

Big Data and Extreme Scale Computing Workshop
February 26-28, 2014 – Fukuoka Japan



PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

PRACE: Expression of Interest Big Data

Dr. Sergi Girona
Managing Director and Chair of the PRACE Board of Directors

PRACE → *the* European HPC Research Infrastructure

- Enabling **world-class science** through large scale simulations
- Providing **HPC services** on **leading edge capability** systems
- Operating as a **single entity** to give access to **world-wide supercomputers**
- **Attract, train** and **retain** competences
- **Lead the integration** of a highly effective **HPC ecosystem**
- Offering its resources through a **single** and **fair** pan-European **peer review process** to **academia** and industry

April, 23rd 2010 creation of the legal entity (AISBL) PRACE with seat location in Brussels, Belgium



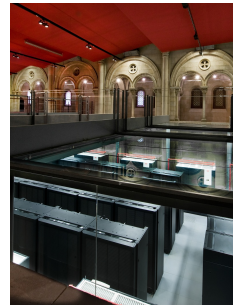
- **25** members, since 2010
- **6** supercomputers in **4** hosting countries
- Research and industrial access
- Nearly **15 Pflop/s**

- **8 billion** hours granted since 2010
- **1.2 billion** hours awarded per call since Nov. 2012
- **303** scientific projects enabled

PRACE's achievements:

In 2013, nearly 15 Pflop/s provided

MareNostrum: IBM IDPX
at BSC, >48 000 cores



JUQUEEN: IBM BlueGene/Q
at GCS partner FZJ,
>458 000 cores

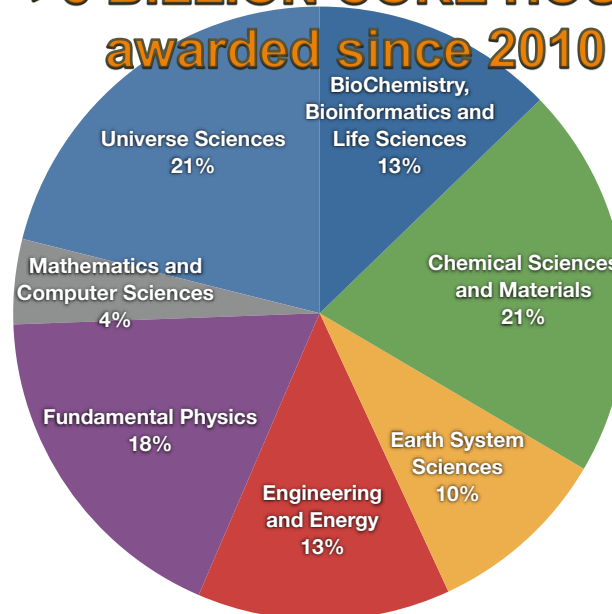


FERMI: IBM BlueGene/Q
at CINECA, >163 000 cores



CURIE: Bull Bullx at
GENCI partner CEA
>90 000 cores.

**>8 BILLION CORE HOURS
awarded since 2010**



SuperMUC: IBM IDPX
at GCS partner LRZ,
>155 000 cores



HERMIT: Cray
at GCS partner HLRS, >113 000 cores

Success Stories

Example 1: Seismology

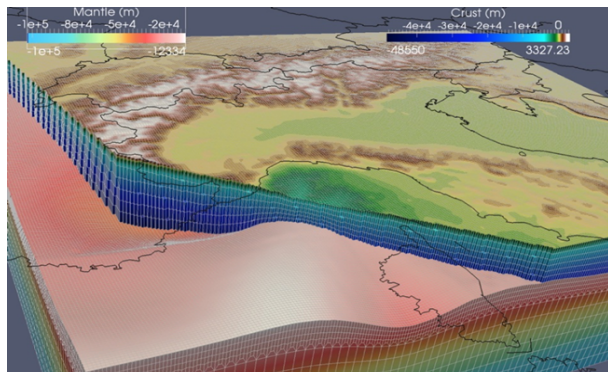
**53,4 million
core hours
on
SuperMUC
(Germany)
for Italy**



The massive allocation of computing resources can be used to **explore the non-linearity involved in the dependence of local ground shaking on geological structure**, by analysing suites of physically consistent, and geologically plausible, models.

