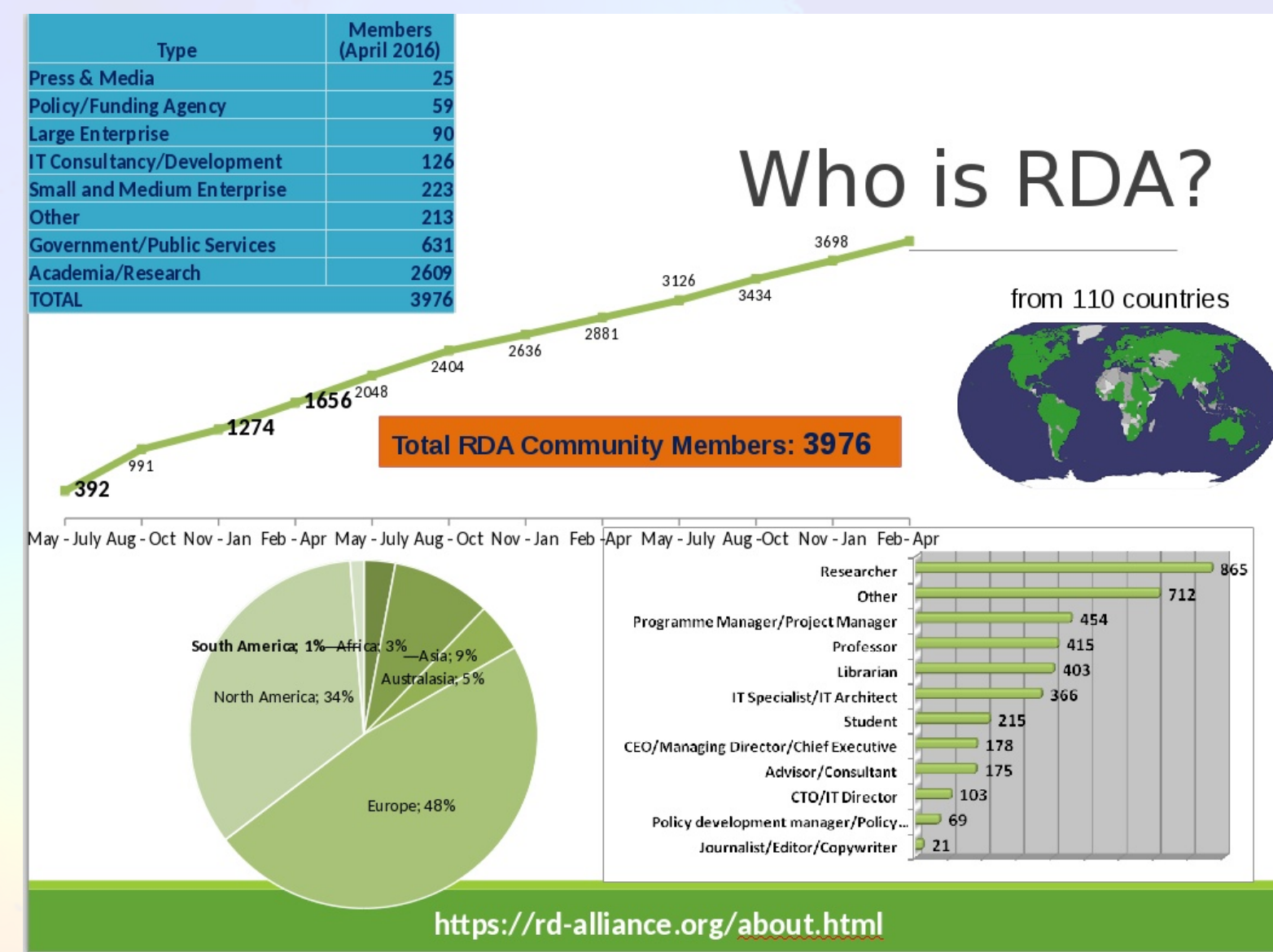


Francesco Benincasa, Pierre-Antoine Bretonnière
Earth Sciences Department - Barcelona Supercomputing Center, Spain

