is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Final General Assembly

16-18 January 2023
Toulouse

Development of Analytics Services

Stephan Kindermann (DKRZ), Alessandro Spinuso (KNMI), Carsten Ehbrecht (DKRZ), Paola Nassisi (CMCC)

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Overview

- ENES CDI Analytics Services
- Analytics service activity (SA, TNA)
- Web processing service
- Future Perspectives

Processing / Analytics Services

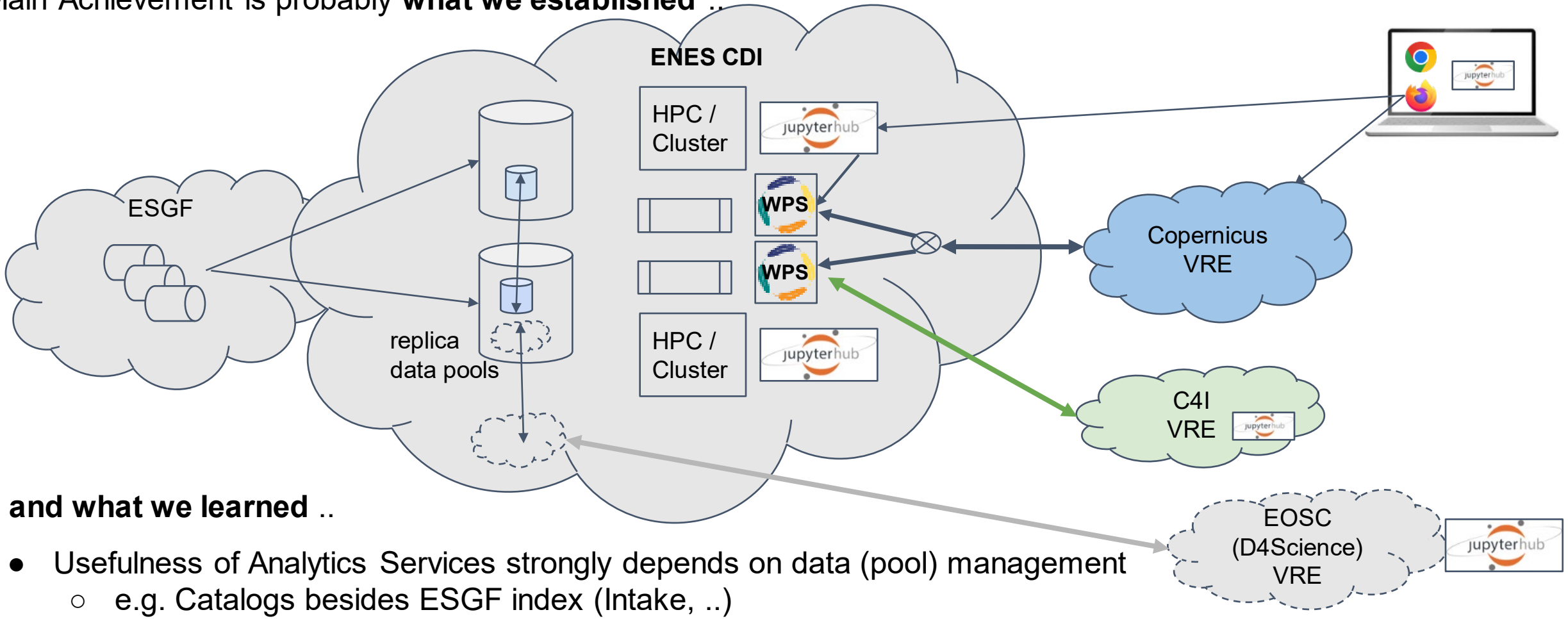
Two complementary analytics service development efforts:

A) Harmonize access to institutional processing capabilities with focus on (interactive) Jupyter-Hub based data analytics
- service provisioning (SA and TNA)
- IPCC WG support
- aligned with broader community efforts (PANGEO SW stack, Intake data catalogs etc.)
- aligned with EOSC efforts (community VRE support, EOSC-pillar example)

A) Standards (OGC) based Processing Web Service provisioning
- integrating data access and data pre-processing behind one interface
- used by C4I portal
    - used in production for Copernicus data provisioning
- based on OGC/GeoPython and bird-house, close collaboration with Canada
- proposed solution for an ESGF compute service (ESGF CWT)

Processing / Analytics Services

Main Achievement is probably **what we established** ..



.. **and what we learned** ..

- Usefulness of Analytics Services strongly depends on data (pool) management
  - e.g. Catalogs besides ESGF index (Intake, ..)
- Interfaces to Analytics services (WPS) are important VRE integration points
  - e.g. one interface for data download and (pre-)processing (WPS)
- Cloud (and analysis ready data) based Analytics Services becoming more important

Interactive Analytics Services: Jupyterhub

Main Achievements (IS-ENES3)

- Provisioning of Jupyter-hub instances at ENES-CDI ESGF tier1 centers (DKRZ, IPSL and STFC) as well as CMCC with access to replica data pools
  - Intake catalog support to exploit data pools at some sites

- Demonstrated usefulness of consistent community SW stack setup at different sites (at DKRZ and CMCC to support summer school with a failsafe environment)
  - Pangeo related Amazon/Google Cloud CMIP6 community notebooks are supported by ENES CDI analytics service
  - Specific support for ESMValTool based analytics (Jupyterhub kernel)

- Demonstrated integration with external (EOSC based) Jupyterhub based VRE (EOSC-pillar project)
- (Usefulness of data pools with associated interactive analytics services at DKRZ and STFC acknowledged in CMIP6 survey)