Team: Dr. Andrea Morelli – Istituto Nazionale di Geofisica e Vulcanologia, Italy

Goal: Produce an estimate of the **impact of ground shaking on Northern Italy after major earthquakes**.
Provide better foundations for decision-making processes for societal preparedness for earth quakes.



Example 2: Climate

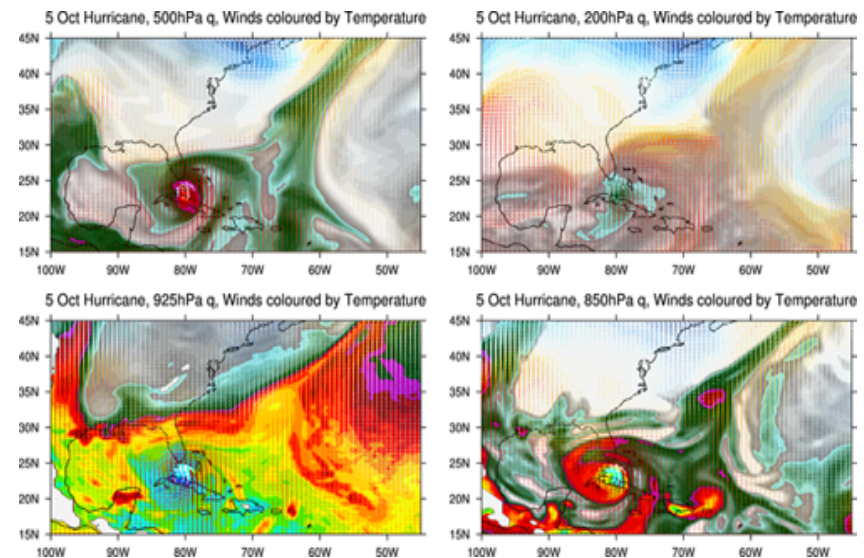
**144 million
core hours
on Hermit
(Germany)
for UK - UB**



PRACE will give to UK **Meteorology** office a 3-year advance in the development of their models.

Team: Prof. Pier Luigi Vidale (NCAS-Climote, Dept of Meteorology, Univ. of Reading and UK Met Office, Exeter, UK)

Goal: to develop high resolution global **weather & climate models (12km)**



Big Data – what about it?

From discussions with Scientific communities PRACE has identified **challenges in the storage, analysis, visualization and transfer of data** connected to large-scale simulations.

PRACE Council in February 2013 decided:

*BoD with the support of the IP project, should prepare an **Expression of Interest regarding Big Data**. This EOI should be broadly distributed for later analysis by the corresponding bodies of the association, for later information to the Council.*


Invitation for Expression of Interest Big Data

- **Scientific justification** and why the handling of data must be closely integrated with the PRACE computing services.
- Expected **amount of data**, **time-frame** for data access and expected computing resources.
- **Management** & use of data.
- **Data open/not open** to the wider scientific community? And how to handle scientific accessibility and usability?
- **Transfer** of data to/from the PRACE resource.

Received Eol Big Data

 **33 proposals** from **8 scientific domains**

 **Earth Sciences and Universe Sciences** have the biggest need for long-term access to data and extensive computing resources

 Storage need in general: **10-200 Tb**, but a few projects **reach the Pb level.**

 **5 years access to stored data.** First with embargo on data, later on **Open Access** for the Scientific community

 A **PRACE Big Data policy** would help to **deploy the full scientific potential** of Tier-0 simulations



33 proposals from 8 scientific domains

Domain	Expressions of Interest
Oceanography/climate/seismology	7 (3/3/1)
Genomics	1
Engineering/CFD	7
Life sciences	2
Universe sciences	10
Plasma	1
Particle physics	1
Chemistry	2

