



Data storage and preservation of high resolution climate experiments

This pilot focuses on facilitating long-term storage and sharing amongst a wide scientific user community of high-resolution climate model output data. It aims at building a repository serving the climate change impact modelling community, providing selected variables at high temporal and spatial resolution, with a focus on climate extremes and the hydrological cycle in areas with complex orography. Potential users include researchers studying the impacts of climate on ecosystems, floods, landslides, fires. The archive will contain high-resolution data from the PRACE project Climate SPHINX and will later be extended with simulations from the projects PRIMAVERA, CRESCENDO and HighResMIP.

The Scientific Challenge

An open issue which is currently being actively investigated is the sensitivity of climate simulations to model resolution and determining if very high resolution is useful for a realistic representation of the main features of climate variability. The advantage of sub-grid parameterizations capable of capturing small-scale variability, such as stochastic parameterizations, has to be determined. To this end extremely high resolution climate integrations are necessary and they are being performed or planned in the framework of several initiatives (Climate SPHINX, HighResMIP, CRESCENDO, PRIMAVERA).

In a first stage the EC-Earth Earth-System model is being used to explore the impact of Stochastic Physics in long climate integrations as a function both of model resolution (from 80km to 16km for the atmosphere). This research will systematically and extensively investigate the impact of resolution and stochastic parameterisations for climate simulations. As a result, we estimate data storage needs around 50-300 TB in this stage. In a subsequent stage, the archive will be further expanded with high-resolution coupled simulations performed mainly with the EC-Earth model in the framework of the CMIP6 HighResMIP initiative and of the PRIMAVERA & CRESCENDO projects.

For this second phase the estimated storage needs around 300-700 TB. Technical issues to be solved include the implementation appropriate tools for the distributing and searching the data, for post-processing and data extraction and for comparing them with available observations from other archives.

The integration of standard tools from the climate research community (such as ESGF nodes) will be explored. The pilot will be used to demonstrate the integration of existing solutions, still under development, with relevant EUDAT services. The size of the potential user base can be estimated as hundreds of scientists in the climate change and climate impact fields.

Who benefits and how?

The pilot will serve as an essential tool for storage, data analysis and scientific investigation for the participants in the Climate SPHINX PRACE project, including a team of researchers from CNR-ISAC (Coordinator) and Oxford University. The simulation data and the services made available in this pilot will be shared with the participants in the PRIMAVERA and CRESCENDO projects (2015-2019).

The results of this project will also integrate with several other efforts currently underway. Climate SPHINX and this pilot will leverage ground-breaking past initiatives in the pioneering use of HPC for climate simulations such as the UPSCALE PRACE and the ATHENA (High Resolution Global Climate Simulations) projects and the HiresCLIM PRACE Tier-0 project (exploring historical runs and seasonal and multi-annual predictions at T359L31 and T511L91 with the ARPEGE and EC-Earth models). The pilot will also allow to share and analyse efficiently the data in the framework of collaboration, currently underway, with the Center for Ocean-Land-Atmosphere studies (COLA) at



George Mason University, USA.

The data repository, data sharing and staging services offered by the pilot will be crucial to allow a wide user base to have access to a set of climate variables at high temporal resolution and at extremely high spatial resolutions, something that is not commonly available now. Potential users include both climate scientists and researchers from a wide range of fields, studying the impacts of climate change and of extremes on topics such as ecosystems, floods, landslides, fires.

Technical Implementation

A THREDDS Data Server has been deployed at CINECA, providing access to data from the CLIMATE SPHINX PRACE project model outputs. The data has been transferred, using B2STAGE/GridFtp from the original production machine (SuperMUC at LRZ/Germany) to CINECA.

The pilot will, in the future, expose stored data using an ESGF (Earth Science Grid Federation) node. The first implementation tests of an ESGF node have been performed by ISAC. This pilot is currently discussing, in the framework of EUDAT, how to expose the ESGF instance through B2FIND for improving data discoverability. The possibility to register data sets either through DOI or PID will be investigated and contacts have been started with DKRZ to explore the best options.

Preliminary Results

Already in its current form, the platform, which has been deployed, is allowing an extensive exploration of the dataset stored in the archive. A scientific paper on the results of the Climate SPHINX PRACE project and on the role of stochastic physics in high-resolution climate simulations (Davini et al. 2016) has already been submitted and prepared mainly using these facilities. The scientific community using these facilities is growing and includes researchers at CNR, Oxford University (UK), KNMI (NL), COLA (US), ENS-LMD (Fr).

The platform is also serving as a basis to develop new automatic postprocessing tools for climate data, which will also serve as a prototype for tools for analysis and diagnostics of climate data envisioned in the Copernicus C3S-MAGIC project. A further project which is benefitting from this infrastructure is the PRIMAVERA H2020 project, involved mainly in the analysis of high-resolution climate model simulations.

The platform has been proposed for use in a European ERA4CS proposal, devoted to the development of climate services for the Mediterranean region (MEDSCOPE). Finally, the platform is also supporting storage of intermediate model results targeted at the development (tuning) of the next version of the EC-Earth Global Earth System model.

Contacts

- Jost von Hardenberg, ISAC-CNR, j.vonhardenberg@isac.cnr.it
- Giovanni Morelli, CINECA, g.morelli@cineca.it

Further Information

[Scientific Challenges behind the Pilot](#)

To learn more about the EUDAT/Data SPHINX collaboration [visit here](#)

[Read more](#)

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.