



**INDIGO - DataCloud**

**WP2**

Definition of Support to Research  
Communities

**Preliminary input regarding the climate domain**

Sandro Fiore  
[sandro.fiore@cmcc.it](mailto:sandro.fiore@cmcc.it)

Advanced Scientific Computing Division

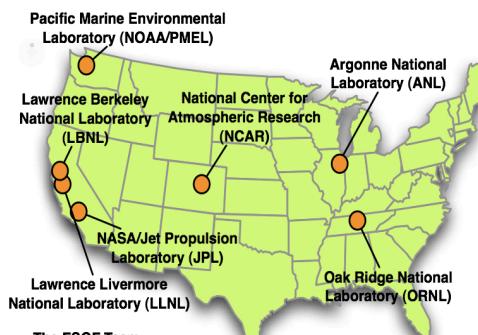
*Euro Mediterranean Center on Climate Change  
(CMCC)*



# ENES & IS-ENES/ESGF eco-system

## II CMIP5 Federated Archive

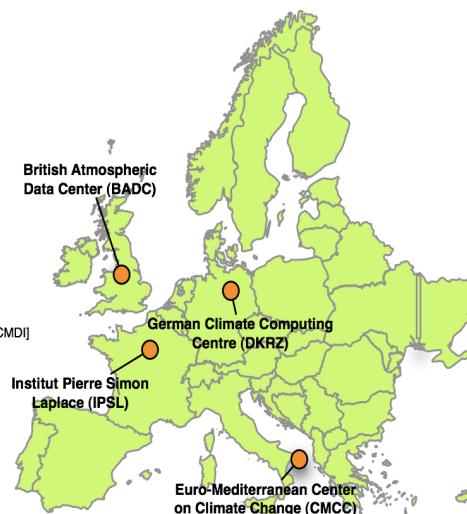
Summary	
Modeling centers	26
Models	51
Experiments	98
Data nodes	20
P2P Index	10
Datasets	44162
Size	1,754.33 TB
Files	3,095,405



The ESGF Team

- Gavin M. Bell, Bob Drach, Charles Doutriaux, Renata McCoy, Dean Williams [LLNL/PCMDI]
- Luca Cincunini, Dan Crichton, Chris Mattmann [NASA/JPL]
- John Harney, Galen Shipman, Feiyi Wang, [ORNL]
- Roland Schweitzer [NOAA/PMEL]
- Rachana Ananthkrishnan, Neill Miller, Ian Foster [ANL]
- Estanti Gonzales, Stephen Kindermann [DKRZ]
- Philip Kershaw, Stephen Pascoe [BADC]
- Sandro Fiore, Giovanni Aloisio [CMCC]
- et al

The ESGF Sponsors



5

Model	# Datasets	Replica	Size (GB)	# Files
CMCC-CESM	48		694.53	9,942
CMCC-CM	519		91,088.84	130,148
CMCC-CMS	46		1,728.73	8,595



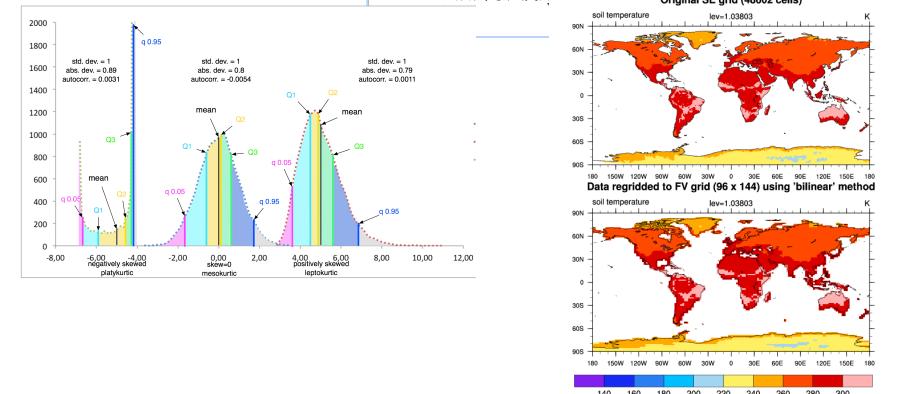
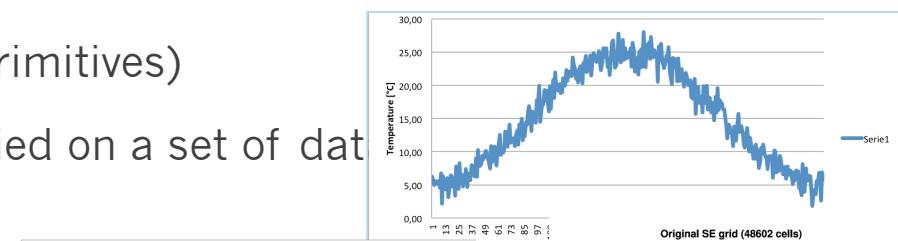
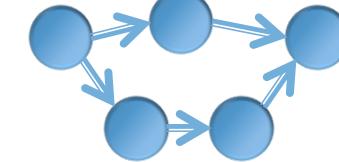
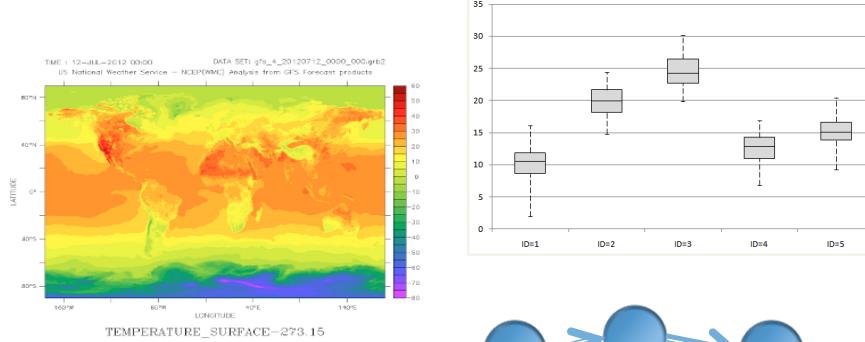