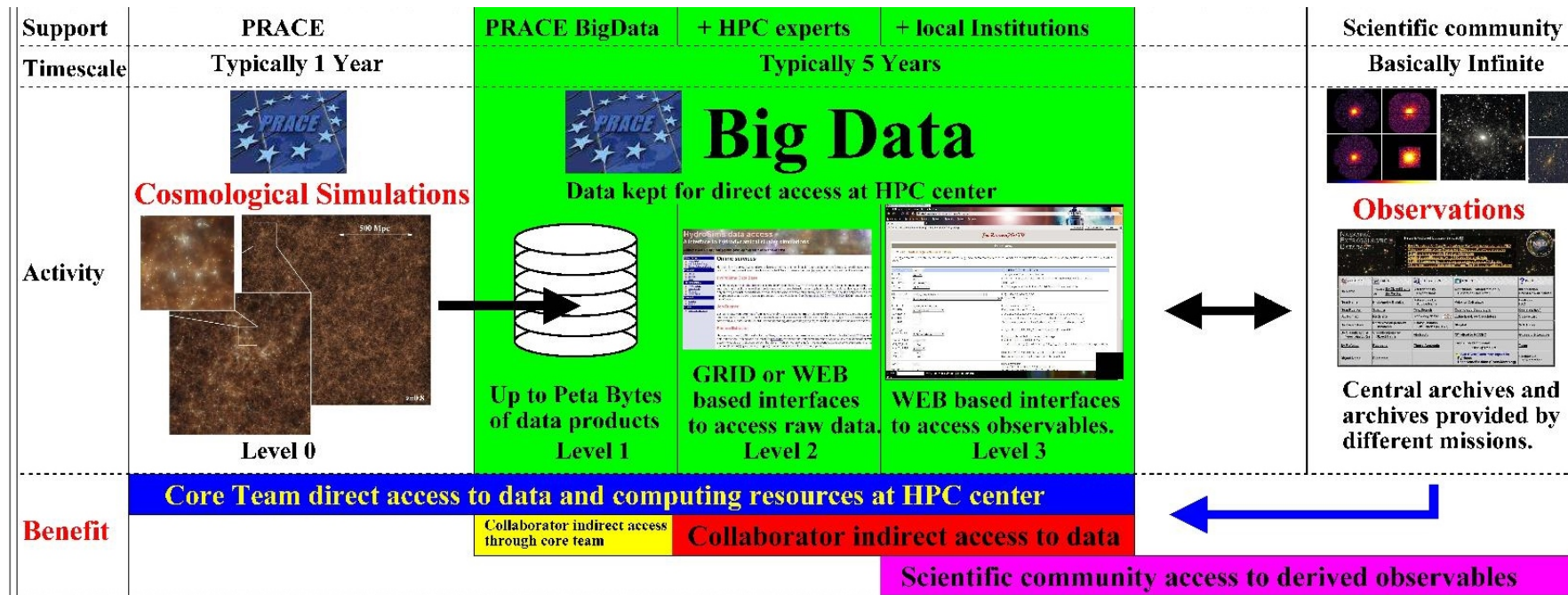


Earth Sciences & Universe Science: needs for long-term access and big resources

- **Universe Science**
“New large scale survey projects (e.g Dark Energy Survey) are moving forward to address cosmic acceleration and the behavior of gravity on the largest scales => **petabyte (PB) scale datasets**. The PRACE computational resources have an important role to play in this.”
- **Earth Sciences/ Climatology**
“Climate integrations need moderate amounts of data for input but produce very large output: Given the output figure of 150---300 GB/1000 core hours, climate jobs filling a 100,000 core machine **would generate 15 to 30TB/hour. We might want to run such jobs for weeks if not months.** “



Access to stored data



Example given by Team Dolag/ Universe science



PRACE Big Data policy

Recommendation of PRACE Scientific Steering Committee (Dec 2013)

PRACE should develop a policy for the handling and access to large datasets arising from the use of PRACE Tier-0 resources.

Decision of PRACE Council (Feb 2014)

