Web Processing Services for Copernicus Climate Change Service

Carsten Ehbrecht[1], Stephan Kindermann[1], Ag Stephens[2], Sébastien Denvil[3]

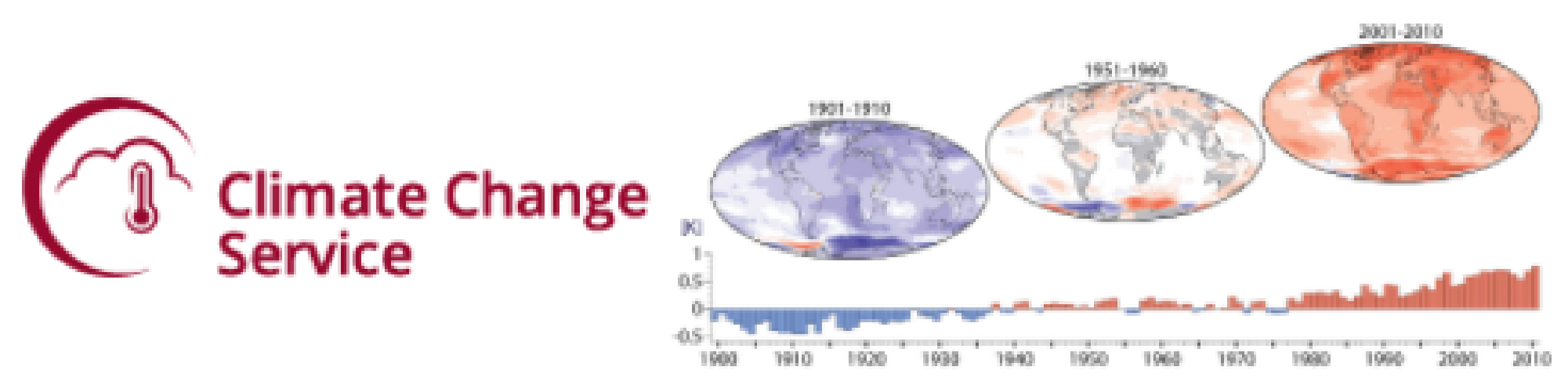
German Climate Compute Centre (DKRZ)[1], Centre for Environmental Data Analysis (CEDA)[2], Institut Pierre Simon Laplace (IPSL)[3]

EGU Vienna, 2018

Copernicus

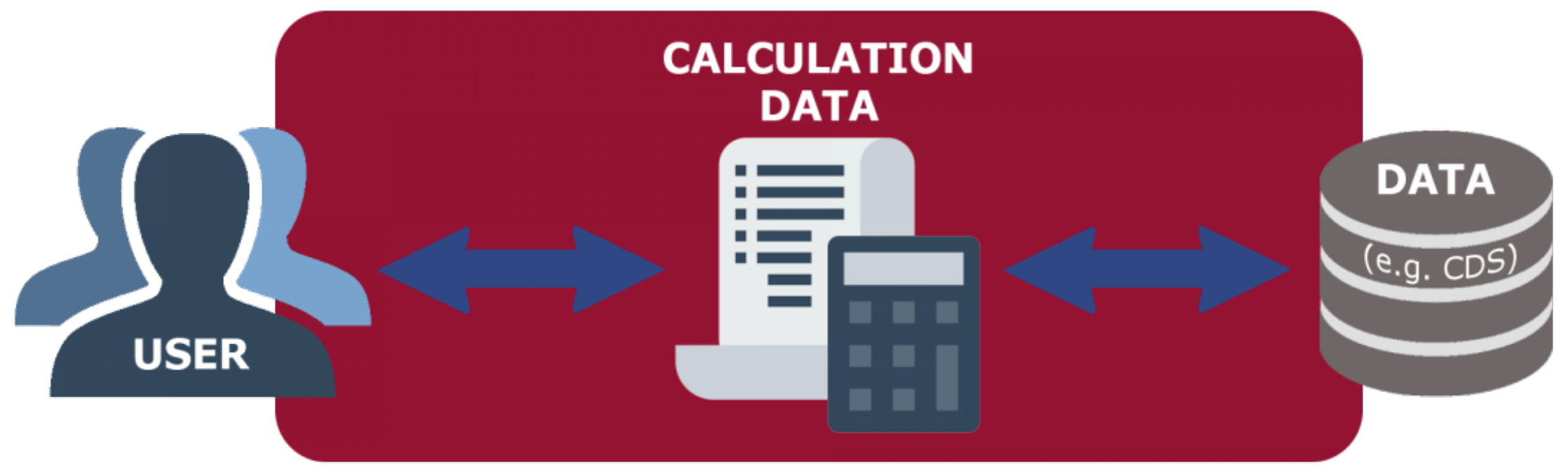
- Copernicus is the **European Union's earth observation programme**
- Data from multiple sources: earth observation, satellites and in situ sensors
- **Thematic areas**: land, marine, atmosphere, **climate change**, emergency management, security
- **Users: policymakers** and public authorities

