cmip6-ipsl/
  - Germany, DKRZ - https://esgf-data.dkrz.de/projects/cmip6-dkrz/
  - UK, CEDA - https://esgf-index1.ceda.ac.uk/projects/cmip6-ceda/

CMIP6 Endorsed MIPs:

- WCRP Endorsed (Model Intercomparison Project) MIPs overview page
- CMIP6 Ocean Model Intercomparison Project (OMIP) overview page

Additional information for CMIP6:

- CMIP6 license and terms of use

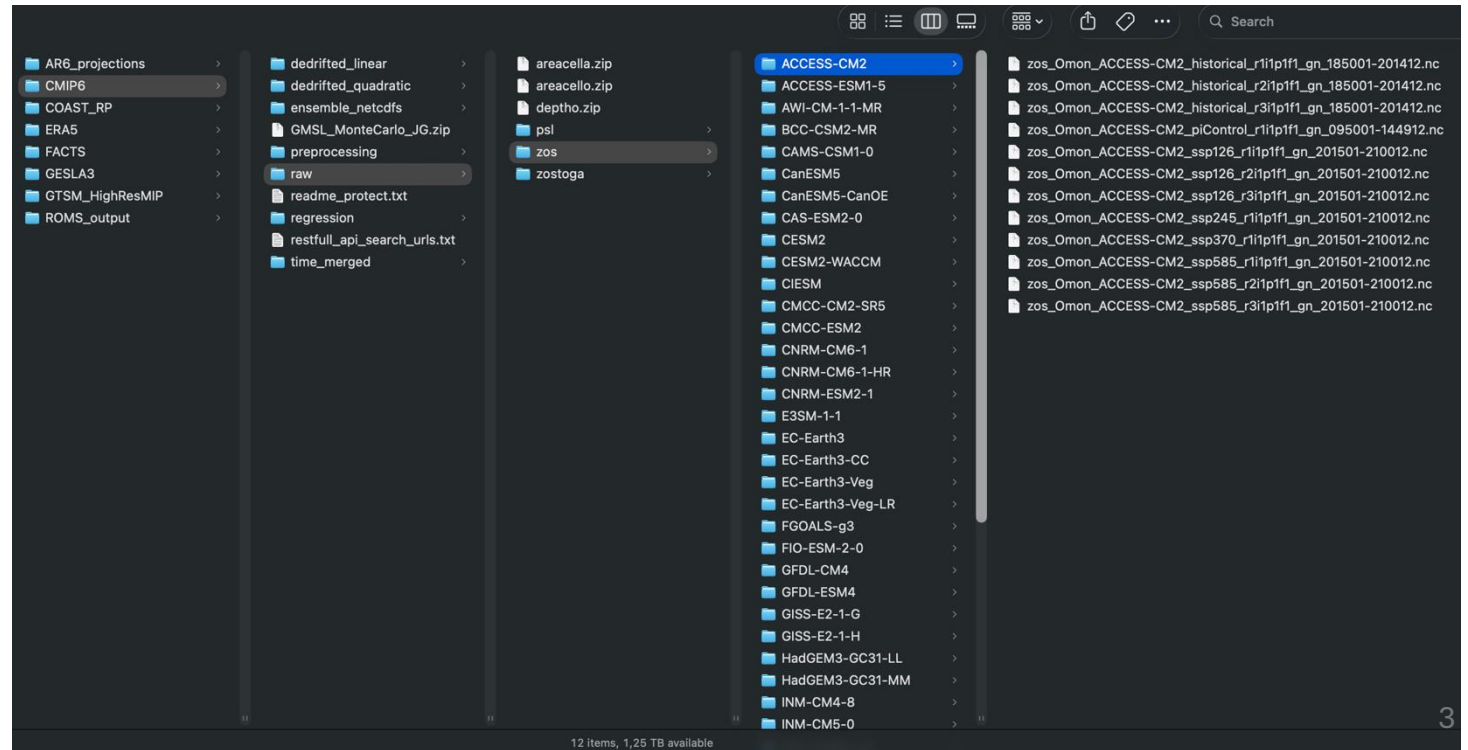## CMIP6 Modeling Groups (click on flags to reveal identity)

