

DATA SPHINX

DATA Storage and Preservation of
High resolution climate eXperiments

*Jost von Hardenberg
ISAC-CNR, Torino, Italy*

with Paolo Davini, Susanna Corti, and many others

*EUDAT User Forum, Rome, Italy
3-4 February, 2016*



PILOT GOALS

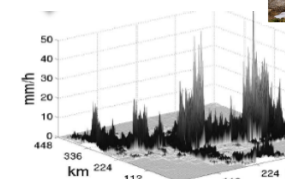
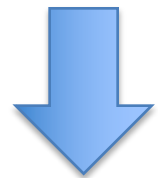
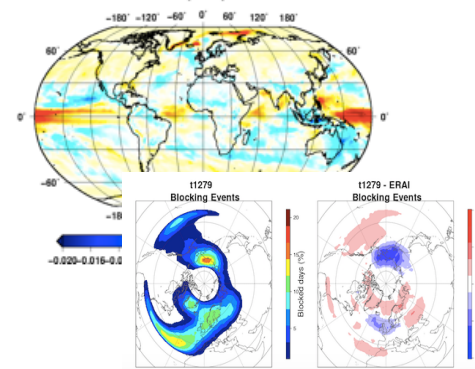
High-resolution global climate modeling has the potential to improve significantly the representation of climate variability, circulation regimes and transport.

Main goal of the pilot:

DATA SPHINX will create a **widely accessible archive of high-resolution climate model outputs** produced through H2020 EU projects (e.g. Primavera, Crescendo) and PRACE grants (e.g. Climate SPHINX) with a particular focus on **climate extremes and the hydrological cycle**.

- **Platform for medium term storage** to facilitate data access and discovery.
- **EUDAT services** used to allow easy and fast access, sharing and analysing efficiently selected variables.
- This will facilitate **scientific collaboration** and will foster research facilitating data analysis and postprocessing.
- Services made available to different **climate research communities**
- Application to the study of **impacts of climate variability** on ecosystems, floods, landslides, fires.

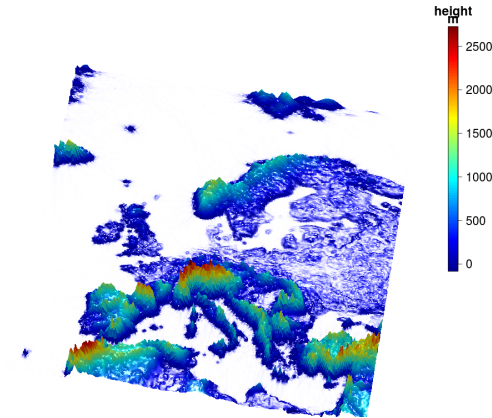
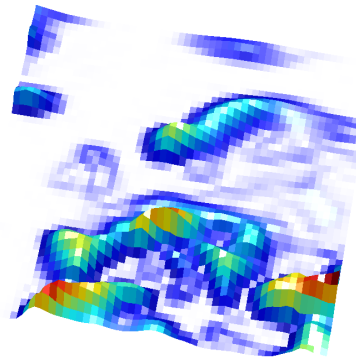
EC-Earth RCP 4.5 - Tot. precipitation trend 2006-2100



CLIMATE SPHINX

Climate SPHINX (Stochastic Physics High Resolution Experiments) is a PRACE EU project (2015-2016) investigating high-resolution climate simulations and the role of stochastic parameterizations

- **20 million of core hours** on **Supermuc** @ LRZ Computing Center, Garching, Germany.
- **About 120 Tb of post-processed data** already available
- More than 110 ensemble members at resolutions from 125 to 16km available.



Model: with additional tuning.

Collaboration between ISAC-CNR (Italy) and Oxford University

CONNECTIONS WITH OTHER HPC PROJECTS AND ARCHIVES



PRIMAVERA H2020 Project (2015-2020)



CRESCENDO H2020 Project (2015-2020)



NextDATA National Project of Interest (2011-2015)



ECMWF special projects (2014-2016)-(2016-2018)



Gauss EXPRESS project (2013-2014)



ECOPOTENTIAL H2020 Project (2015-2019)

PLAN OF THE PILOT



- **Stage1:** Archive at **CINECA** populated with high-resolution simulations with the EC-Earth climate model, mainly from CLIMATE SPHINX integrations. Estimate: 150-300 TB
- **Stage2:** archive 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 and CRESCENDO H2020 projects. Estimate: 300-700TB

Technical goals:

- implementation of tools for distributing and searching the data, for postprocessing and data extraction
- Integration with standard tools from the climate research community (such as ESGF nodes)
- Integration with relevant EUDAT services.



TECHNICAL ISSUES

- Expose the archive through an **ESGF** instance (**Thredds** Data Server)
- Evaluate the possibility to register the data sets either through the DOI or the PID.
- Expose the ESGF instance through **B2FIND** for improving data discoverability
- Evaluate **B2SHARE** as catalogue where to store meta-data records only. The size of the data sets (few GBs) does not permit the use of B2SHARE to store binary objects
- Make part of the archive be accessible to other collaborators via GridFTP (large files) and Thredds
- Implementation of authentication and authorization through **B2ACCESS**