Copernicus Climate Change Service



- Information for **monitoring and predicting climate change**
- Helps to support adaptation and mitigation
- **Access to Global Climate Model projections** using **well-established metrics and manipulation tools** and receive outputs **tailored to specfic user needs**
- Products for coastal, water, insurance and energy sectors

Climate Data Store (CDS)



- A climate data store will contain the **geophysical information needed to analyse the climate change indicators** in a consistent and harmonised way.
- This will combine the functions of a **distributed data centre** with a set of services and facilities for users and content developers.
- The store will **provide data resources and computing facilities** that can be utilised, for example, to develop improved climate reanalyses and seasonal forecasts.

References

[1] Copernicus Climate Change Service, `http://climate.copernicus.eu/`

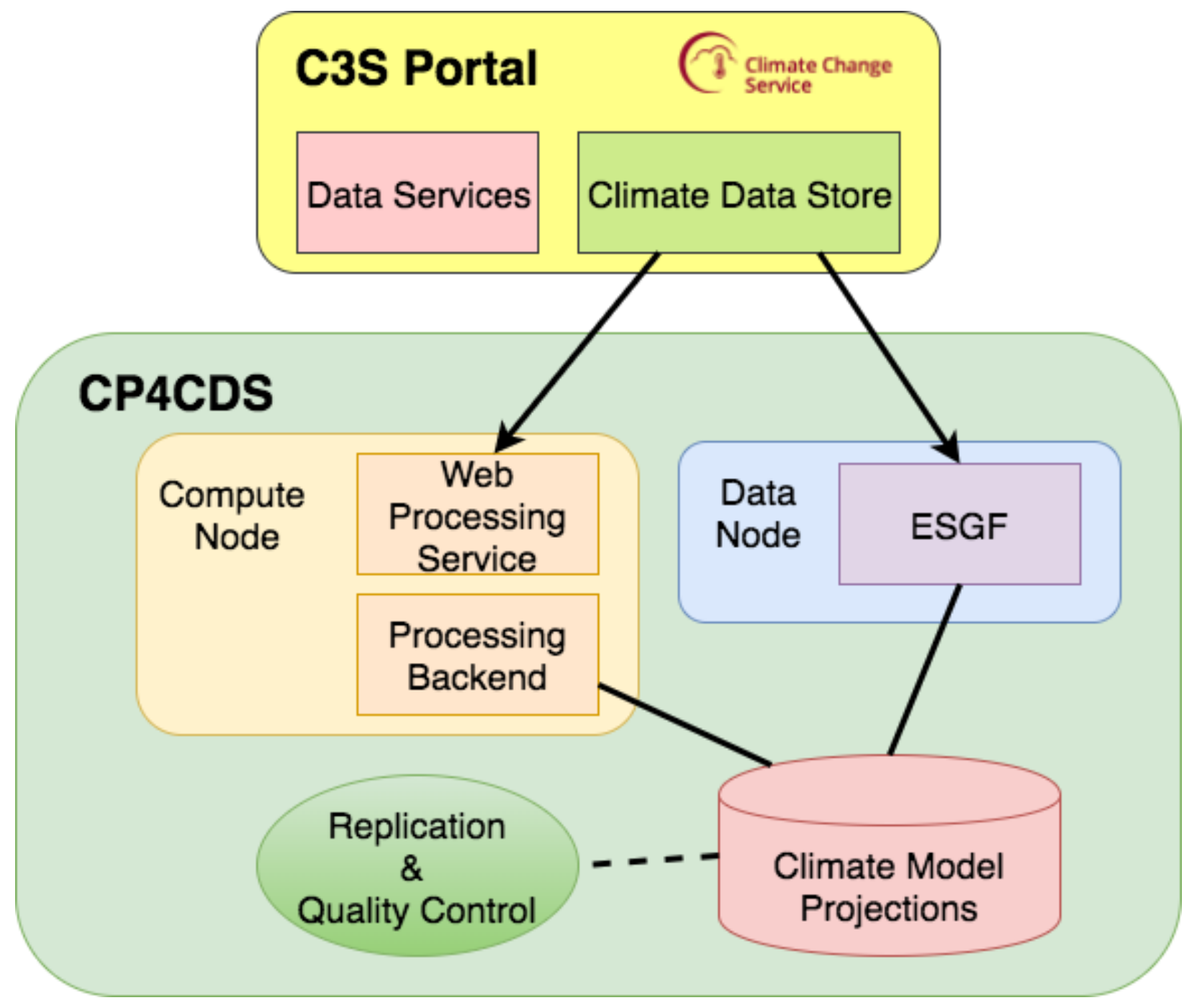
[2] Birdhouse, `http://bird-house.github.io/`