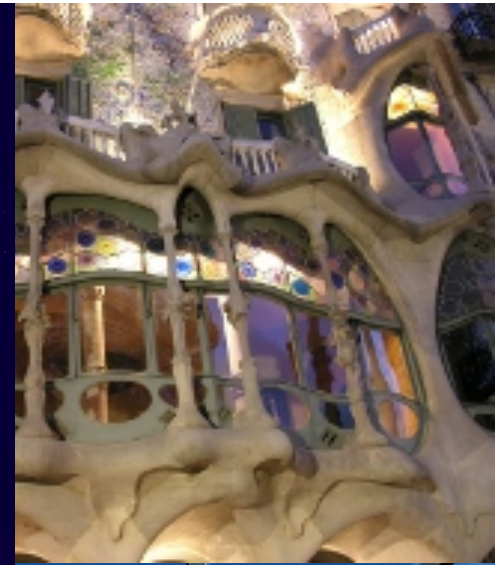
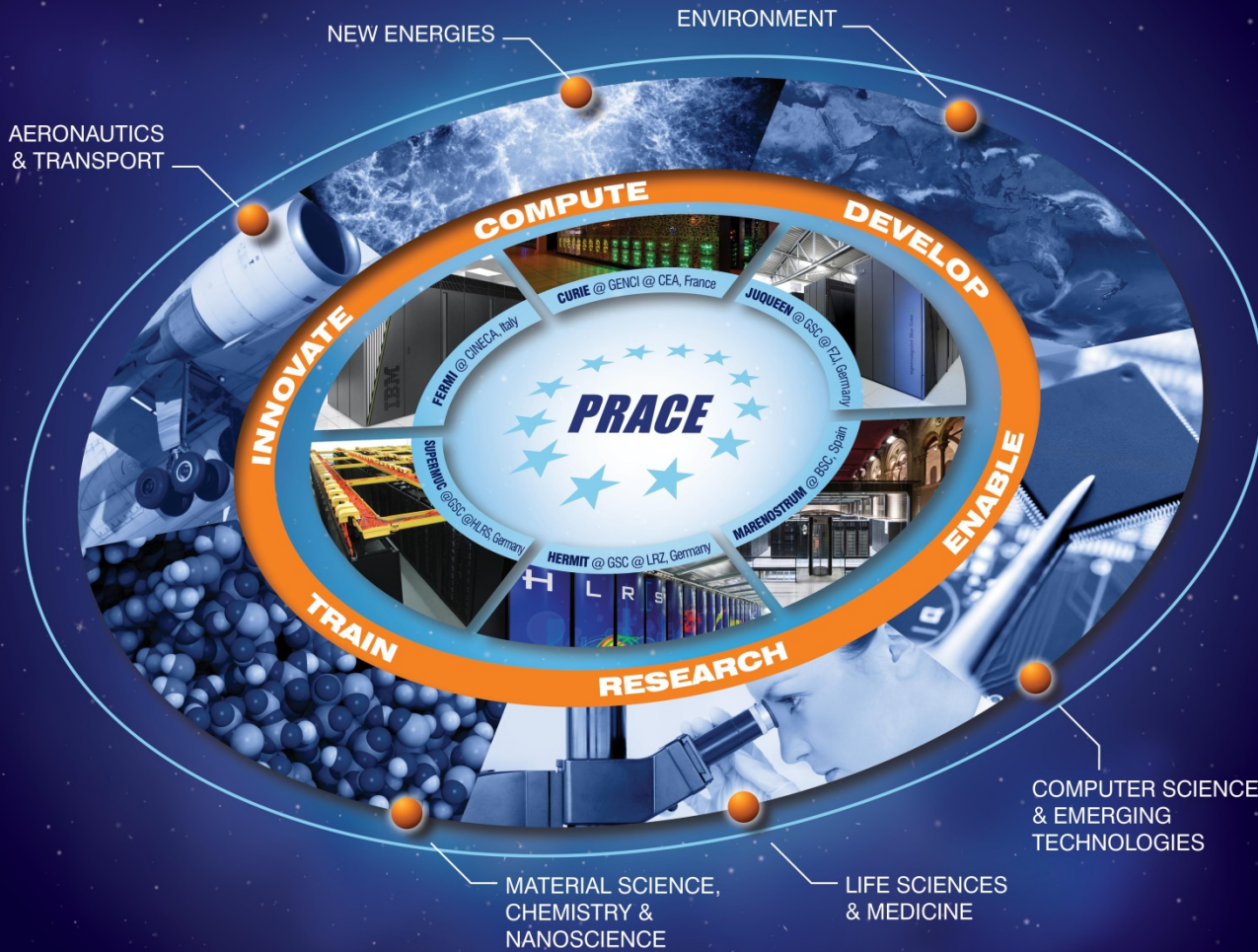
Action-2014-01-04: BoD shall develop a policy for the handling and access to large datasets arising from PRACE Tier-0 resource usage and present it to the Council for approval.



www.prace-ri.eu

HPC for Innovation

WHEN SCIENCE MEETS INDUSTRY



BARCELONA
20 - 22 May 2014





Thanks for your attention

If you have any further questions, don't hesitate to
contact me.

director@prace-ri.eu