



EUDAT

Support to scientific research on seasonal-to-decadal climate and air quality modelling



INSTITUT CATALÀ
DE CIÈNCIES DEL
CLIMA



**Barcelona
Supercomputing
Center**

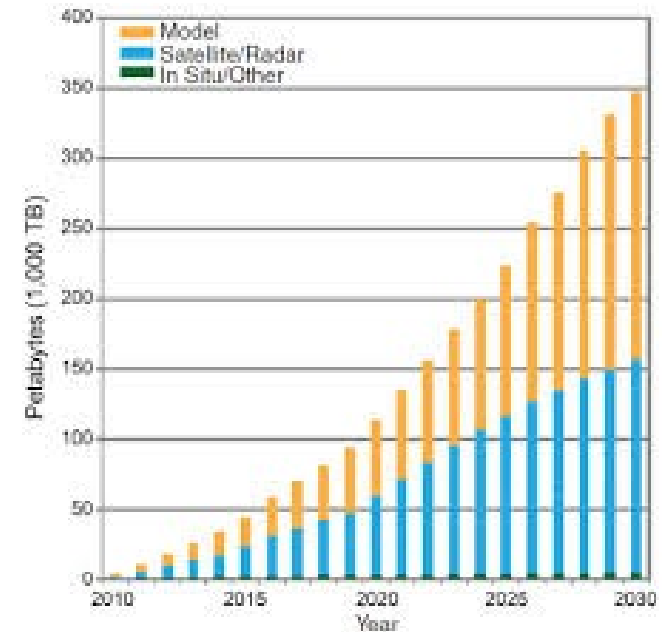
Centro Nacional
de Supercomputación

Pierre-Antoine Bretonnière
Francesco Benincasa
IC3-BSC - Spain

pierre-antoine.bretonniere@ic3.cat,
francesco.benincasa@bsc.es

3-4 February 2015, Rome, Italy

- Cope with exponential growth of data volume because of:
 - increase of spatial and temporal model and instrument resolutions
 - larger variety of data sources: observations, model outputs, reanalysis, sensors,...



Source: Overpeck et al. (2011), "Climate Data Challenges" in the 21st Century, Science

One factor: increase in model resolution

Air Quality Model (NMMB/BSC-CTM)

	Horizontal resolution (grid cell size)	Output size of one year of 48h daily forecasts, global fields (including meteorology, aerosols and gas-phase chemistry)
Standard Resolution	10 km	4.6 PB
High Resolution	4 km	18.2 PB
Ultra High Resolution	1 km	73 PB

Climate Model (EC-Earth)

	Horizontal resolution (atmosphere/ocean)	Output size of a decadal climate prediction experiment (6000 years of simulation)
Standard Resolution	T255/ORCA1 60km/100km	26x72.000 GB
High Resolution	T511/ORCA025 40km/25km	120.72.000 GB
Ultra High Resolution	T1279/ORCA012 25km/12km	1x72.000 TB

Another factor: increase in the number of modelling institutions involved in Climate Model Intercomparison Projects (MIP)

	CMIP (1996)	CMIP2 (1997)	CMIP3 (2005)	CMIP5 (2010)
Number of experiments	1	2	12	110
Centers participating	16	18	15	24
Number of models	19	24	21	45
Total dataset size	1GB	540GB	36TB	3.3PB

Potential consequences without a solution



Scientists and users have to live with problems like:

- Having data “stuck” locally and difficult to share among institutions
- Data repositories too big to be indexed/explored
- Softwares demanding unrealistically large amounts of memory to efficiently compute diagnostics, metrics and products

Why EUDAT services



- Need for new tools for data management to tackle weather, climate and air quality issues inside and across institutions



- Unique opportunity for sharing data knowledge with other communities
- Strong pressure from a large user community (urban air quality, climate-change adaptation, industry sensitive to environmental pressures)