https://pcmdi.llnl.gov/CMIP6/
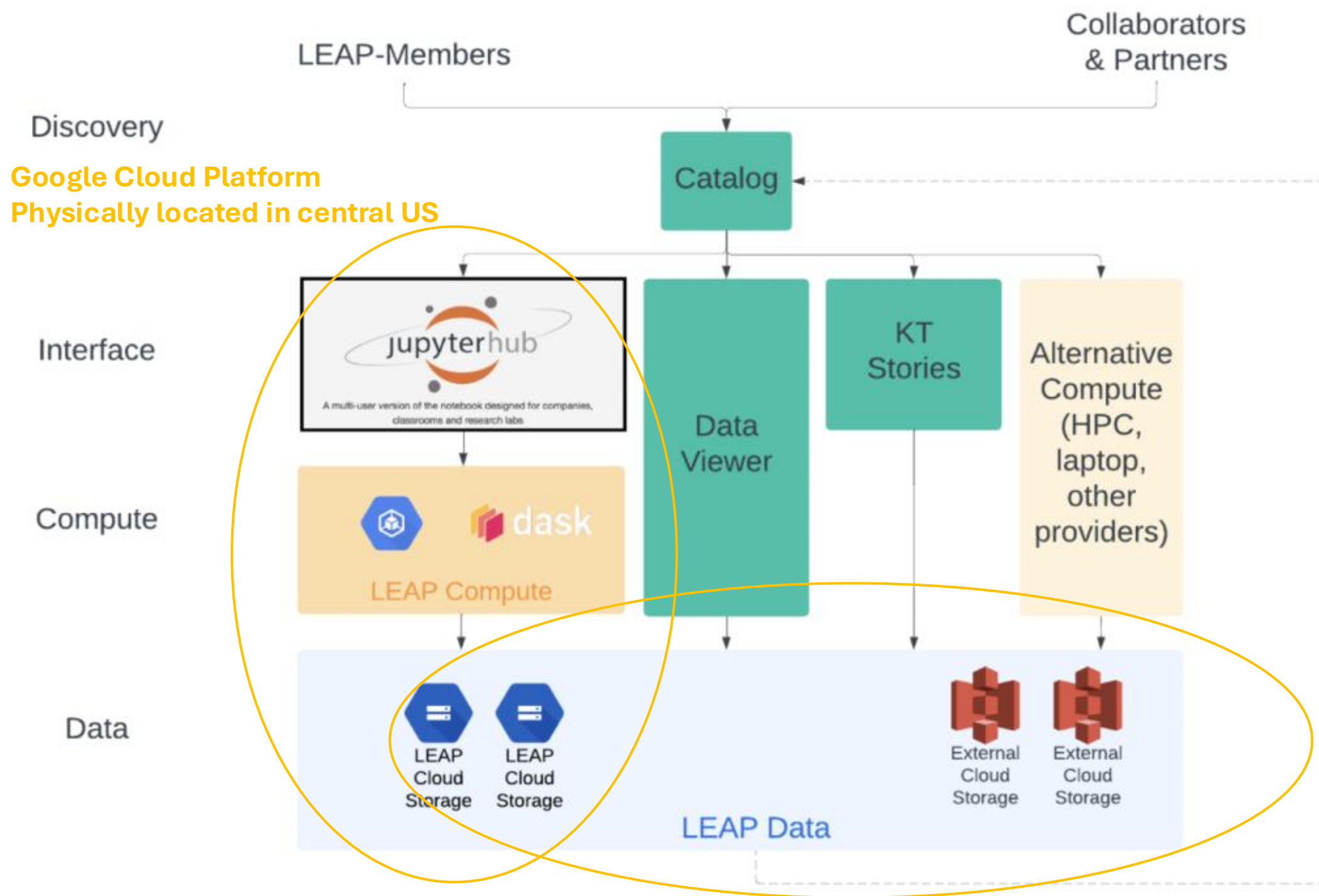
Typical workflow