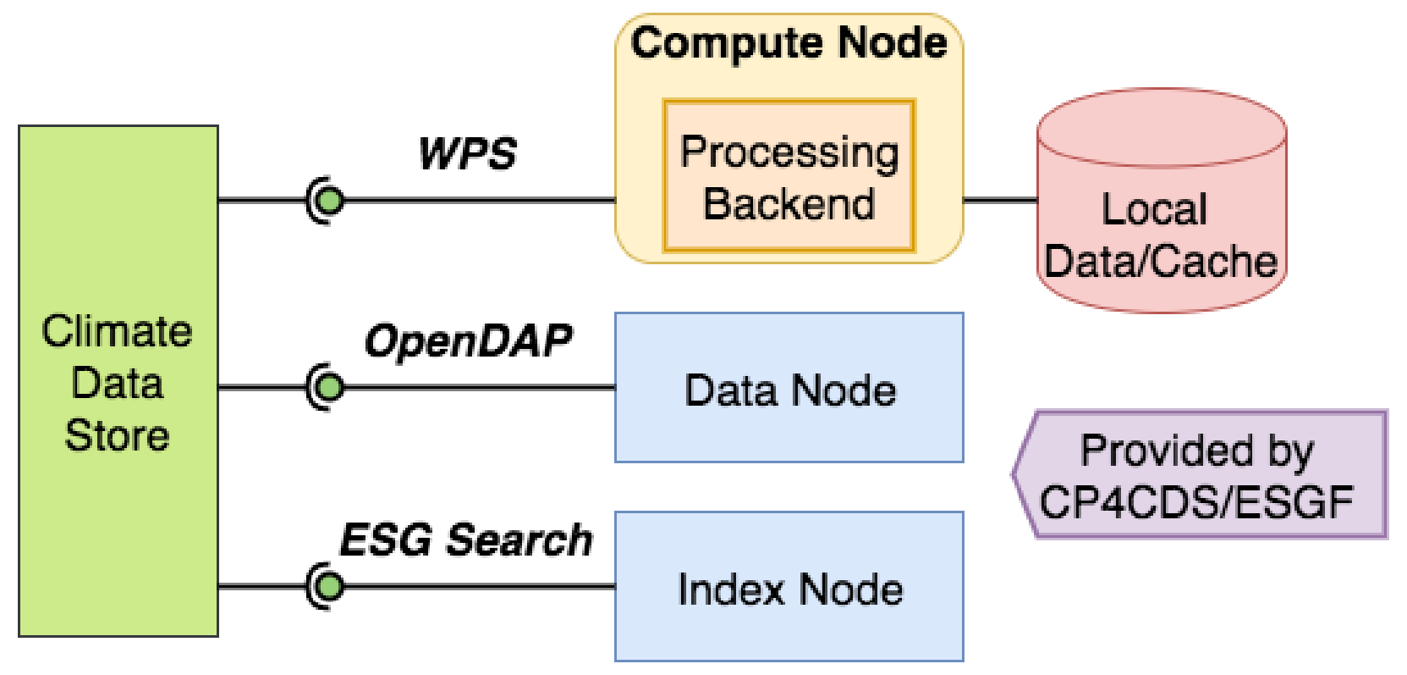
[3] ESGF, `https://esgf.llnl.gov/`

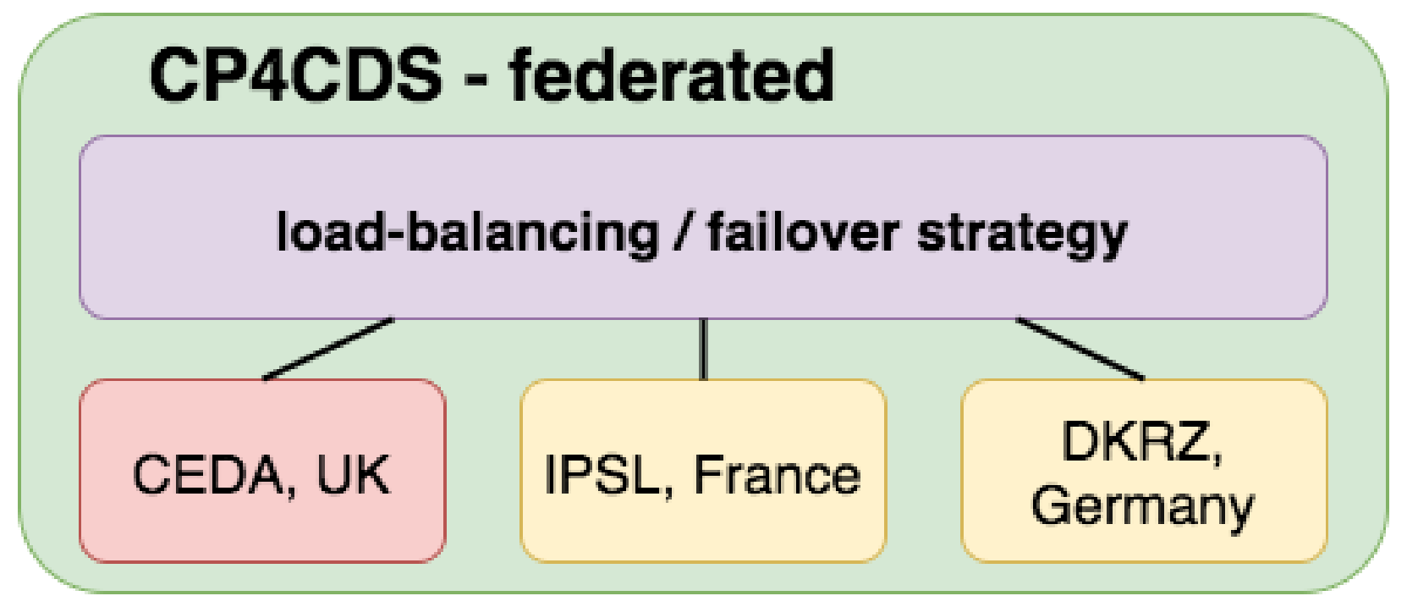
[4] CP4CDS on GitHub, `https://github.com/cp4cds`

[5] C3S MAGIC, `https://github.com/c3s-magic`

Climate Projections for the Climate Data Store (CP4CDS)