Analytics Services: Service Aspects (SA and TNA)

Main Achivements:

- Supported 15 group applications during IS-ENES3 (TNA)
- > 4000 registered users for VA compute service
- Service Activities helped to harmonize the Analytics Environments
  - Coordinated Training Material, Coordinated compute environments (xarray,..)
  - Shared approach for local data catalogs: Intake (ESGF catalog not so usefull ..)
- Lessons learned:
  - TNA:
    - relatively low CPU resource requirements
    - relatively small research groups (mostly individuals, exception: Primavera support)
    - some requirements for longer term support e.g. SMEs (Geoskop)
  - VA:
    - short term analytics / on demand requirements → lightweight quick application procedure
    - jupyter hub + data catalog (intake) for local data useful

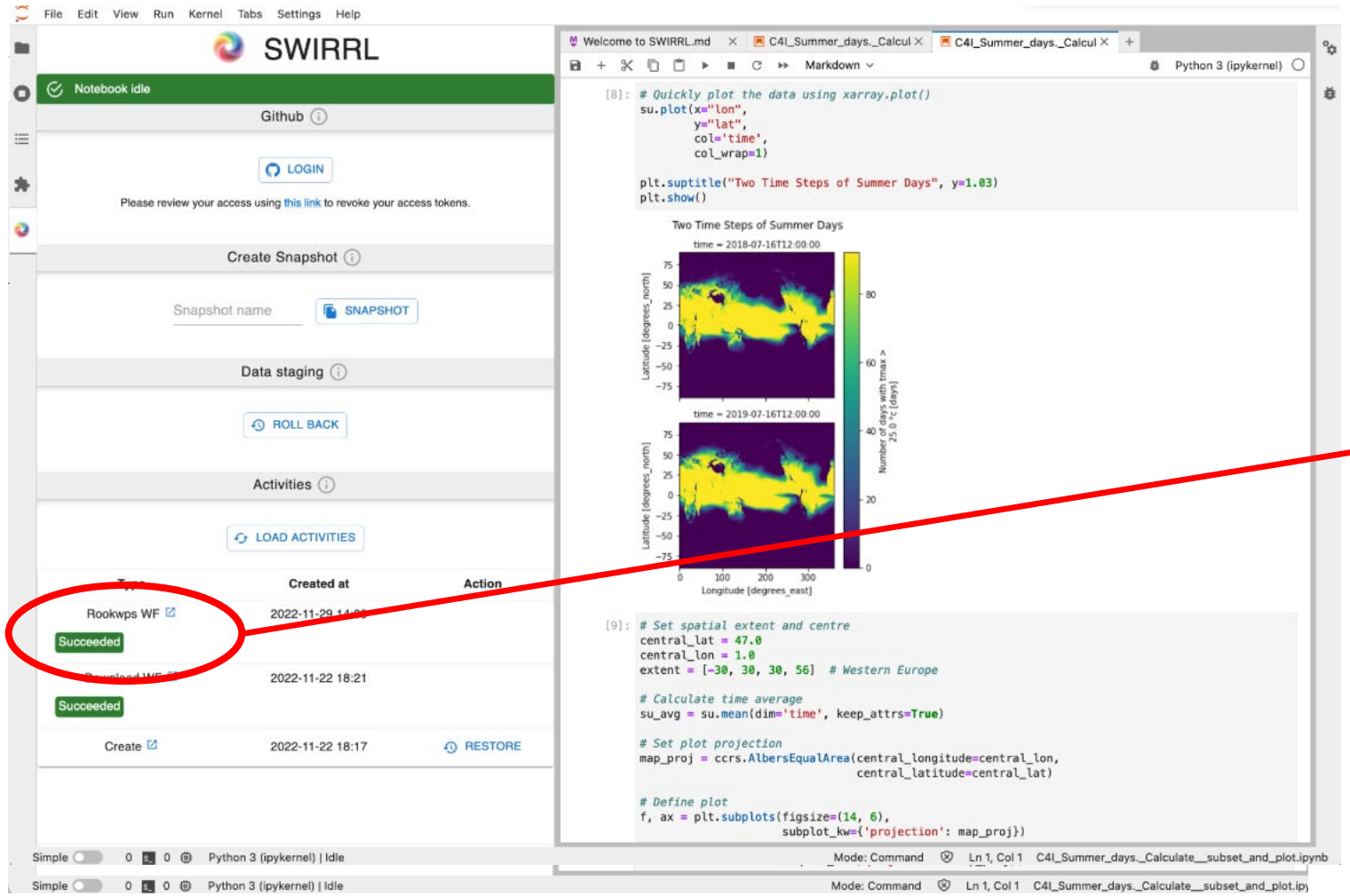
→ requirement to sustain a coordinated “ECAS” service beyond IS-ENES3

OGC Web processing interface based analytics service (WPS) developments

Main achievements:

- Demonstrated suitability as a uniform interface solution for data download and data processing in ENES CDI.
- Stable cooperation around different parts of the SW stack and deployment approach (Copernicus, Canada, OGC)
- Demonstrated production deployment readiness of WPS developments in Copernicus service provisioning
  - load balanced deployment at DKRZ and IPSL (and STFC)
  - rooks WPS (subsetting, averaging, …)
  - relies on distributed ENES CDI data management
- Integrated with C4I
- Provenance reporting for WPS calls (exploited in Copernicus and C4I integration)

