



Microsoft® Research

FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS



Microsoft® Research

FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

Cloud computing for science in Europe and
VENUS-C

Fabrizio Gagliardi
Director, MRC EMEA

Cloud Computing for Science

- Cloud computing as an emerging approach to computational science
- Distributed Computing in Europe massively supported by EU FPs over the last 10 years (following from earlier massive investment in HPC and parallel computing)
- Funding in excess of 1 Billion Euros over the last 10 years
- VENUS-C (FP7 Computing Infrastructure 7th call) funded for exploring the complementary/alternative role of industrial cloud computing in the present DCI EU infrastructure

The Cloud

- A model of computation and data storage based on “pay as you go” access to “unlimited” remote data center capabilities
- A cloud infrastructure provides a framework to manage scalable, reliable, on-demand access to applications
- A cloud is the “invisible” backend to many of our mobile applications
- Historical roots in today’s Internet apps and previous DCI computing (Cluster, Grid etc.)

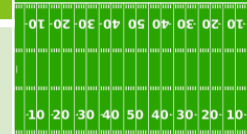


The Cloud is built on massive data centers

Essentially driven by economies of scale

- Approximate costs for a small size center (1K servers) and a larger, 100K server center.

Technology	Cost in small-sized Data Center	Cost in Large Data Center	Ratio
Network	\$95 per Mbps/ Month	\$13 per Mbps/ month	7.1
Storage	\$2.20 per GB/ Month	\$0.40 per GB/ month	5.7
Administrat	140	> 1000	7.1



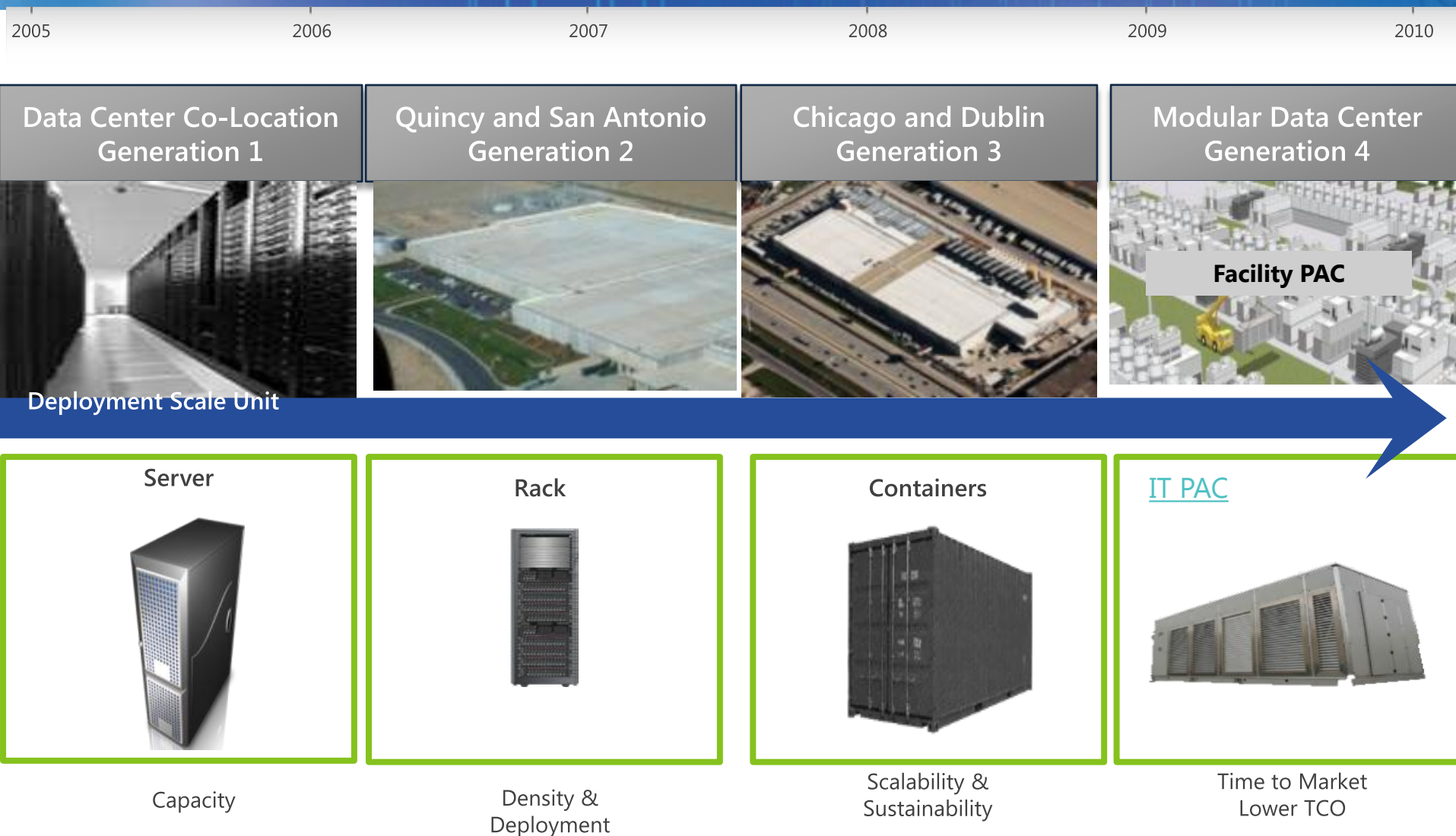
Each data center is
11.5 times
the size of a football field