CP4CDS Overview

- Providing the required data and services for global climate projections to the **Climate Data Store** (CDS) of the Climate Change Service (C3S) **portal hosted at ECMWF, UK**
- Data Node - Consisting of vanilla **Earth System Grid Federation** (ESGF) index and data node
- Compute Node - Providing compute facilities using the Web Processing Service (WPS) standard interface
- Processing Backend - **External software toolbox** to analyse climate model projections
- Climate Model Projections (CMIP5, CORDEX) in filesystem cache
- Quality Control - **Climate Model Projections** are selected for C3S and **quality checked**
- Replication - Using **Synda** Python library for managing data movement

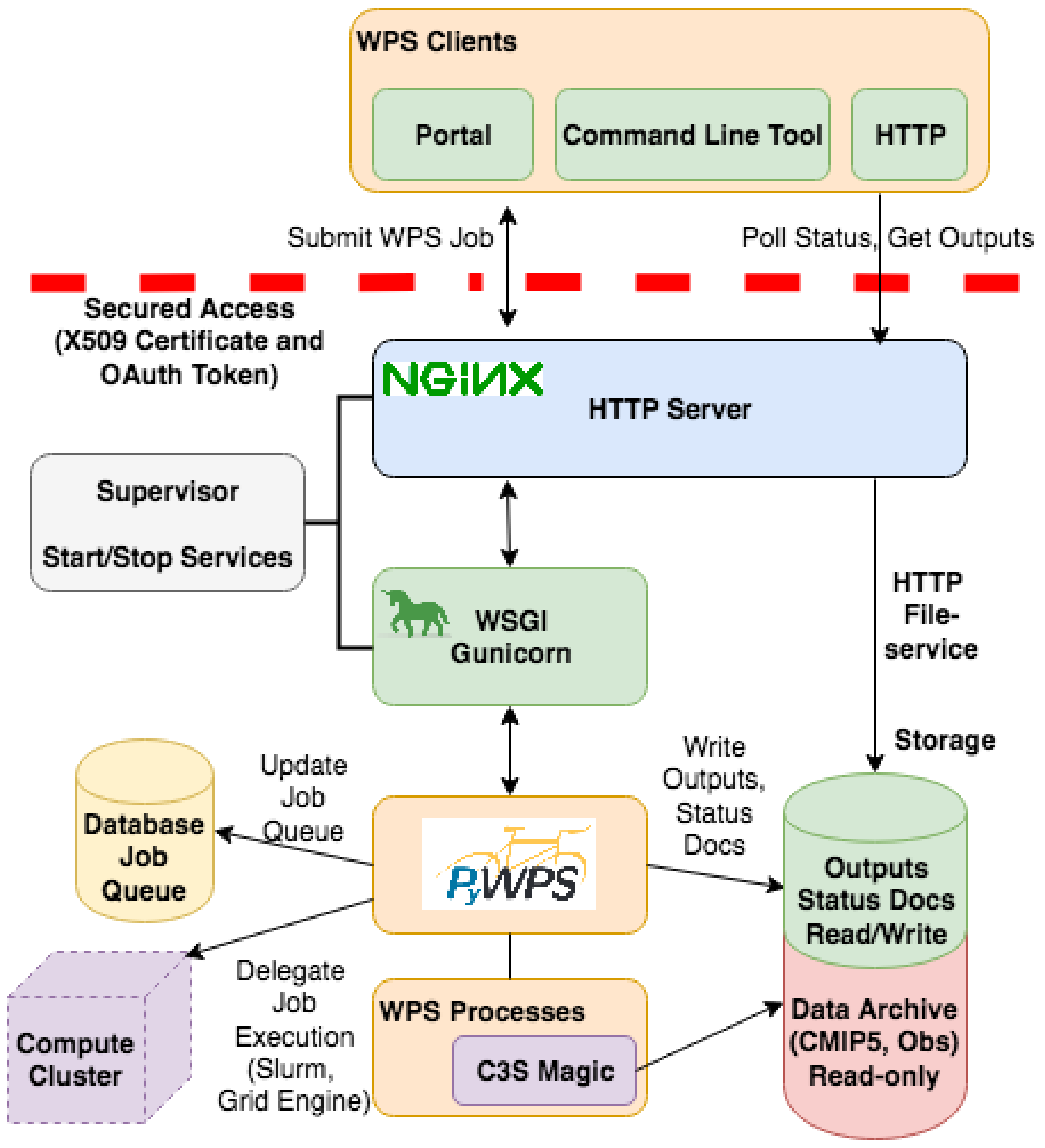
Service Interfaces exposed to Climate Data Store (CDS)