Climate4Impact Workspaces - WPS Integration




```
{
  ▸ "prov:used": […],
  ▾ "provone:hadPart": [
    ▾ {
      ▸ "prov:wasAssociatedWith": […],
      ▾ "@type": [
          "Resource",
          "prov:Activity",
          "provone:Execution"
        ],
        "rdfs:label": "orchestrate",
        "prov:startedAtTime": "2022-11-29T14:13:02Z",
        "@id": "urn:roocs:orchestrate_16ca3e1f-fee6-4419-a264-65d273a801bf",
        "prov:endedAtTime": "2022-11-29T14:13:45Z"
      },
    ▾ {
      ▾ "prov:wasAssociatedWith": [
          ▸ {…}
        ],
      ▸ "prov:wasActivityOfInfluence": […],
      ▾ "@type": [
          "Resource",
          "prov:Activity",
          "provone:Execution"
        ],
        "rdfs:label": "average_ta_1",
        "roocs:apply_fixes": false,
        "@id": "urn:roocs:average_ta_1_6eabac6b-444e-4e9a-a71b-bcc1b16b9fb1",
        "roocs:dims": "time"
      },
    ▸ {…}
  ],
  ▸ "@type": […],
  ▸ "@context": {…},
  ▸ "prov:generated": {…},
    "swirrl:sessionId": "13abfb95-fd4a-463f-ba61-fc96d40a9b6c",
    "swirrl:message": "Succeeded",
  ▸ "prov:wasAssociatedWith": […],
    "@id": "urn:uuid:49f0d7f5-5cc8-4893-8e17-2a18ed870540",
    "swirrl:jobId": "49f0d7f5-5cc8-4893-8e17-2a18ed870540",
    "prov:endedAtTime": "2022-11-29T13:13:54.220Z",
    "prov:atLocation": "POST /workflow/rookwps/run/",
    "prov:startedAtTime": "2022-11-29T13:09:03.676Z"
}
```


Thanks to the use of PROV by Rooks and SWIRRL, provenance is easily merged, stored and made available to users and machines (interoperable)

WPS sustainability achievements

Based on modular Ecosystem

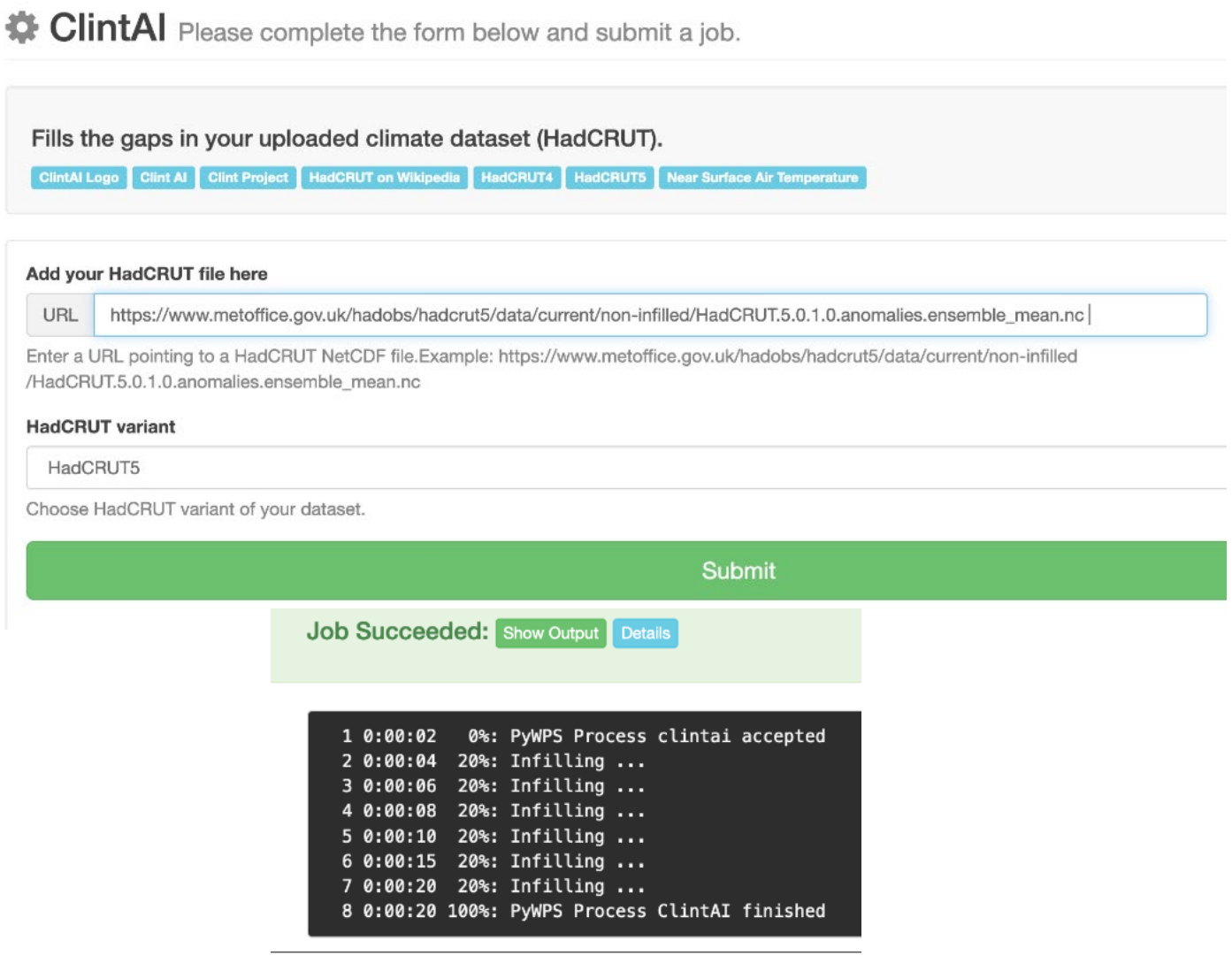
- Client side
  - OGC/OWSLib: WPS client module
  - Birdy: based on OWSLib, used in notebooks
  - Rooki: based on Birdy, used by CDS
- Server side
  - OGC/PyWPS: Python WPS implementation
  - Rook: subsetting service for CMIP and CORDEX
- Libraries:
  - clisops: subset, average, regrid with xarray for CMIP and CORDEX
- Deployment:
  - Ansible playbook
  - Docker

