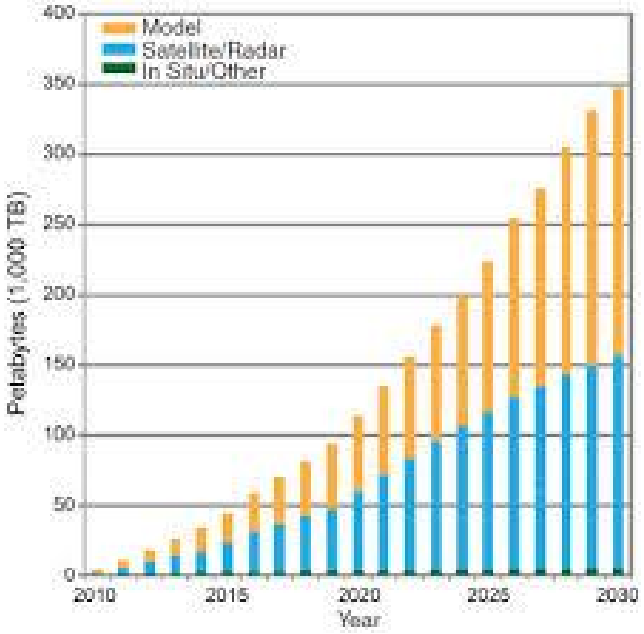
Support to scientific research on seasonal-to-decadal climate and air quality modelling

User Forum
3-4 February 2016, Rome, Italy

Pierre-Antoine Bretonnière
Francesco Benincasa
IC3-BSC - Spain
pierre-antoine.bretonniere@ic3.cat,
francesco.benincasa@bsc.es

Scientific challenges

- 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

Scientific challenges

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

Scientific challenges

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

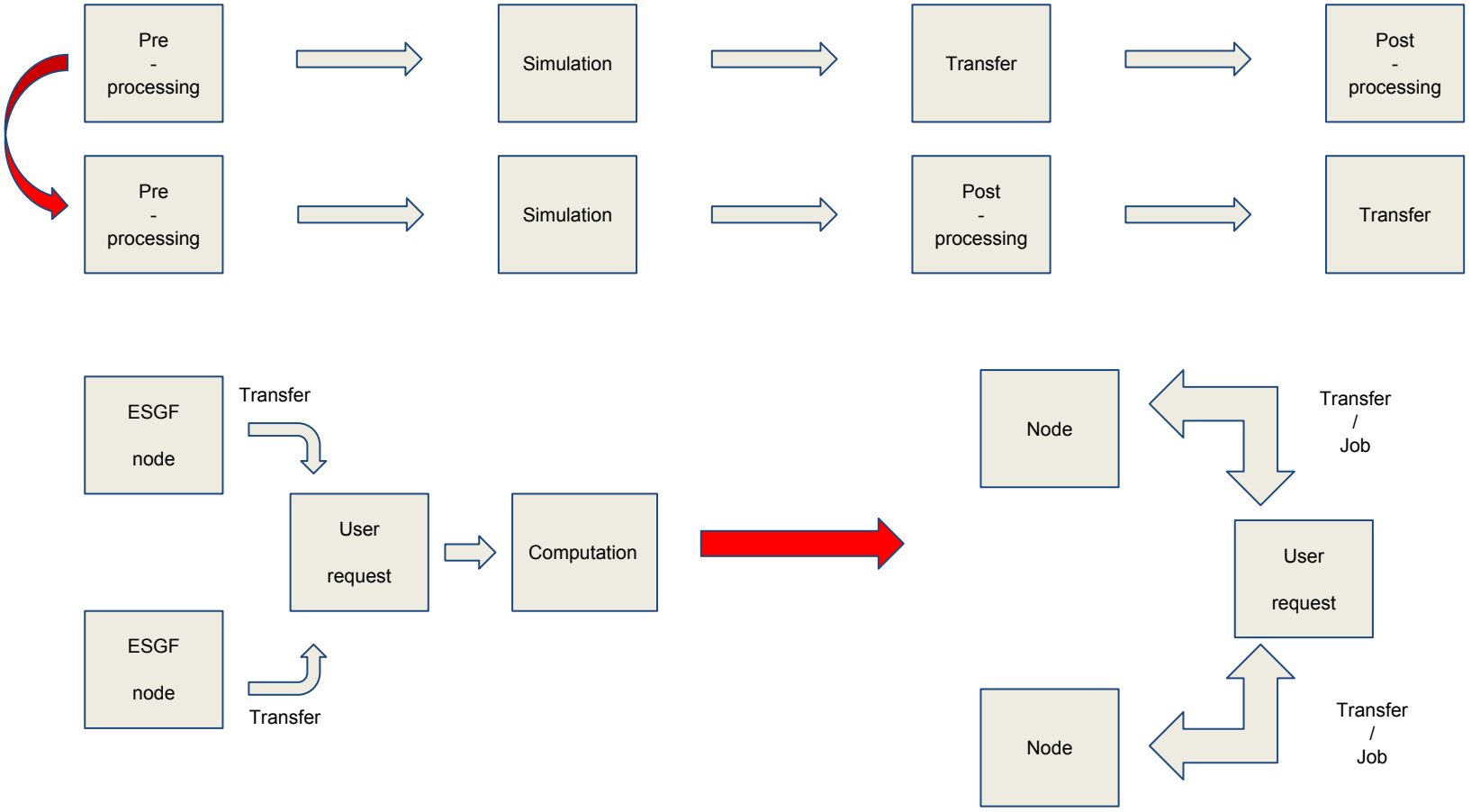
- Unique opportunity for sharing data knowledge with other communities
- Strong pressure from a large user community (urban air quality, climate-change adaptation, industry sensitive to environmental pressures)

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,...)

Specific benefits expected from EUDAT services

- "Bring the compute to the data": improve the data workflow



Expected future impact

- 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

EUDAT

Thank you!

User Forum
3-4 February 2016, Rome, Italy

Pierre-Antoine Bretonnière - Francesco Benincasa
IC3-BSC - Spain
pierre-antoine.bretonniere@ic3.cat francesco.benincasa@bsc.es