

CP4CDS: Climate Predictions for the Copernicus Climate Data Store

An extended ESGF sub-system populated and configured to support the Copernicus Climate Data Store

Martin Jukes ..

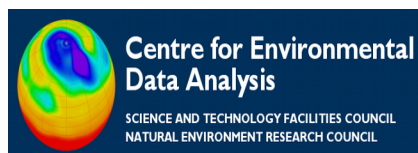
CS3_34a : Global Climate Projections: data access, product generation and impact of front-line developments

Lot 1: Provision of support to one Earth System Grid Federation (ESGF) node in Europe

Feeding global climate projections into the Climate Data Store.



Victoria Bennett, Sarah Callaghan, Sebastien Denvil,
Carsten Ehbrecht, Xia Jin, Martin Jukes, Phil
Kershaw, Stephan Kindermann, Bryan Lawrence,
Guillaume Levavasseur, Ag Stephens, Ruth Petrie.



Aims of CP4CDS

CP4CDS will develop a pre-operational system that will:

- run dedicated CP4CDS ESGF nodes
- deliver a quality assured subset of CMIP5 data to the CDS
- provide compute services,
- be flexible and within a scalable environment,
- provide high standards of reliability and performance (collaborating with IPSL, DKRZ)

for the stakeholders: ECMWF/CDS, Lot2: MAGIC and the ESGF community.

Data from WCRP projections

GFDL Geophysical
Fluid
Dynamics
Laboratory



Beijing Climate Center



Building on ESGF and IS-ENES



ESGF is a global collaboration led by the Program for Climate Model Diagnosis and Intercomparison (PCMDI).



IS-ENES projects, led by IPSL, coordinate European activities contributing to ESGF and beyond.

Data from WCRP projections

- Data is available through ESGF

But

- It is a distributed system with all components running on a best-effort basis;
- Data quality standards are set, but compliance cannot be enforced;
- Variations in range of services provided;

Data Management

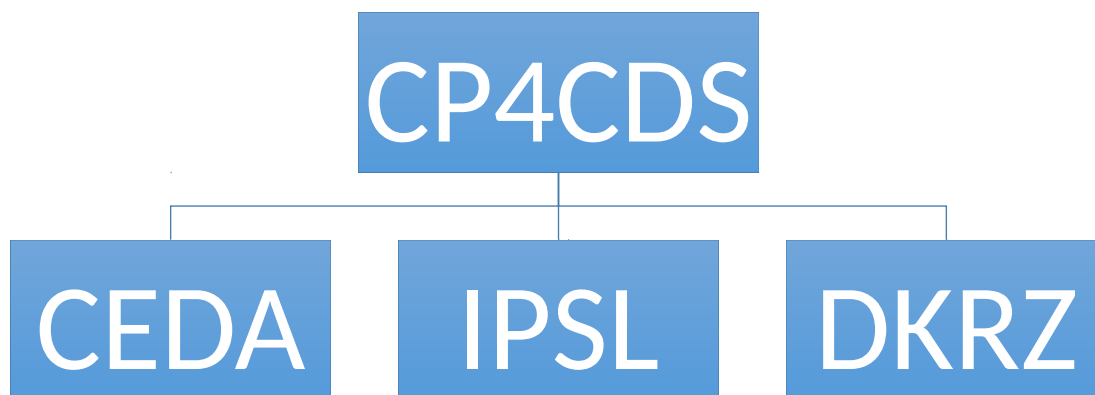
- Provide a data management plan for a quality controlled subset of CMIP5 data
- Provide a database with quality control information on the subset of CMIP5 data
- Where possible data will be corrected before being provided to the CDS.



System Integration and operation

The data and compute services will be deployed at three sites, requiring:

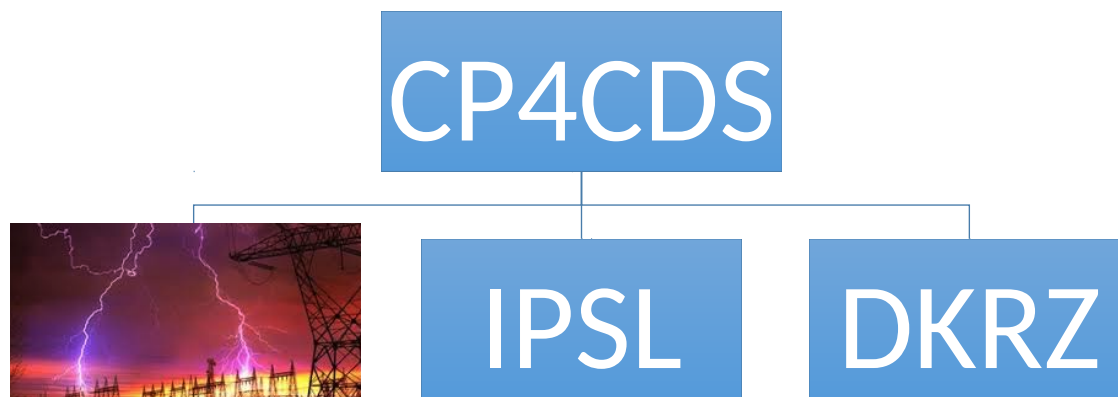
- Load balancing;
- Aggregated up time to 99% (production level);
- A “master” index node that contains the most up to date version of the catalogue and ensuring all nodes are fully in sync.