WPS Sustainability

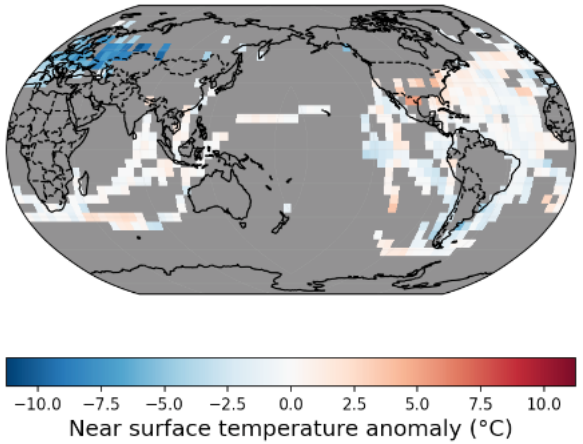
- ENES CDI WPS approach promoted as an ESGF compute service solution
- Community based development: OGC, Canada (Ouranos), STFC, IPSL, DKRZ, …
- Demonstrated usefulness of ENES CDI WPS approach in external projects:
  - Copernicus Climate Data Store (data access for CMIP, CORDEX)
  - H2020 Climate Intelligence Project (CLINT): WPS based AI infilling service

CLINT- WPS based AI Climate Service

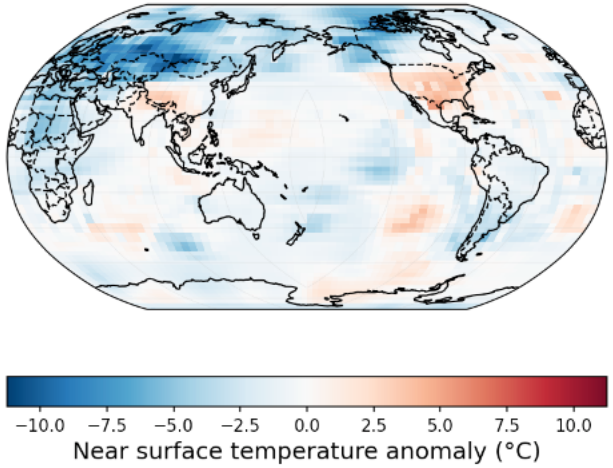
Fill gaps in climate dataset using a web processing service



Before



After



Future Perspectives:

- Good Integration of analytics services with data catalogs is missing
  - changing data pool content not reflected in cross site catalogs (e.g. ESGF)
  - automatic institutional catalog generation vs. manual ESGF publication
  - also hinders cross-institutional data analytics workflows

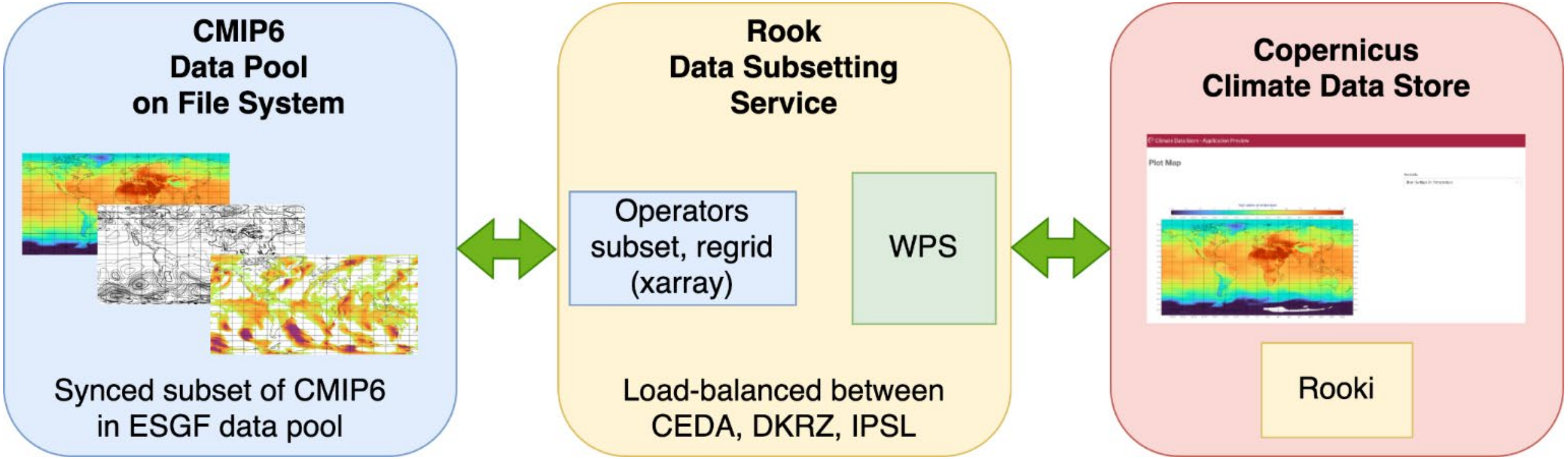
  → future collaboration on Intake/STAC exploitation “beyond” ESGF