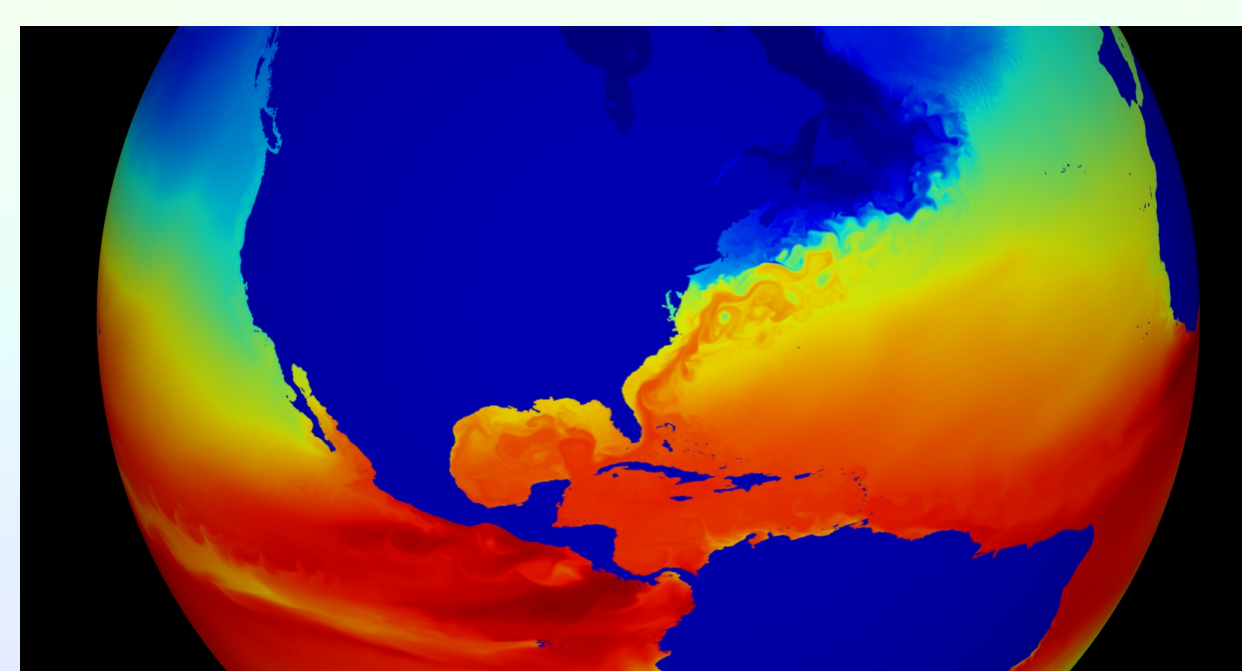
1. What is RDA



- Community-driven organization launched by institutions like the European Commission and US Government's National Science Foundation
- Goal of building the social and technical infrastructure to enable open sharing of data
- Promote data-sharing and data-driven research
- Based on Working and Interest Groups

2. The (Big) Data challenge

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



NEMO model ocean SST simulation at 12 km 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

One factor: increase in model horizontal resolution

Scientists and users have to approach several types of problems:

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

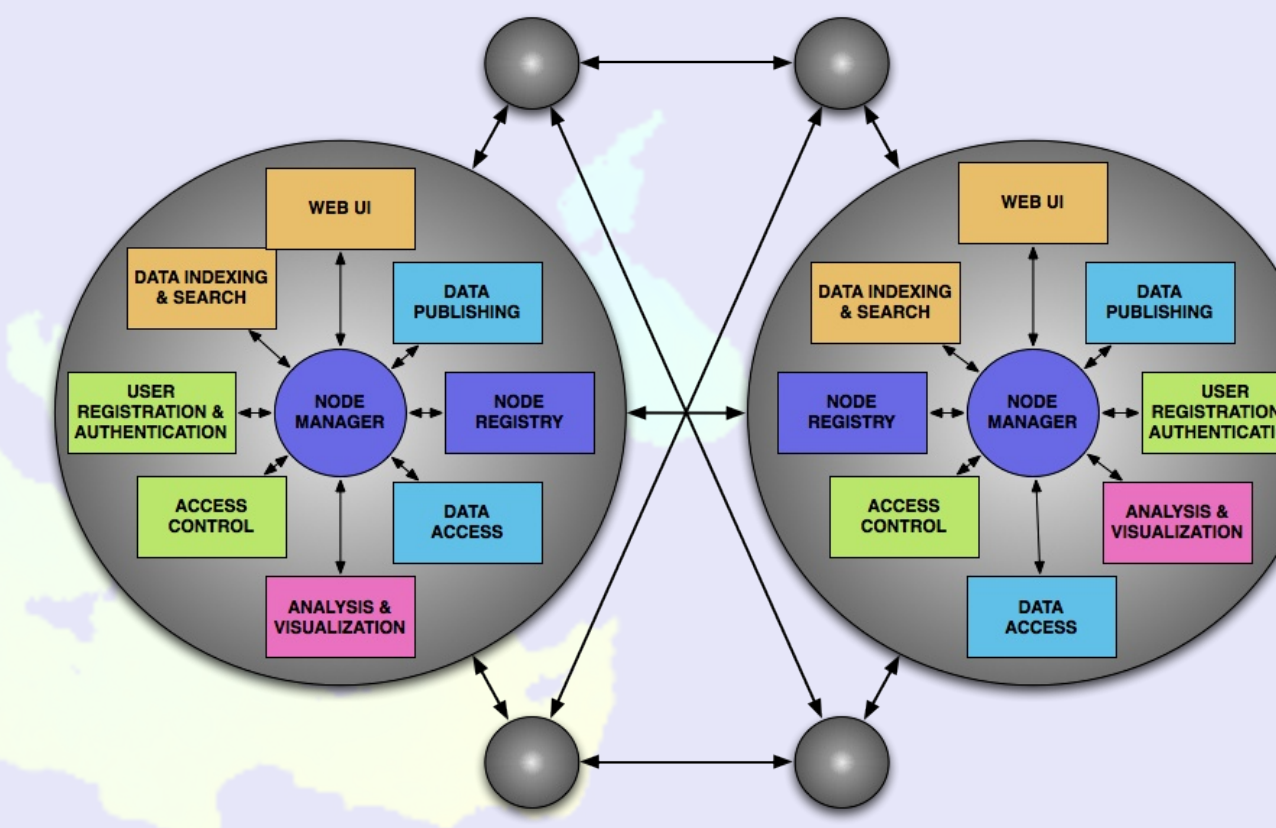
	CMIP (1996)	CMIP2 (1997)	CMIP3 (2005)	CMIP5 (2010)
Number of experiments	1	2	12	110
Centers participating	16	18	15	24
Number of models	19	24	21	45
Total dataset size	1GB	540GB	36TB	3.3PB

