ENES DTF Telco 16-09-23, 15 – 16:30

Attendees:
Bryan Lawrence (NCAS), Sylvie Joussaume (IPSL), Martin Juckes (CEDA), Sebastien Denvil (IPSL), Frank Toussaint (DKRZ), Stephan Kindermann (DKRZ), Christian Page (CERFACS), Francesca Guglielmo (IPSL), Michael Lautenschlager (DKRZ, Minutes)
Excused: Sandro Fiore (CMCC), Wim Som de Cerff (KNMI), Michael Kolax (SMHI)

Topics that have been discussed:

CMIP6 data volume: European contributions, European data node resources and European replication strategy
CMIP6 data volume estimates exist from two different approaches (August 2016), from the WIP and from CDNOT.

The WIP estimated the data volume via a Phython script which evaluates CMIP6 modelling group input via WGCM. The results are provided at the WIP web pages under the topic "Ressources" and the sub-topic "CMIP6 Data". The CMIP6 data count at model resolution results in 18 PB, and the count at a standard resolution gives 1.8 PB except for the HighRes-MIP which is not included.
At the WIP telco September 20th the suggestion was to specify a core data set of 10% of the CMIP6 data volume for replication at tier-1 sites. The core data set may contain almost all CMIP6 data on a standard grid or selected variables on the model grid. The specification of CMIP6 core data is planned to be discussed at the upcoming WGCM meeting begin of November.

The CDNOT direct CMIP6 data volume estimate from telco attendees sums up to 18 PB which corresponds to the WIP estimate but it covers only a subset of modelling groups. The European contribution is 8 PB of CMIP6 data while DKRZ is missing. Including DKRZ Europe plans to contribute 10 PB to the CMIP6 data volume.
The CMIP6 data production should be balanced by ESGF storage resources for initial data publication and replication of CMIP6 core data. The CDNOT direct storage resource estimates gives a total: 28 PB disc space which is planned to increase to 43 PB by 2020. The European contribution is presently 6 PB disc space.

Status for Europe (August 2016) is that a CMIP6 data production of 10 PB faces CMIP6 disc storage resources of 6 PB. CMIP6 data production and available disc space is presently unbalanced in Europe. The gap can be closed by integration of tertiary storage (tapes). The balance between CMIP6 data production and available disc space looks better at the global scale.

Specification of ESGF tier1/tier2
The discussion about the specification of tier1/2 data nodes started at ENES-DTF Face-to-face meeting in Paris (June 2016) and should be separated into specification with respect to the ESGF infrastructure development and maintenance (more generic) and with respect to data projects like CMIP6 (more resource oriented). A compilation of bullet points contains:
- ESGF infrastructure
    - Level of service:
        - uptime more than 90% / 95%,
        - installation of most recent version of SW (tier 2 as well),
        - installation of complete SW stack,
        - contribution to development and maintenance
    - Support for tier 2 data nodes

- o Support for data providers
- Data projects (like CMIP6)
  - o Spinning disks
    - Tier 1 for initial data publication and data replication (CMIP6: 4-5 PB)
    - Tier2 for initial data publication (CMIP6: about 2 PB)
  - o Compute resources (tier 1)
    - Hard to specify, no experience yet
  - o Network connection
    - Tier 1: optimisation of nominal bandwidth of 10 GBit/s results 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 shows that each data n odes provides 10 times the data they host over a period of 4 years and the average available network bandwidth should cover this.
  - 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

C3S Regional Climate PIN

A request from SMHI reached the ERA4CS consortium for participation and specifically for coordination the C3S Regional Climate PIN LOT1 by one of the European partners. In C3S are lot of more work is expected for the CORDEX data in comparison to ERA4CS for CMIP5 data because of the inhomogeneity of CORDEX data and the additional requirement to serve forcing data form global climate models as well. The ERA4CS consortium members are happy to participate but the coordination of C3S LOT1 is seen at SMHI as first option because of synergy with LOT2 and with the coordination of CORDEX itself.

ENES Strategy and Convergence Workshops, October 24 - 27th in Reading

Bryan summarised briefly the outline of the planned IS-ENES workshops in Reading end of October. It was decided The ENES DTF will present status and strategy of the ENES climate data infrastructure. This includes storage, network, processing as well as global (ESGF) and European (EOSC) integration. Stephan and Michael prepare a draft presentation to be coordinated with the DTF members.

Other Business

- ESGF release management
  The ESGF V2.4 has been prepared and tested over summer but integration of the ESGF installation package is still missing. A single point of failure has been identified in the ESGF release management workflow at least at this point. Sebastien will try to solve to current problem with ESGF 2.4 and Michael will communicate the discussion to the ESGF XC in order to aim for a more robust ESGF release management.
- Sylvie reported from her recent discussions at PCMDI and the idea to invite the CMOR developer at PCMDI to Europe for more intensive exchange about the CMOR software stack. Possible hosting institutes could be Hadley Centre, CEDA, IPSL or DKRZ. In parallel o common workshop in the frame of the IS-ENES coding sprints might be helpful.
  Within the IS-ENES2 project funds for EU-US cooperation are still available. Sylvie will coordinate this with PCMDI and with the ENES DTF.