- Download locally or on HPC via https (OPeNDAP urls) or generate wget script (via search URL:
  - https://esgf-data.dkrz.de/esg-search/wget?query=CMIP6&variable=zos&frequency=mon&variant_label=r1i1p1f1&variant_label=r1i1p1f2&variant_label=r1i1p1f3&experiment_id=piControl&experiment_id=historical&experiment_id=ssp126&experiment_id=ssp245&experiment_id=ssp585&download_structure=variable,source_id&limit=10000

- Analyze from there

Alternative: use cloud-based data

- Google Public Datasets hosts CMIP6 data: https://cloud.google.com/blog/products/data-analytics/new-climate-model-data-now-google-public-datasets
- Advantages:
  - No need to store data locally
  - Fast, & really fast if combined with data-proximate cloud-based compute
  - Less involved, more transparent, easier to reproduce
  - Scale up with dask (package for parallel, distributed computing)

Example: what can you do with this?

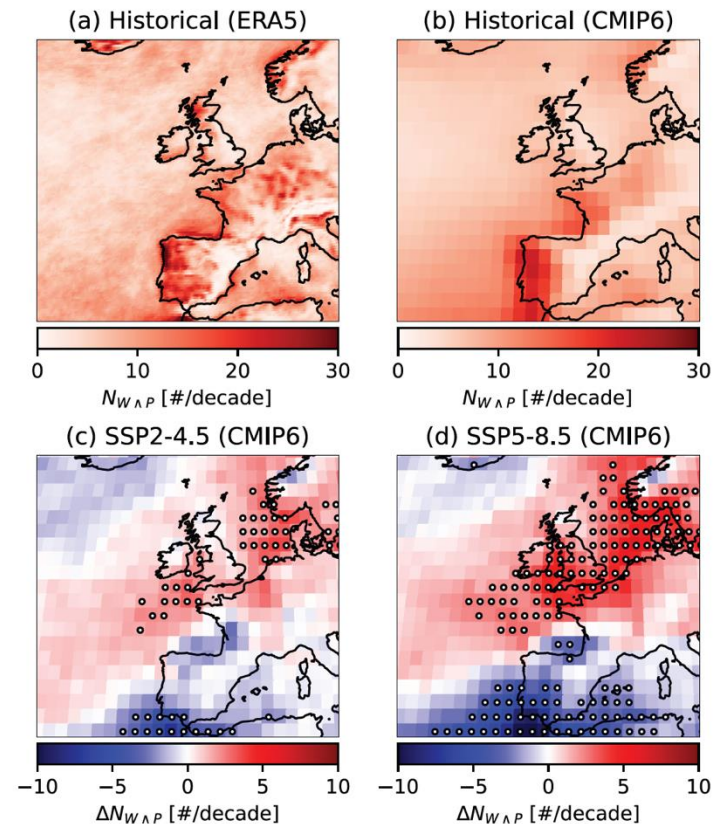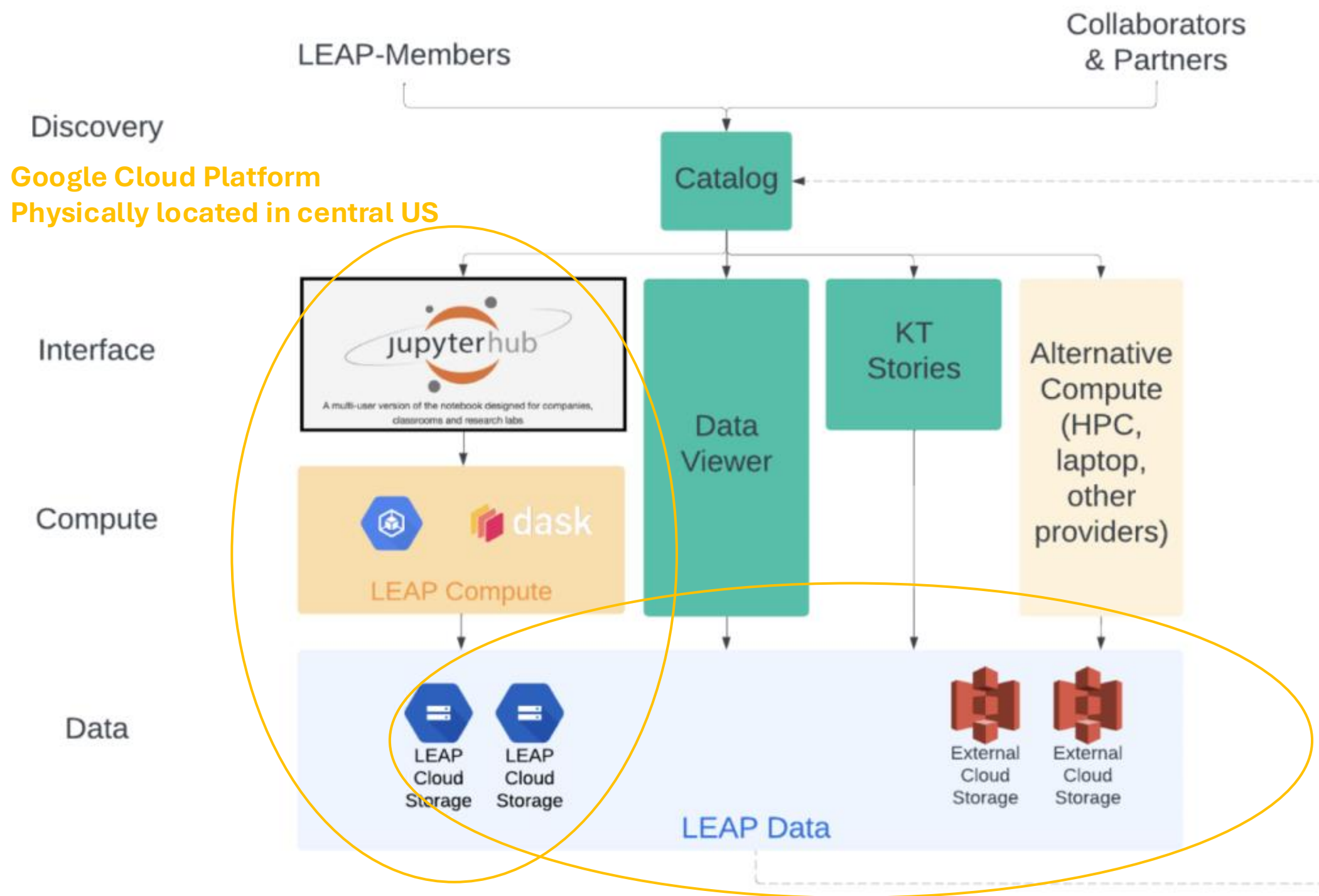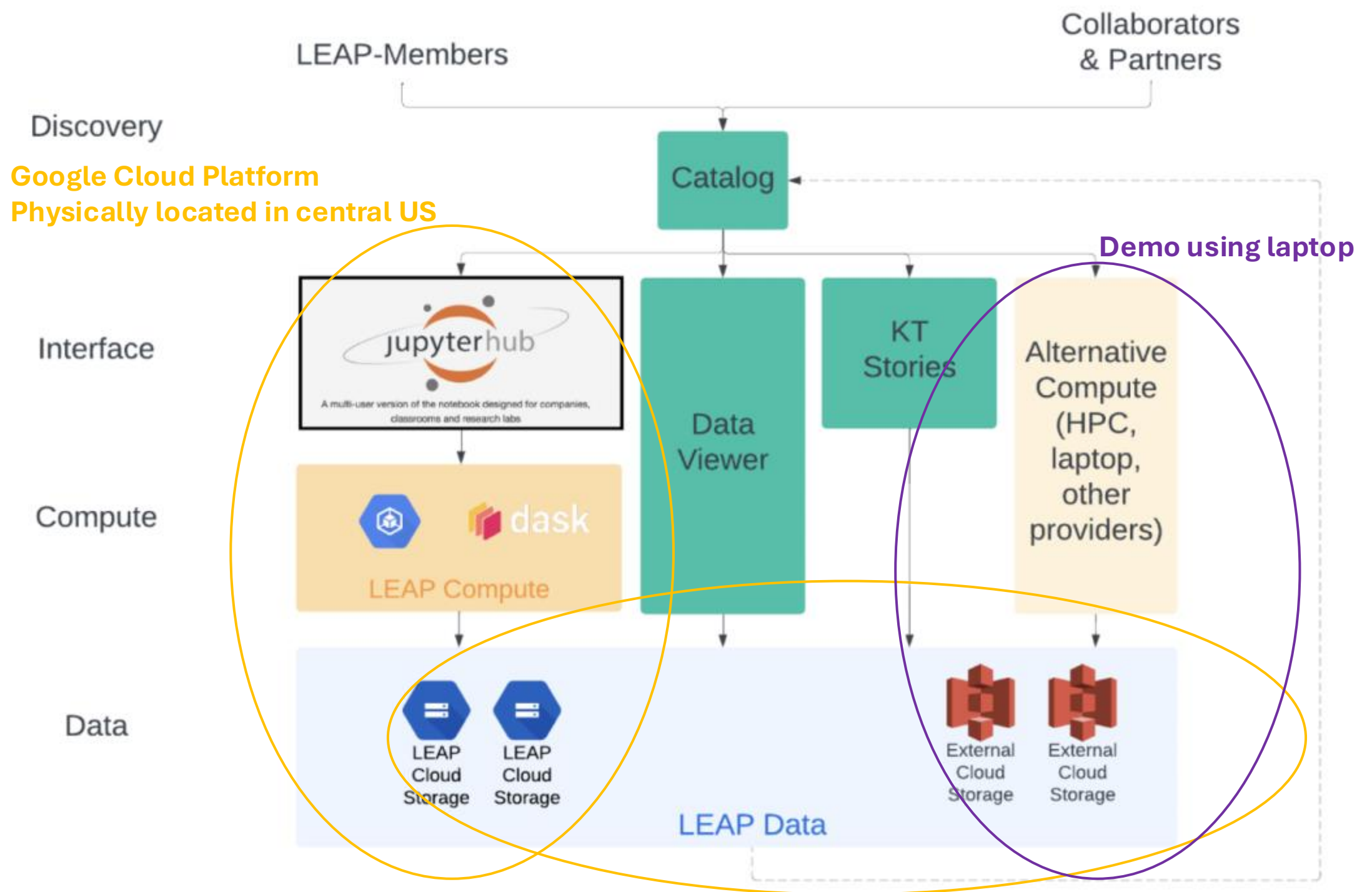- Analysis of daily precipitation, wind and sea-level pressure simulations (20 TB+ of CMIP6 data):
  - https://github.com/Timh37/CMIP6cex

