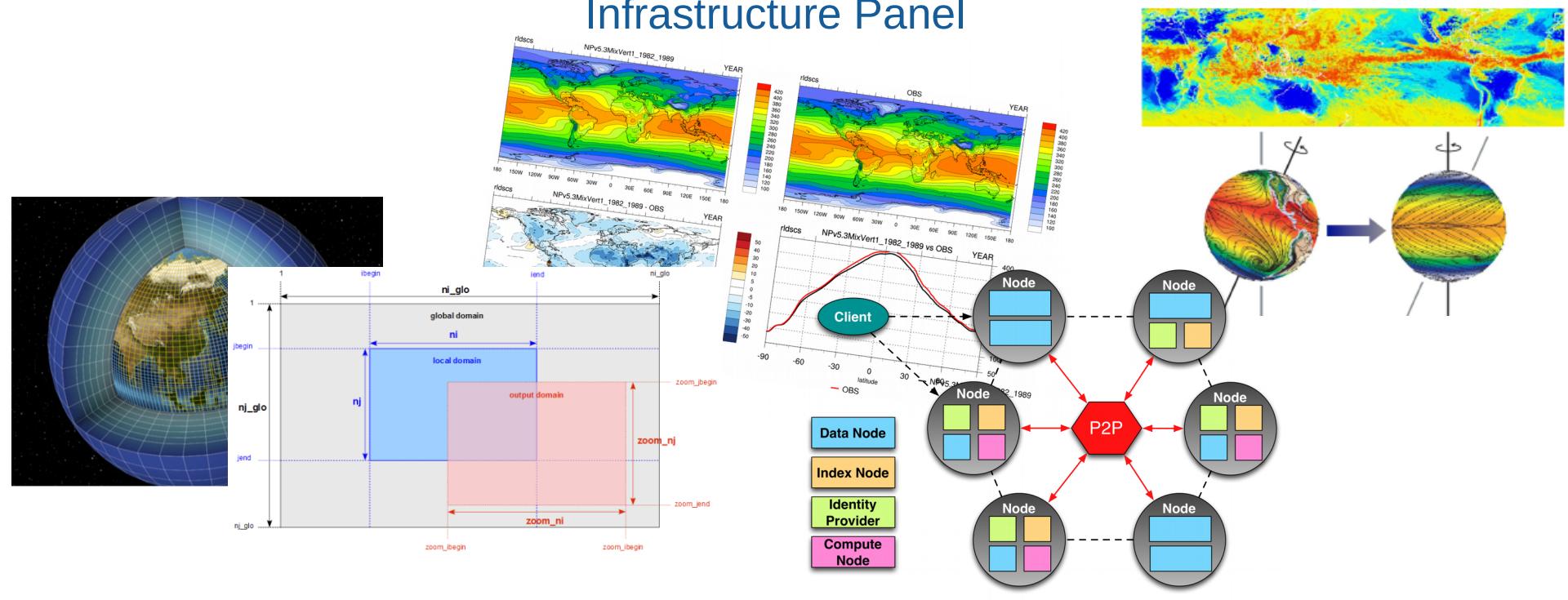


# CMIP6 ESGF Tier 1 and Tier 2 Nodes

Sébastien Denvil (IPSL), Michael Lautenschlager (DKRZ).  
 With contributions from ESGF Executive Committee and WGCM  
 Infrastructure Panel