Major motivations

- Environmental responsibility
 - Managing energy efficiently
 - Adaptive systems management
- Provisioning 100,000 servers
 - Hardware: at most one week after delivery
 - Software: at most a few hours
- Resilience during a blackout/disaster
 - Service rollover for millions of customers
- Software and services
 - End-to-end communication
 - Security, reliability, performance, reliability

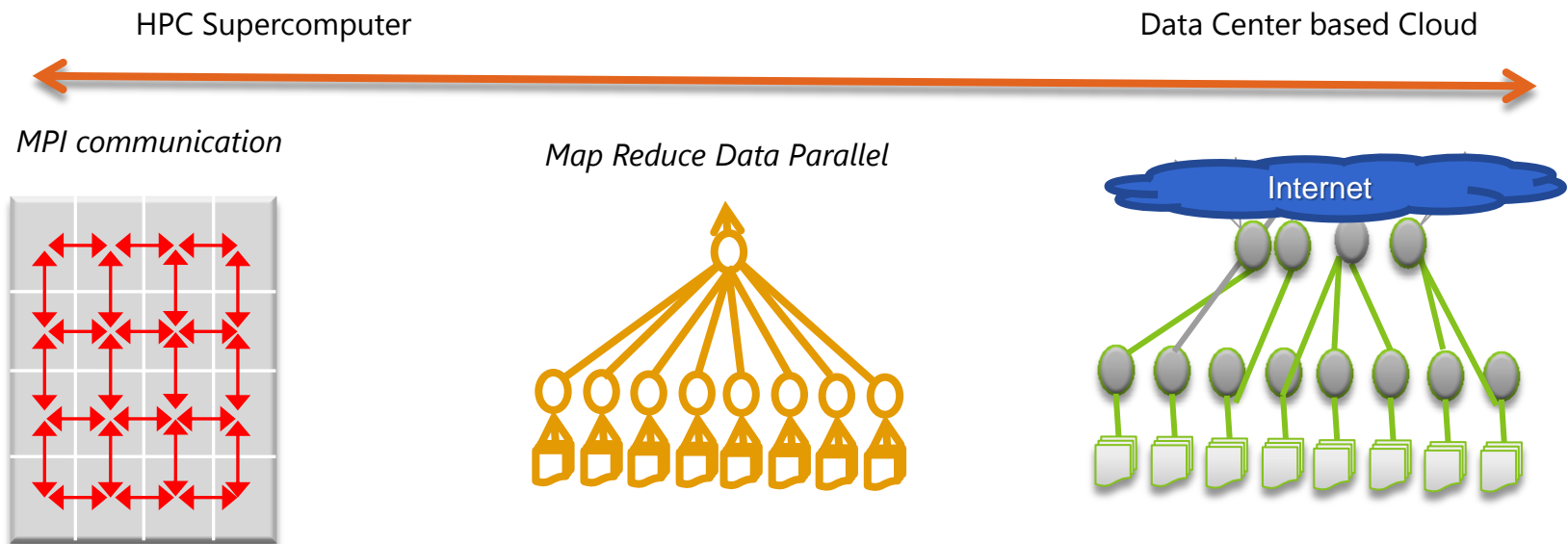


Microsoft's Data Center Evolution And Economics

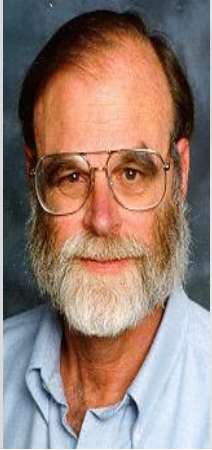


DCs vs Grids, Clusters and Supercomputer Apps

- Supercomputers
 - High parallel, tightly synchronized MPI simulations
- Clusters
 - Gross grain parallelism, single administrative domains
- Grids
 - Job parallelism, throughput computing, heterogeneous administrative domains
- Cloud
 - Scalable, parallel, resilient web services



Orders of Magnitude Always Matter



Tools must empower, not frustrate

These are systemic problems
An insight from Jim Gray ...