(a) Historical (ERA5)   (b) Historical (CMIP6)

(c) SSP2-4.5 (CMIP6)   (d) SSP5-8.5 (CMIP6)

Table 1
Coupled Model Intercomparison Project 6 Simulations Used

	Model	SSP2-4.5 [#]	SSP5-8.5 [#]	Both [#]	°Lon × °Lat	Reference
1	ACCESS-CM2	5	6	4	1.875 × 1.25	Bi et al. (2020)
2	ACCESS-ESM1-5	38	35	33	1.875 × 1.25	Bi et al. (2020)
3	CanESM5	25	25	25	2.8 × 2.8	Swart et al. (2019)
4	CESM2	2	2	2	1.25 × 0.9	Danabasoglu et al. (2020)
5	CESM2-WACCM	3	3	3	1.25 × 0.9	Danabasoglu et al. (2020)
6	CMCC-ESM2	1	1	1	1.25 × 0.9	Lovato et al. (2022)
7	CMCC-CM2-SR5	1	1	1	1.25 × 0.9	Cherchi et al. (2019)
8	EC-Earth3	59	1	1	0.75 × 0.75	Döscher et al. (2022)
9	EC-Earth3-Veg	1	0	0	0.75 × 0.75	Döscher et al. (2022)
10	FGOALS-g3	1	0	0	2 × 2	L. Li et al. (2020)
11	GFDL-CM4	1	1	1	1 × 1	Held et al. (2019)
12	GFDL-ESM4	1	1	1	1 × 1	Dunne et al. (2020)
13	HadGEM3-GC31-LL	5	4	4	1.875 × 1.25	Andrews et al. (2020)
14	HadGEM3-GC31-MM	0	4	0	0.83 × 0.56	Andrews et al. (2020)
15	INM-CM4-8	1	1	1	2 × 1.5	Volodin and Gritsun (2018)
16	INM-CM5-8	1	1	1	2 × 1.5	Volodin et al. (2017)
17	IPSL-CM6A-LR	11	7	6	2.5 × 1.3	Boucher et al. (2020)
18	KACE-1-0-G	3	3	3	Not reported	Lee et al. (2020)
19	MIROC6	43	50	43	1.4 × 1.4	Tatebe et al. (2019)
20	MIROC6-ES2L	10	1	1	2.8 × 2.8	Hajima et al. (2020)
21	MPI-ESM1-2-LR	24	24	24	1.88 × 1.88	Mauritsen et al. (2019)
22	MPI-ESM1-2-HR	2	2	2	0.93 × 0.93	Mauritsen et al. (2019)
23	MRI-ESM2-0	1	1	1	0.75 × 0.75	Yukimoto et al. (2019)
24	NorESM2-LL	3	1	1	2.5 × 1.88	Seland et al. (2020)
25	NorESM2-MM	2	1	1	1.25 × 0.94	Seland et al. (2020)
26	TaiESM1	1	1	1	1.25 × 0.9	Wang et al. (2021)
27	UKESM1-0-LL	5	5	5	1.875 × 1.25	Sellar et al. (2020)

LEAP-Members

Collaborators & Partners

Discovery

Google Cloud Platform
Physically located in central US

Catalog

Demo using laptop

Interface

jupyterhub

A multi-user version of the notebook designed for companies, classrooms and research labs

Data Viewer

KT Stories

Alternative Compute (HPC, laptop, other providers)