- **Web Processing Service** (WPS) - standard interface for processing
- OpenDAP - remote data access interface for NetCDF files
- ESG Search - adapted Solr search interface by ESGF for data discovery

Federated CP4CDS Nodes

- **Geographically distributed and highly available** set of data and compute services
- Federated between the leading European institutes: CEDA, IPSL and DKRZ
- Using **load-balancing across sites** / failover strategy
- All 3 sites of the **same replicated local data pool**
- All 3 sites have the **same (exact version) software stack** using a common software deployment (**SDDS**)
- **CEDA hosts the main node**, IPSL and DKRZ take over service when needed

Architecture of Compute Node



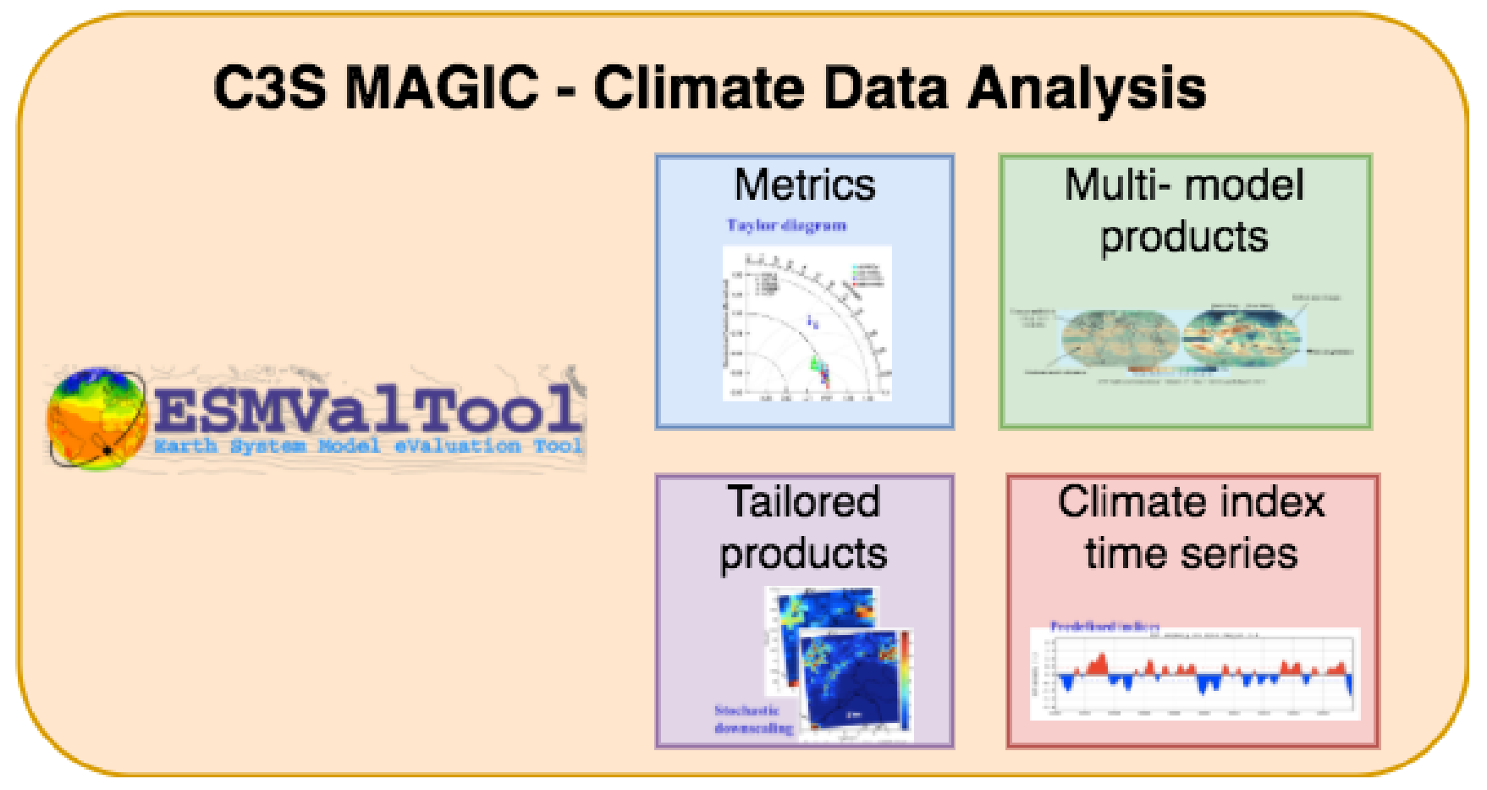
Software Components of WPS Service

- A WPS request (HTTP GET/POST) comes from a WPS client.
- The **Nginx/Gunicorn combination** delegates the request to the PyWPS WSGI application
- Gunicorn - spawns several workers to use the available CPUs on a single compute node
- **PyWPS** - Python implementation of OGC Web Processing Standard
- Supervisor - used to start/stop and monitor services
- Processing outputs and status documents are web accessible by the Nginx file-service
- **Access control** (X509 Certificate and OAuth token) for WPS service
- WPS Processes are defined for project analysis toolbox, like **C3S MAGIC** diagnostics
- **Processing Backend** has read-only access to the climate data pool on file-system with CMIP5 climate model projections and observational data.
- Using **PyWPS scheduler extension** (Slurm, GridEngine) to run process on a compute-cluster for scalability