A computation task has four characteristic demands:

Networking

Delivering
questions
and answers

Computation

Transforming
information to produce
new information

Data access

Access to information
needed by the
computation

Data storage

Long term storage
of information

The ratios among these *and their* costs are critical

Reaching Out: MSR Azure Research Engagement project

In the U.S.

Memorandum of Understanding with the National Science Foundation (NYT article February 5th, 2010)

- Several projects now starting (http://www.hpcinthecloud.com/hpccloud/2011-04-22/national_science_foundation_partnership_funds_cloud-society_research.html)

In Asia

- Agreement with National Institute of Informatics (NII) in Japan

In Europe

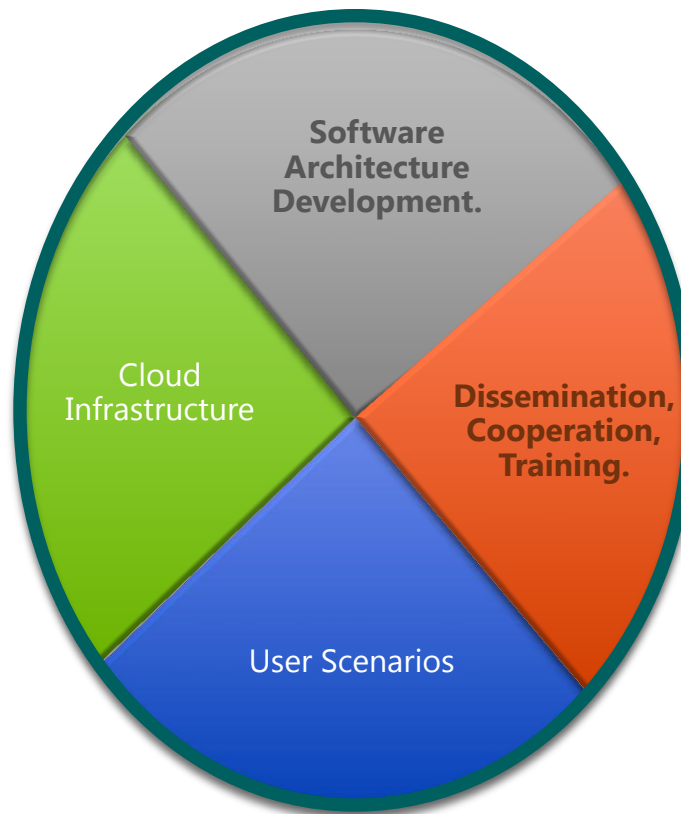
- Direct engagement with science leading institutes in the U.K., France and Germany
- EC engagement in FP7 (VENUS-C)

Virtual multidisciplinary Environments Using Cloud infrastructures

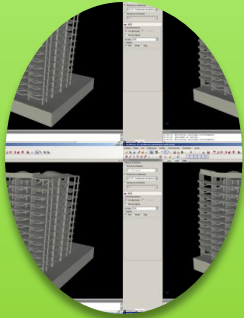
- VENUS-C¹ is developing and deploying a Cloud computing service for research and industry communities in Europe by offering an industrial-quality, service-oriented platform based on virtualisation technologies facilitating a range of research fields through easy deployment of end-user services.
- VENUS-C aims at supporting user communities with the development and deployment of user-friendly services to support the production of successful cloud applications.

(1) VENUS-C is co-funded by the GÉANT and e-Infrastructures Unit, DG Information Society and Media, European Commission. VENUS-C brings together 14 partners from Europe. Microsoft invests in Azure resources and manpower through Redmond and its European research centres.

VENUS-C Partnership



Seven scenarios



Structural
Analysys



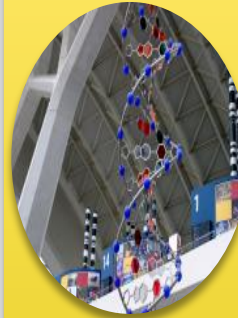
Building Inf.
Manag.



Biodiversity
Aquamaps



Fire Risk
Propagat.