INDIGO - DataCloud

# Some preliminary requirements

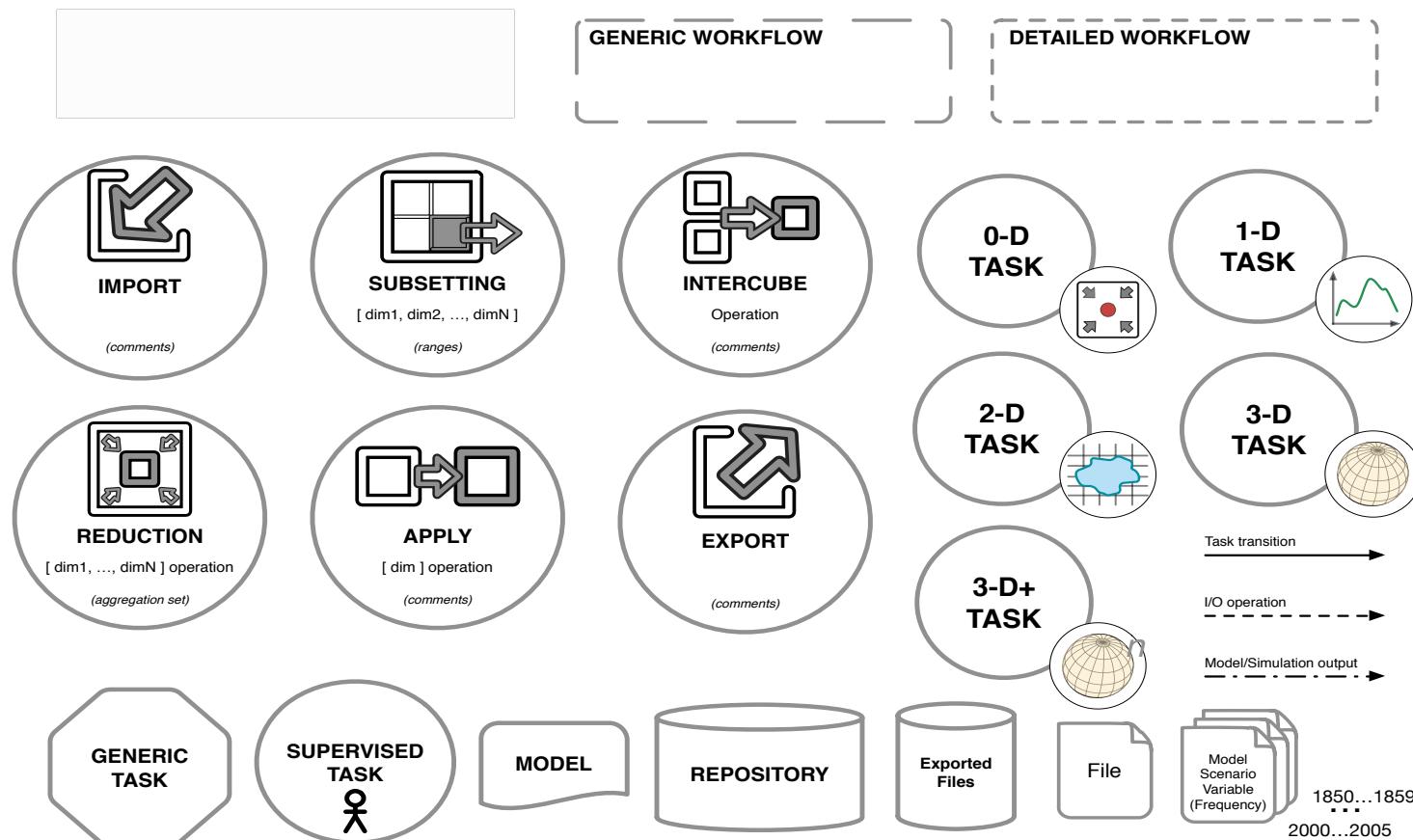
Requirements from the ENES community:

- Time series analysis
- Data subsetting
- Model intercomparison
- Multimodel means
- Massive data reduction
- Data transformation (through array-based primitives)
- Param. Sweep experiments (same task applied on a set of data)
- Climate change signal
- Maps generation
- Ensemble analysis
- Data analytics workflow support
- But also...
- Performance, re-usability, extensibility





# Data Analytics Modeling language





INDIGO - DataCloud

# Example of data analytics use case