Compute

dask

LEAP Compute

Data

LEAP Cloud Storage

LEAP Cloud Storage

External Cloud Storage

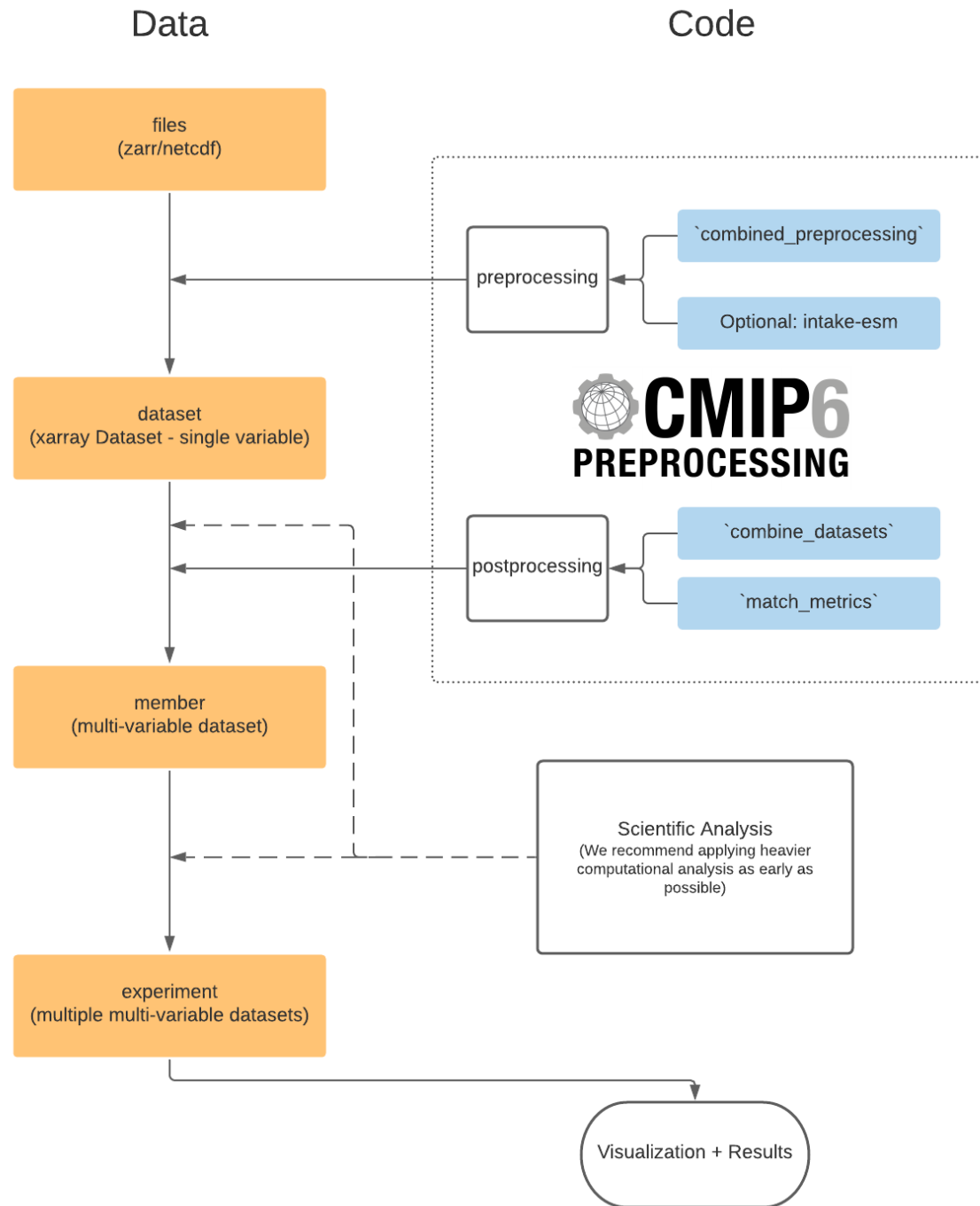External Cloud Storage

LEAP Data

Demo notebook (inspired by pangeo gallery)

- Follow the steps on:
  https://github.com/Timh37/python4lunch_cmip6gc

Getting started

- xMIP documentation: https://cmip6-preprocessing.readthedocs.io/en/latest/?badge=latest
  - https://cmip6-preprocessing.readthedocs.io/en/latest/tutorial.html
- Some basic and more advanced demo notebooks: https://gallery.pangeo.io/repos/pangeo-gallery/cmip6/
- More complex repository using cloud-based CMIP6 data: https://github.com/Timh37/aslc_cmip6/tree/main/preprocessing
- About LEAP Pangeo: https://leap-stc.github.io/introduction/what_is_the_hub/