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

3. The Interest Group (IG)

Objectives

- Explore and discuss the challenges for the use and efficient analysis of large and diverse datasets from the weather, climate and air quality communities
- Foster the collaboration among research institutes, taking into account inputs from private (renewable energies, satellite and agriculture sectors for example) and public sectors
- Suggest practical and applicable solutions for Big Data issues encountered by these communities, both at technological and policy level



Earth System Grid Federation architecture

Kick-off meeting in Barcelona

- Held on 11th of February 2015 at the Barcelona Supercomputing Center (BSC)
- 25 people from different research institutes, private and public companies and governmental organizations
- Discussion about the importance and opportunity to create the IG
- Speakers from BSC, JRC (Joint Research Centre), ENEA (Italian National Agency), Cantabria University



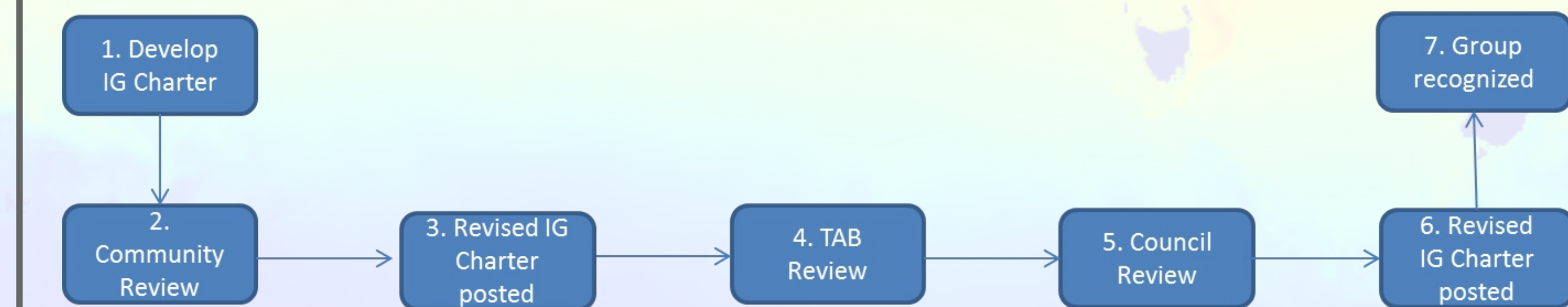
The MareNostrum supercomputer at the Barcelona Supercomputing Center

Birds of a Feather

- Held at 7th RDA Plenary Meeting in Tokyo on 1-3 March 2016
- Around 15 people attended the BoF, from different branches of the Earth Sciences (climate, air quality, climate, services, ...) and from several institutes around the world

Main conclusions

- Even if storage is cheap, it has to be taken into account not only about centralized data repositories, but also about normal storage used by scientists where they download data
- The conclusion that data should be brought to the computing has been emphasized, the actual need is not only data centres but "data analysis centres". In the same way as there are international calls for HPC computing hours, there should be also calls to get "data analysis hours"
- A correct selection about reproducibility of data and which data should be kept and which simulation can be re-executed
- Metadata standards should be defined, several sets can be used: one general and cross community (following the work of the metadata IG) and one at a community level
- Finally, the general opinion was that such IG would be useful to discuss data issues specific of our domain(s) and gather people from these sectors



Steps for the creation of the Interest Group

Next steps

- Those interested please subscribe to the mailing-list: earthsciences-rda-ig@bsc.es
- IG case statement review by RDA community
- Selection of members and co-chairs
- Set up administrative RDA work
- Participate to Barcelona 9th Plenary on April 2017!



Contacts

1. Authors
francesco.benincasa@bsc.es
pierre-antoine.bretonniere@bsc.es
2. Mailing-list
earthsciences-rda-ig@bsc.es
3. RDA
<http://rd-alliance.org>
4. BSC Earth Sciences Department
<http://www.bsc.es/earth-sciences>