# Preamble

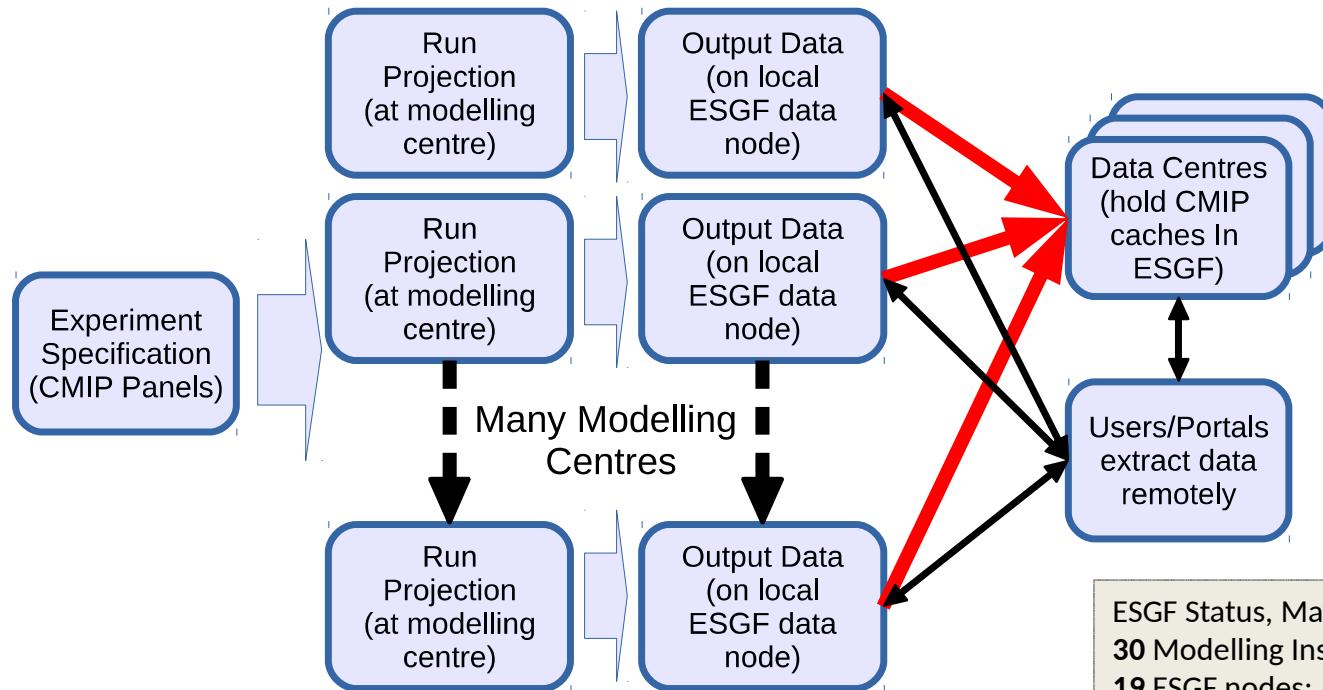
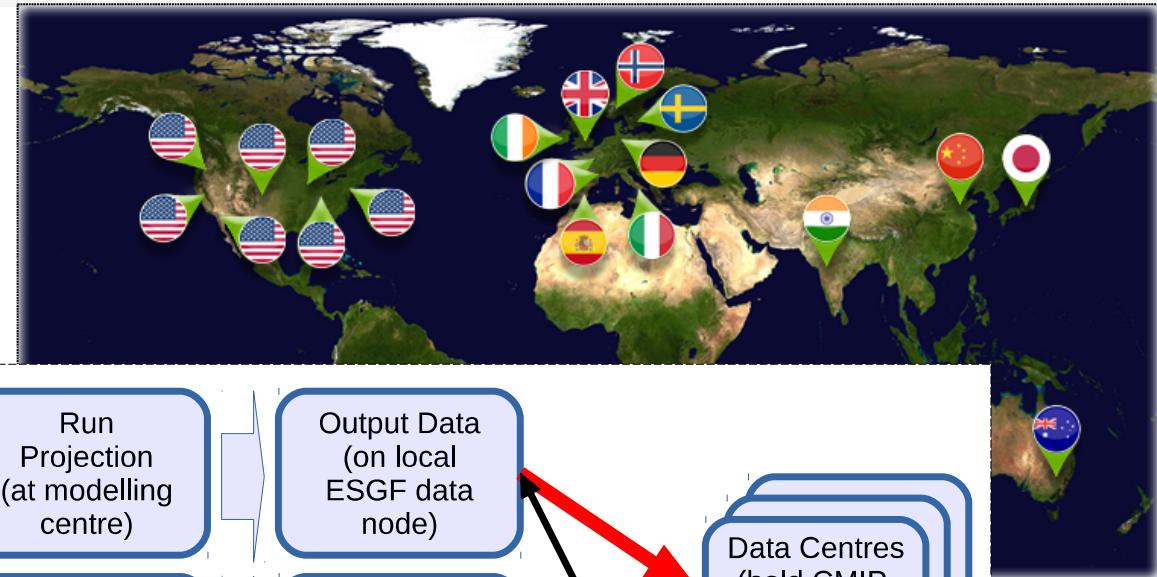
Tiers 1 & Tiers 2 requirements for CMIP6. State of play and what are your views and questions on this.

- The ESGF Executive Committee tasked Michael Lautenschlager and Sébastien Denvil to draw the Tiers 1 / Tiers 2 requirement for ESGF.
- After discussions and iterations within the ENES Data Task Force we came up with the basis I show today.
- We try now to collect feedbacks from group like the CDNOT and you.
- As mentioned we should also have an inventory of the amount of cores a Tier 1 candidate will make available for the community.