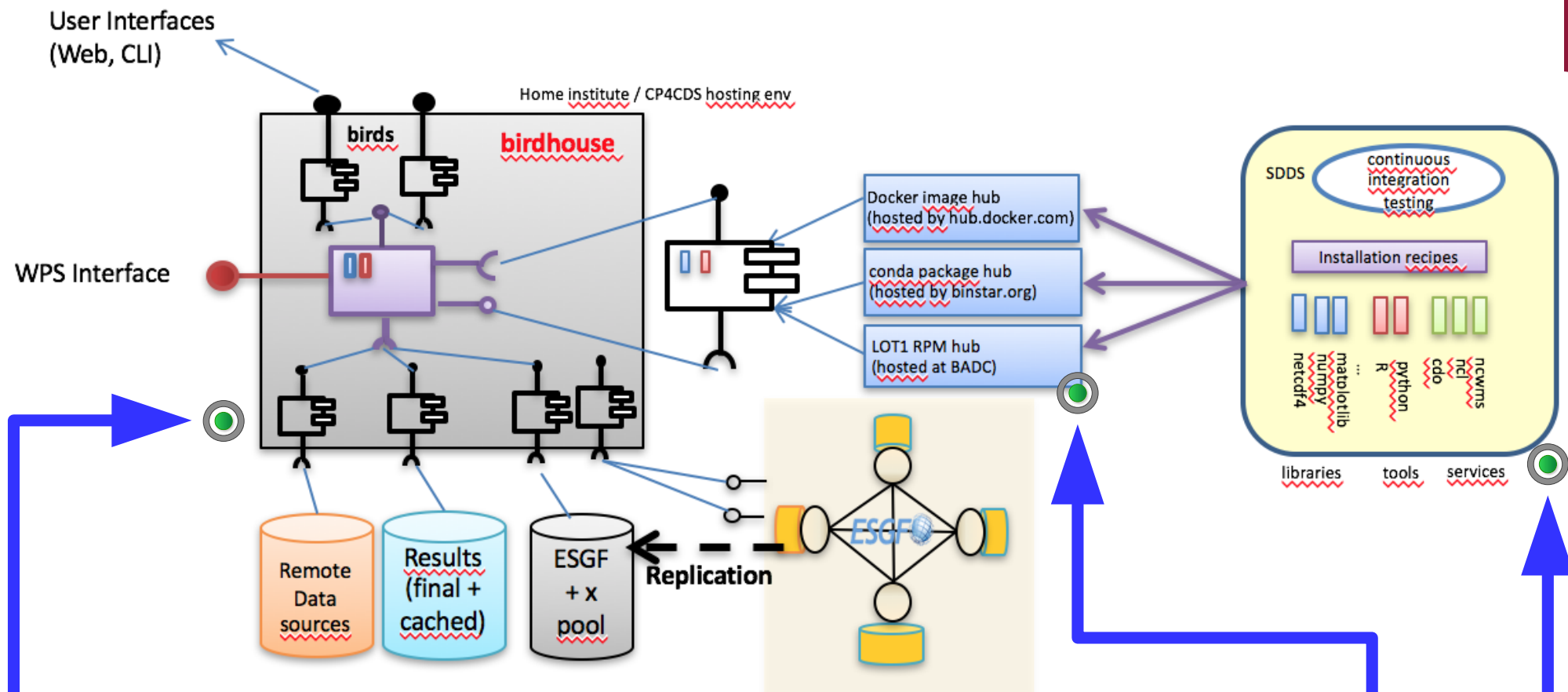
System Integration and operation

The data and compute services will be deployed at three sites, requiring:

- Load balancing;
- Aggregated up time to 99% (production level);
- A “master” index node that contains the most up to date version of the catalogue and ensuring all nodes are fully in sync.



Compute Node



Virtualisation ("containerisation") of the complete processing environment to enable deployment at multiple sites;
 Scientific analysis codes embedded in the containers;
 Data and processing suite brought together following user requests;

Summary

For C3S:

- Providing a quality controlled suite of global climate; projections (cmip5);
- Robust services with 99% availability;
- Processing environment through containerisation.

For the community:

- Improved QA information;
- Tools and software available for ESGF;

Project structure

WP	Name - Leader	Description
1	Management and CDS liaison	Responsible for overall architecture, reporting, workshops, engagement.
2	Data Management	Responsible for ensuring appropriate data available and fit for purpose.
3	Data Node Software	Responsible for the software deployed in the data node (and elsewhere)
4	Compute Node Software	Responsible for compute node software (which may or may not be used elsewhere)
5	Interface and Tools	Responsible for tools and interfaces used by those building code to deploy in compute nodes.
6	Integration and Operation	Responsible for the integration and deployment of the running system(s)