Specific benefits expected from EUDAT services



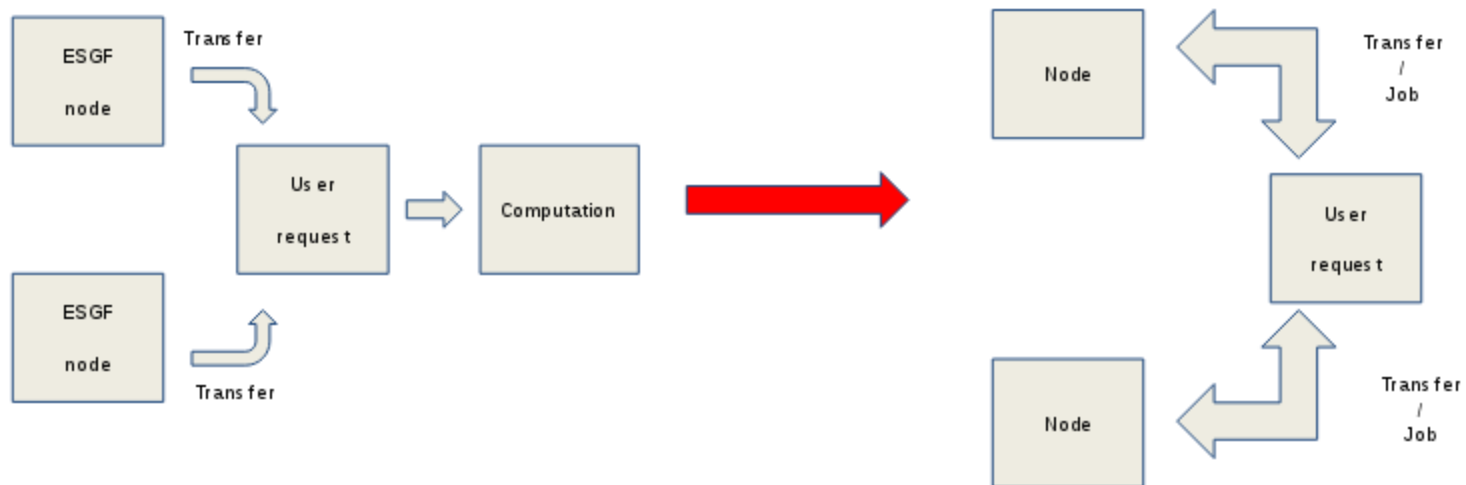
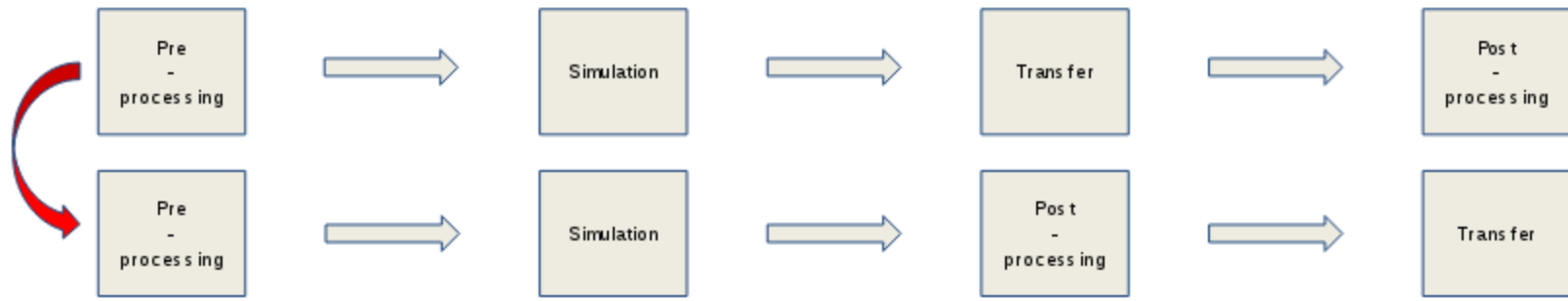
- Increase velocity and efficiency (B2SHARE) in data transfers
- Develop “ESGF-type” innovative solutions for data indexing and discovery (B2FIND)
- Benchmarking with tools used traditionally by the community (gridftp, globus,...)

The screenshot shows the ESGF website interface. At the top, there are logos for ESGF (Earth System Grid Federation), PCMDI, and the University of Reading. A navigation bar includes links for Home, Search, Tools, Login, and Help. Below this, there are sections for 'Current Selections' with links to remove all, project specs, and experiment family Decadal. A 'Search Categories' sidebar lists various filters like Project, Institute, Model, SubModel, Instrument, Experiment Family, Experiment, SubExperiment, Time Frequency, Product, Realm, and Variable, with counts for each. The main search area contains a search box, a search button, and a search bar with a magnifying glass icon. Below the search bar, there are checkboxes for 'Search All Sites', 'Show All Replicas', and 'Show All Versions'. The search results are displayed as a list of entries, each with a link to the data, the data node, version, and options to add to the cart. The results shown are for 'specs_output_IPSL_IPSL-CM5A-LR_decadal_S19610101_day_atmos_day_clt_r1i1p1' and similar entries for different time steps (r2i1p1, r3i1p1, r2i1n1).

Specific benefits expected from EUDAT services



- “Bring the compute to the data”: improve the data workflow



- Interest of our pilot to other communities in our scientific domain:
 - The data transfer and replica issues found in the Earth Sciences community are common to many communities having to share data. Even if the indexing and file organization are very specific to the community (variables, models,...) some solutions could be easily extrapolated to other types of data management procedures.
- Further expected support from EUDAT solving the problem/s we may envisage in the future:
 - The development of generic tools for data transfer and staging and research for cross-community tools by EUDAT can allow to think “out of the box” and transfer innovative solutions to our specific domain: eg management of tapes and disks to respond timely to a wide range of requests.



INSTITUT CATALÀ
DE CIÈNCIES DEL
CLIMA



**Barcelona
Supercomputing
Center**
*Centro Nacional
de Supercomputación*



Thank you!

Pierre-Antoine Bretonnière - Francesco Benincasa
IC3-BSC - Spain

pierre-antoine.bretonniere@ic3.cat

francesco.benincasa@bsc.es

User Forum

3-4 February 2016, Rome, Italy