Status

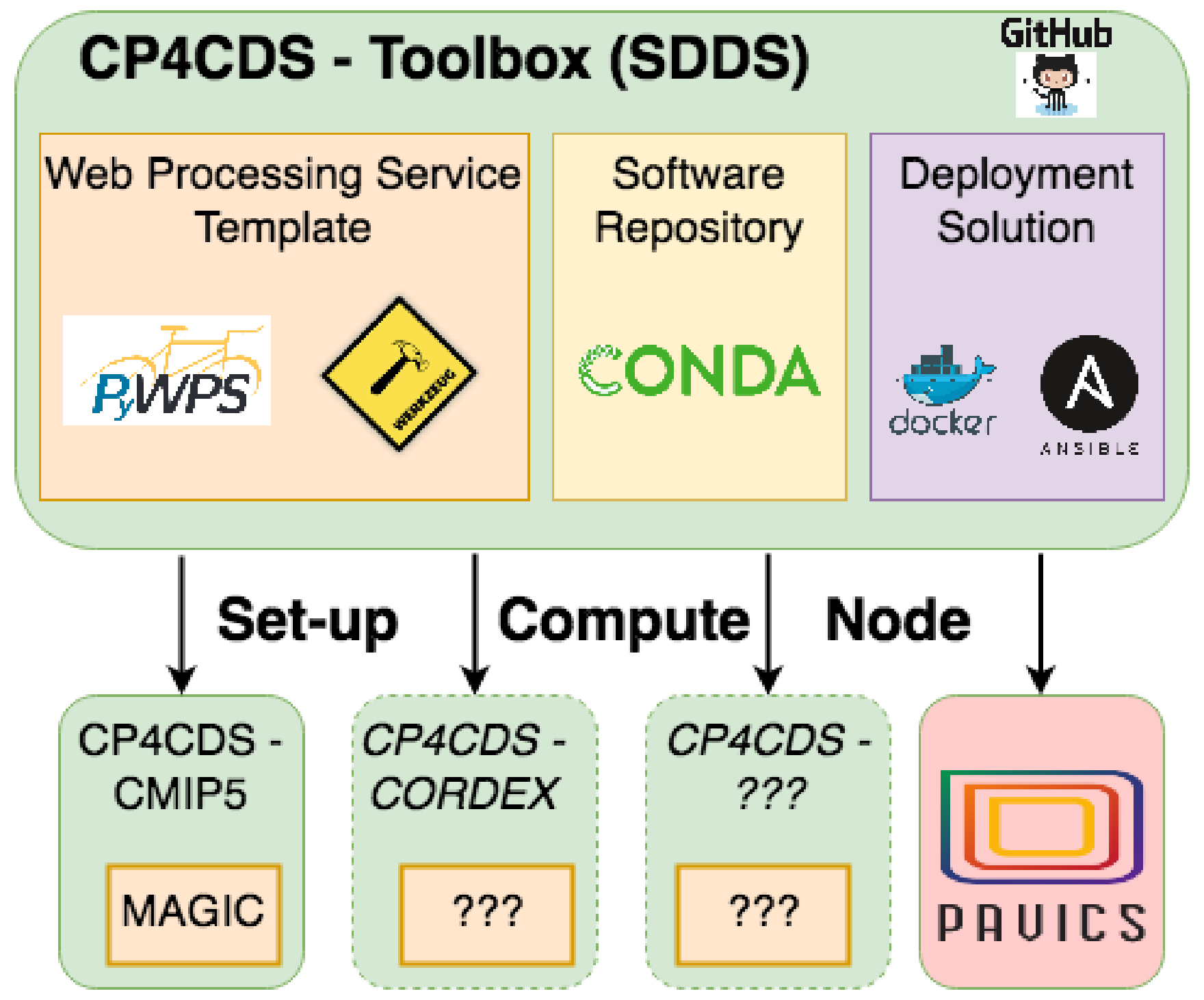
- A Copernicus WPS demo service is available on GitHub (CP4CDS repo)
- The new version of ESMValTool 2.x is used in CP4CDS via Conda package
- A PyWPS scheduler extension for Slurm and Grid-Engine was developed
- A security proxy for WPS services is available using X509 certificates and simple tokens
- A quality checked CMIP5 data store is available at CEDA and partly replicated to other sites

C3S MAGIC - Climate Data Analysis


C3S MAGIC - Climate Data Analysis
ESMValTool
Earth System Model eValuation Tool
Metrics
Taylor diagram
Multi- model products
Tailored products
Stochastic downscaling
Climate index time series

- **C3S MAGIC** - data manipulation and analysis package
- **Developed by KNMI, eScience Center, DLR** and others
- Used by CP4CDS as **Processing Backend** for CMIP5 climate model projections
- To **calculate standardized characteristics** from available climate model output
- **ESMValTool** - To develop and deliver an enhanced version of the ESMValTool software
- **Metrics** - computes and displays a wide set of performance metrics and diagnostics
- **Multi-model products** - To combine the climate information generated by various climate models into a single estimate of any future climate signal
- **Tailored products** - To assure that specific needs of envisaged end users in the selected economic sectors are facilitated by the software

SDDS - Deployment Solution



- **SDDS** - Software Deployment and Dependency Solution
- To **manage and deploy codes** from external projects, such as **C3S MAGIC / ESMValTool**, into the CP4CDS Compute Node
- Consists of a software environment and application, **managed through a GitHub repository**, which includes **a basic template of a working WPS service** (PyWPS)
- **Conda** - Used to record the software dependencies
- **Docker** - Used to provide the Compute Node through containers
- **Ansible** - Ansible and Buildout is used to setup a WPS (PyWPS)
- Used by **PAVICS** (Climate service institute Ouranos, Canada): `https://ouranosinc.github.io/pavics-sdi/`

Next Steps

- Further integration of MAGIC codes
- Roll-out of CP4CDS at all three sites
- Improved SDDS - using a template generator, replacing Buildout by Ansible, deployment in Docker Cluster
- Using ESGF OAuth service for security tokens