- Exploitation of cloud storage in analytics services:
  - currently local, institutional first steps
  - ENES-CDI collaboration important as well as coordination with European infras (e.g. EOSC, EGI, ..) and broader community efforts (Pangeo)

Add on slides …

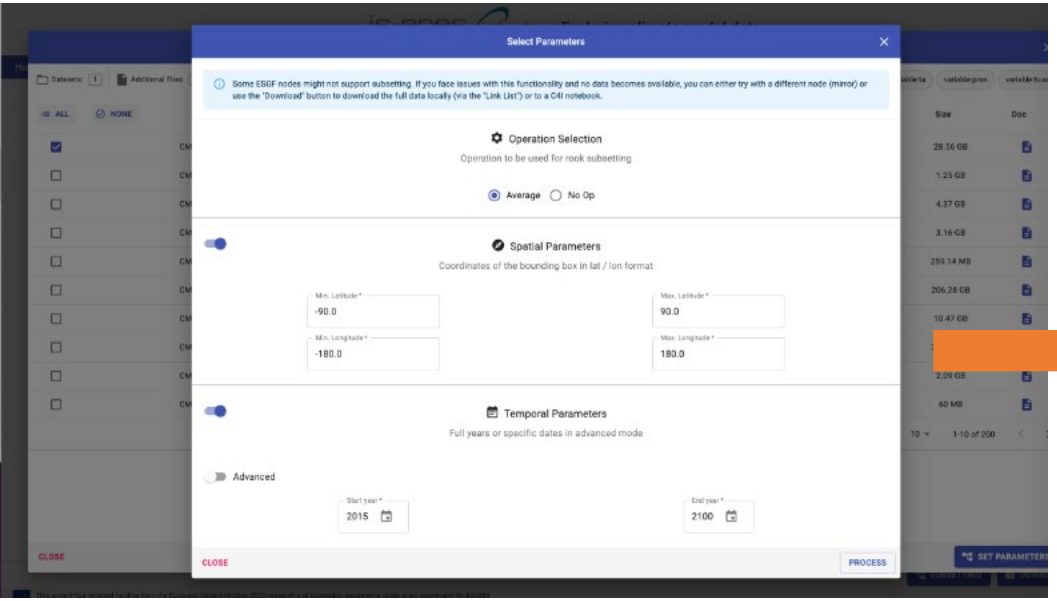
Copernicus Climate Data Store

Rook WPS used by CDS to access CMIP and CORDEX



Climate4Impact Workspaces - WPS Integration

Data Selection and Subsetting configuration enabled for ESGF/WPS nodes



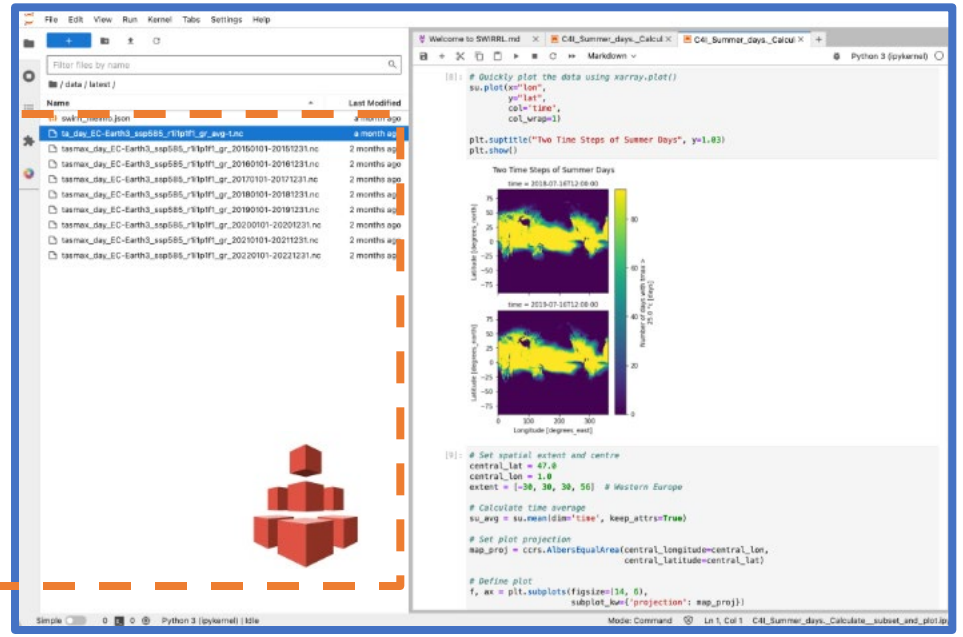
subsets are staged to the user workspace