# ESGF Data Infrastructure

**Institute**

- BCC (6891)
- BNU (512)
- CCCMA (21879)
- CMCC (1189)
- CNRM-CERFACS (5422)
- COLA-CFS (884)
- CSIRO-BOM (625)
- CSIRO-QCCCE (3120)
- FIO (230)
- ICHEC (2966)
- INM (486)
- INPE (24)
- IPSL (10637)
- LASG-CESS (2568)
- LASG-IAP (418)
- MIROC (11509)
- MOHC (23452)
- MPI-M (11654)
- MRI (4804)
- NASA-GISS (4781)
- NASA-GMAO (1620)
- NCAR (4922)
- NCC (1014)
- NCEP (870)
- NICAM (15)
- NIMR-KMA (63)
- NOAA-GFDL (14932)
- NSF-DOE-NCAR (949)
- SMHI (1840)
- UNSW (26)



ESGF Status, May 2015:  
**30** Modelling Institutes;  
**19** ESGF nodes;  
**6** index nodes.  
**140,032** datasets of which  
**58,174** European (42%)

# Data provider information (22 PB)

**IPSL** provisioned 2 PB. CMIP6 simulation starts January 2017

**GFDL** provisioned 1.5 PB CMIP6 simulation starts January 2017

**NCAR** 2 to 4 PB. CMIP6 simulation starts January 2017

**CNRM** 2 PB. simulation starts (probably end of 2016)

**ANU** 2 PB. simulation starts January 2017

**UR** 2 PB. simulation starts January 2017

**BCC** 1 PB. simulation starts January 2017

**CMCC** 1 PB. expected end of 2016/begin of January 2017.

**CEDA**. 1-2 PB. CMIP6 simulations expected end of 2016.

**DKRZ** 2-3 PB. simulations start 2017

**BNU** 1PB. CMIP6 simulation starts January 2017.

**FGOALS** 1-2 PB. CMIP6 simulation starts January 2017.

**DIAS** 1-2 PB. and possible extension to 2-4PB. CMIP6 protocol starting january 2017



# Contribution for replication

**PCMDI** 5 years plan. Up to 20 PB by late 2020

**DKRZ** 5 PB online (disk), up to 50 PB offline (HPSS)

**CEDA** Hope to contribute up to 2 PB, > 2 PB offline

**ANU** Hope to contribute up to 10 PB

**IPSL** Hope to contribute up to 3 PB

**BNU** Hope to contribute up to 3PB

**FGOALS** Hope to contribute up to 3PB

**TOTAL** : 45 PB spinning and 42 PB on Tape



Tier 1/2 requirements for ESGF infrastructure  
and on a per project basis.



# Tiers 1 for ESGF infrastructure

- o Level of service:
  - + uptime more than 90% / 95%,
  - + installation of most recent version of SW (Tier 2 as well, time limit for installation to be defined),
  - + installation of complete SW stack,
  - + contribution to development and maintenance
- o Support for Tier 2 data nodes
- o Support for data providers

# CMIP6

- o Spinning disks
  - + Tier 1 for initial data publication and data replication  
(CMIP6: around 4-5 PB)
  - + Tier 2 for initial data publication (CMIP6: about 2 PB)
- o Compute resources (Tier 1)  
Hard to specify, no experience yet



# CMIP6

- o Network connection

- + Tier 1: optimisation of nominal bandwidth of 10 GBit/s will result in 30 – 50 % for real bandwidth for replication. This together with the specification of the core data set defines the CMIP6 replication strategy.
- + Tier 2: 1 – 2 GBit/s for data provision. CMIP5 experience showed that each data node provides 10 times the data it hosts over a period of 4 years and the average available network bandwidth should cover this.

# CMIP6

- o Tapes

- + Single Tier 1: about 20 PB plus for long-term archiving of reference data from the CMIP6 data (volume not clear yet)
- + Tier 1: tapes to fill the storage gap in case of insufficient disc space for initial data publication and data replication



# Global data infrastructure

- NAGIOS (<https://www.nagios.org>) is the Industry Standard In IT Infrastructure Monitoring
- NAGIOS have many plug-ins that we can use to monitor our infrastructure
- The proposal is to use NAGIOS to monitor CMIP6 nodes and to guard against
  - Expired certificates (host & globus certificates)
  - http/https endpoints unavailable
  - Gridftp endpoints unavailable
  - Etc ....

# Global data infrastructure

- Tiers 1 node could take the responsibility to monitor data node publishing to them
- Tiers 1 node could also monitor themselves
- There is an ongoing proposition to enable ESGF to exclude a data node that does not satisfy all the CMIP6 requirements or a data node that will degrade the federation usability. The implementation of this is currently under discussion.
- NAGIOS like monitoring is necessary to ease our lives providing a high level of service for CMIP6.



# Thank you for your attention