WPS

SWIRRL

Amazon EKS

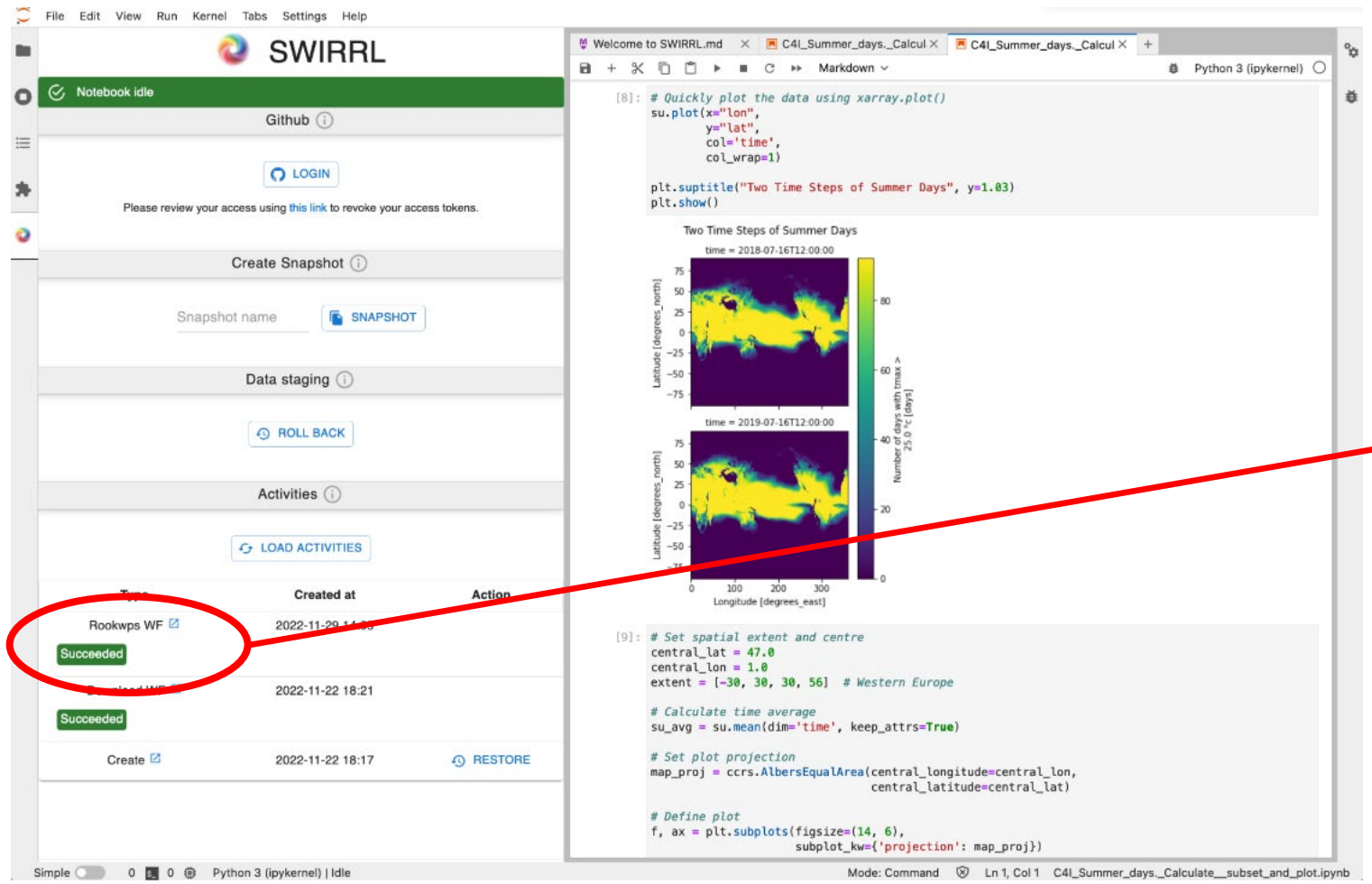
Subsetting Workflow



AWS Fargate

SWIRRL allows to setup shared data volumes, enabling collaboration

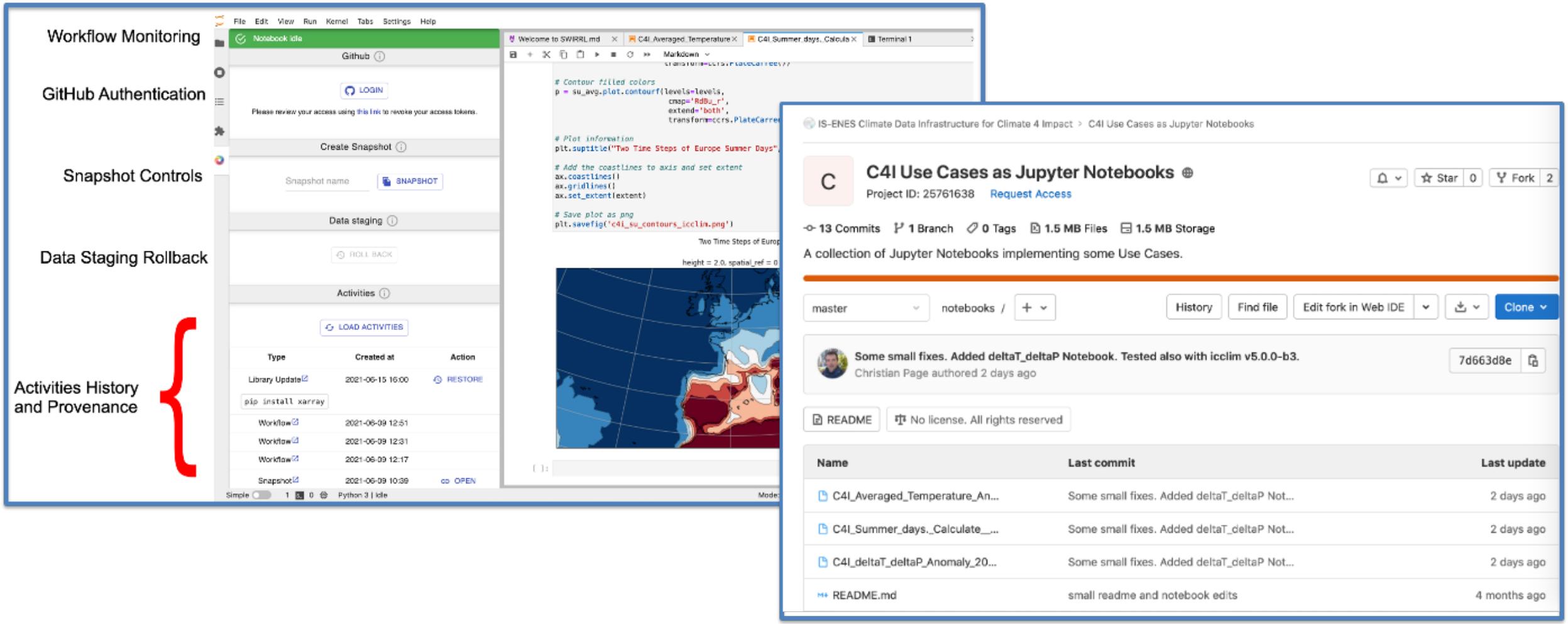
Climate4Impact Workspaces - WPS Integration




```
{
  ▸ "prov:used": […],
  ▾ "provone:hadPart": [
    ▾ {
      ▸ "prov:wasAssociatedWith": […],
      ▾ "@type": [
          "Resource",
          "prov:Activity",
          "provone:Execution"
        ],
        "rdfs:label": "orchestrate",
        "prov:startedAtTime": "2022-11-29T14:13:02Z",
        "@id": "urn:roocs:orchestrate_16ca3e1f-fee6-4419-a264-65d273a801bf",
        "prov:endedAtTime": "2022-11-29T14:13:45Z"
      },
    ▾ {
      ▾ "prov:wasAssociatedWith": [
          ▸ {…}
        ],
      ▸ "prov:wasActivityOfInfluence": […],
      ▾ "@type": [
          "Resource",
          "prov:Activity",
          "provone:Execution"
        ],
        "rdfs:label": "average_ta_1",
        "roocs:apply_fixes": false,
        "@id": "urn:roocs:average_ta_1_6eabac6b-444e-4e9a-a71b-bcc1b16b9fb1",
        "roocs:dims": "time"
      },
    ▸ {…}
  ],
  ▸ "@type": […],
  ▸ "@context": {…},
  ▸ "prov:generated": {…},
    "swirrl:sessionId": "13abfb95-fd4a-463f-ba61-fc96d40a9b6c",
    "swirrl:message": "Succeeded",
  ▸ "prov:wasAssociatedWith": […],
    "@id": "urn:uuid:49f0d7f5-5cc8-4893-8e17-2a18ed870540",
    "swirrl:jobId": "49f0d7f5-5cc8-4893-8e17-2a18ed870540",
    "prov:endedAtTime": "2022-11-29T13:13:54.220Z",
    "prov:atLocation": "POST /workflow/rookwps/run/",
    "prov:startedAtTime": "2022-11-29T13:09:03.676Z"
}
```


Thanks to the use of PROV by Rooks and SWIRRL, provenance is easily merged, stored and made available to users and machines (interoperable)

Climate4Impact Workspaces - on-demand calculations - icclim

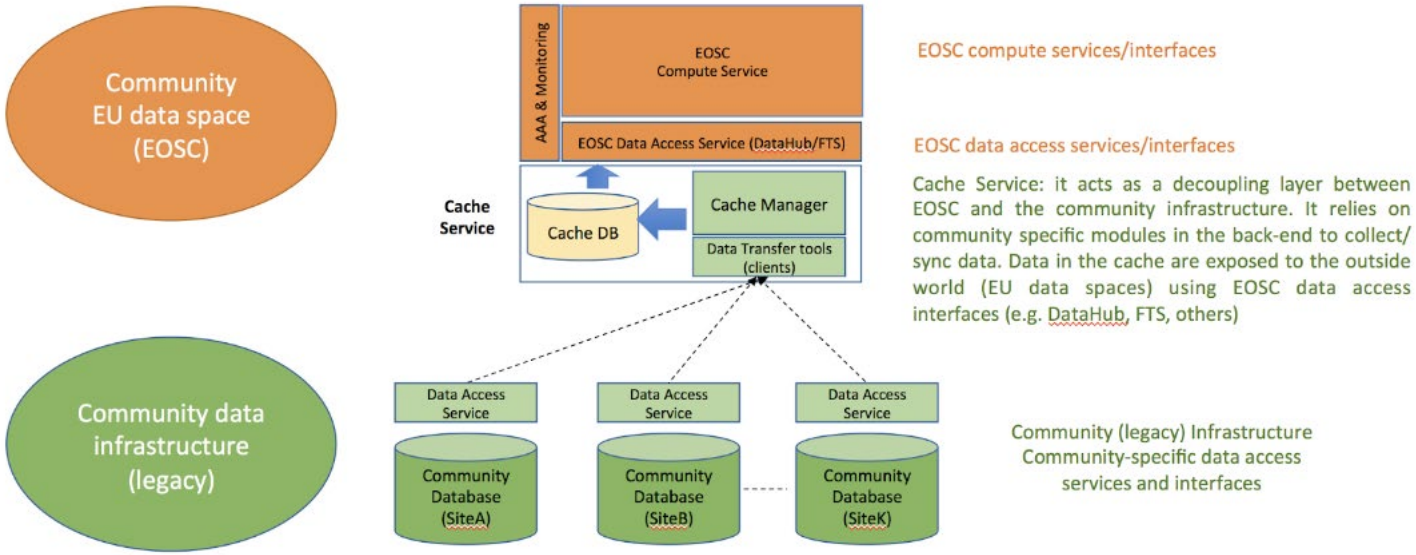


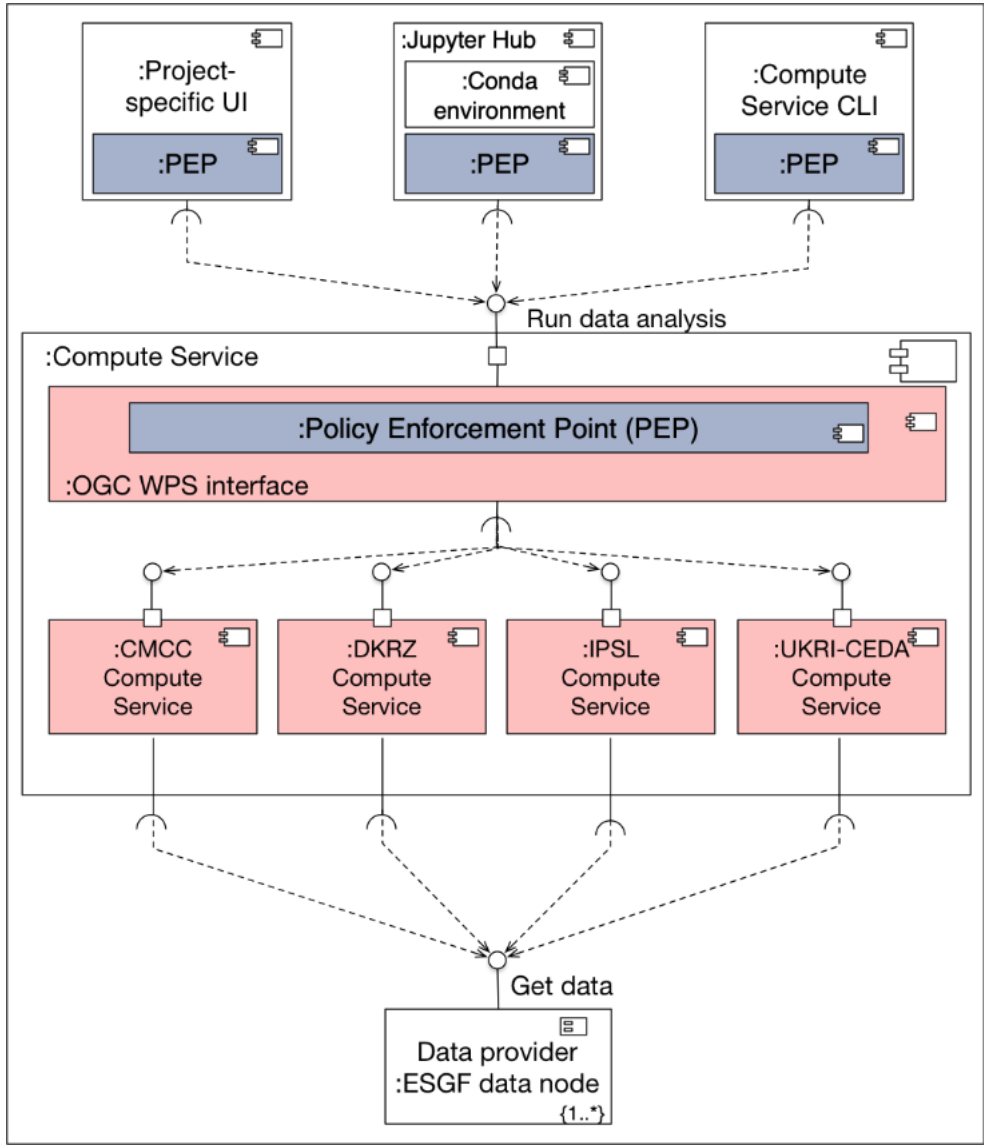
https://gitlab.com/is-enes-cdi-c4i/notebooks

processing service roadmap

Milestones and Deliverables

https://docs.google.com/document/d/1gvcRyeAvO10rk-sfzo0kfdmGW-BfMAud/edit#
https://marketplace.eosc-portal.eu/services/enes-data-space
https://docs.google.com/document/d/1y1izcPtSnCv8HlJIJa9g8BJU8orzK-sn/edit





is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

THE CONSORTIUM

Coordinated by **CNRS-IPSL**, the IS-ENES3 project gathers **22 partners** in **11 countries**

CNRS with CEA, Sorbonne Université
Met Office
National Centre for Atmospheric Science – Natural Environment Research Council
DKRZ – Deutsches Klimarechenzentrum
BSC
CERFACS
CMCC – Centro Euro-Mediterraneo sui Cambiamenti Climatici
Koninklijk Nederlands Meteorologisch Instituut – Ministerie van Infrastructuur en Waterstaat
SMHI
UK Research and Innovation
UC – Universidad de Cantabria
DLR
netherlands eScience
Norwegian Meteorological Institute
METEO FRANCE – Toujours un temps d'avance
MANCHESTER 1824 – The University of Manchester
ΔΗΜΟΚΡΙΤΟΣ
WAGENINGEN UNIVERSITY & RESEARCH
CHARLES UNIVERSITY
NORCE
LINKÖPING UNIVERSITY

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°824084

Our website
https://is.enes.org/

Follow us on Twitter !
@ISENES_RI

Contact us at
is-enes@ipsl.fr

Follow our channel
IS-ENES3 H2020