Bioinform.



System
Biology



Drug
Discovery



Open Call for New Pilots

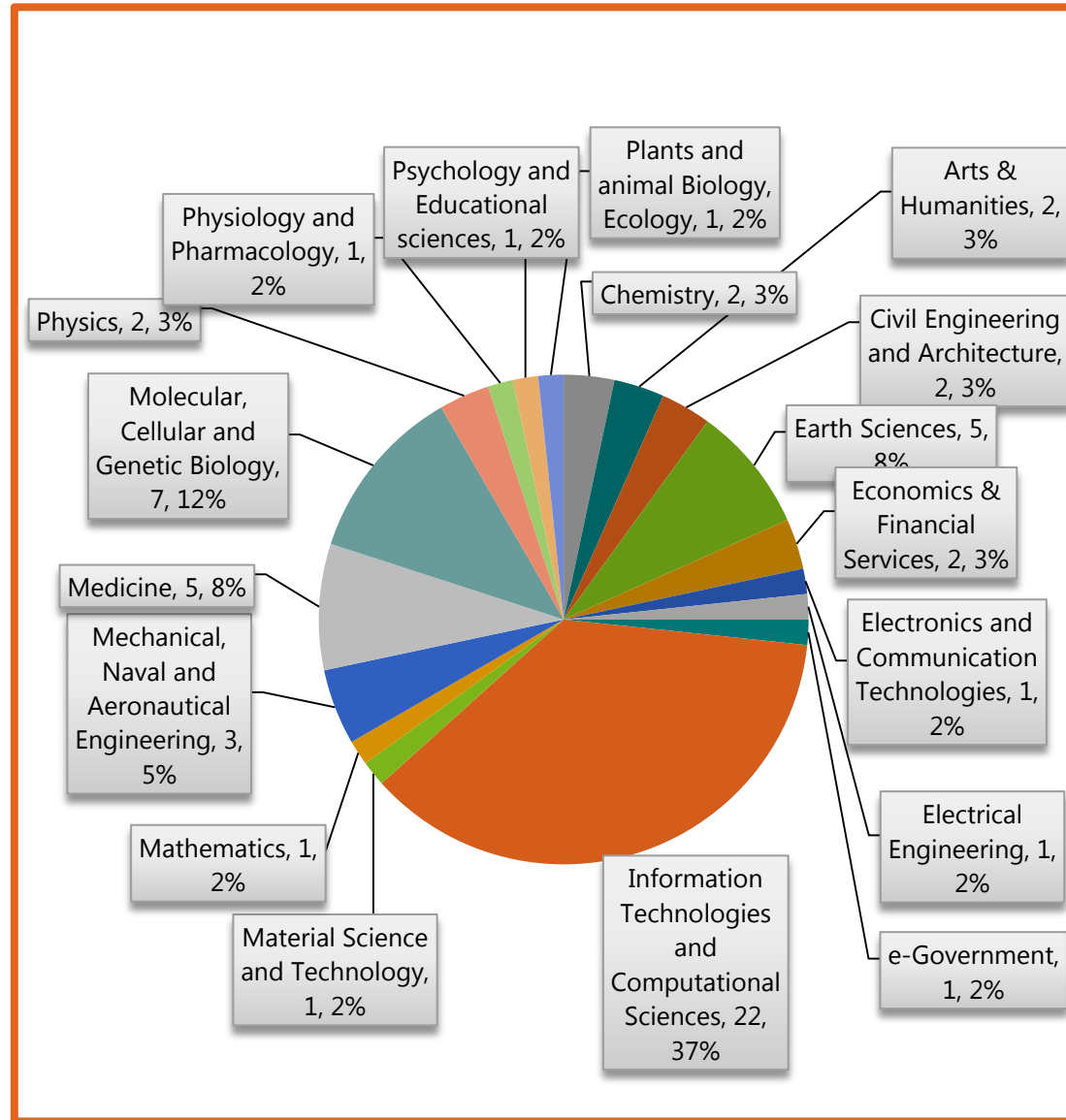
- **Aims: Extend the current scenarios with new pilots leveraging VENUS-C**
 - 400K € for around 15 pilots
 - Access to Microsoft's Azure resources (6 million of CPU hours yearly in total).
 - Access to part of the resources provided by **Barcelona Supercomputing Center, Royal Institute of Technology – Sweden** and Engineering.
 - Support porting of applications and Training.
- **60 valid applications received**
- **Selection and kick-off by June 10th**



Results of the Open Call

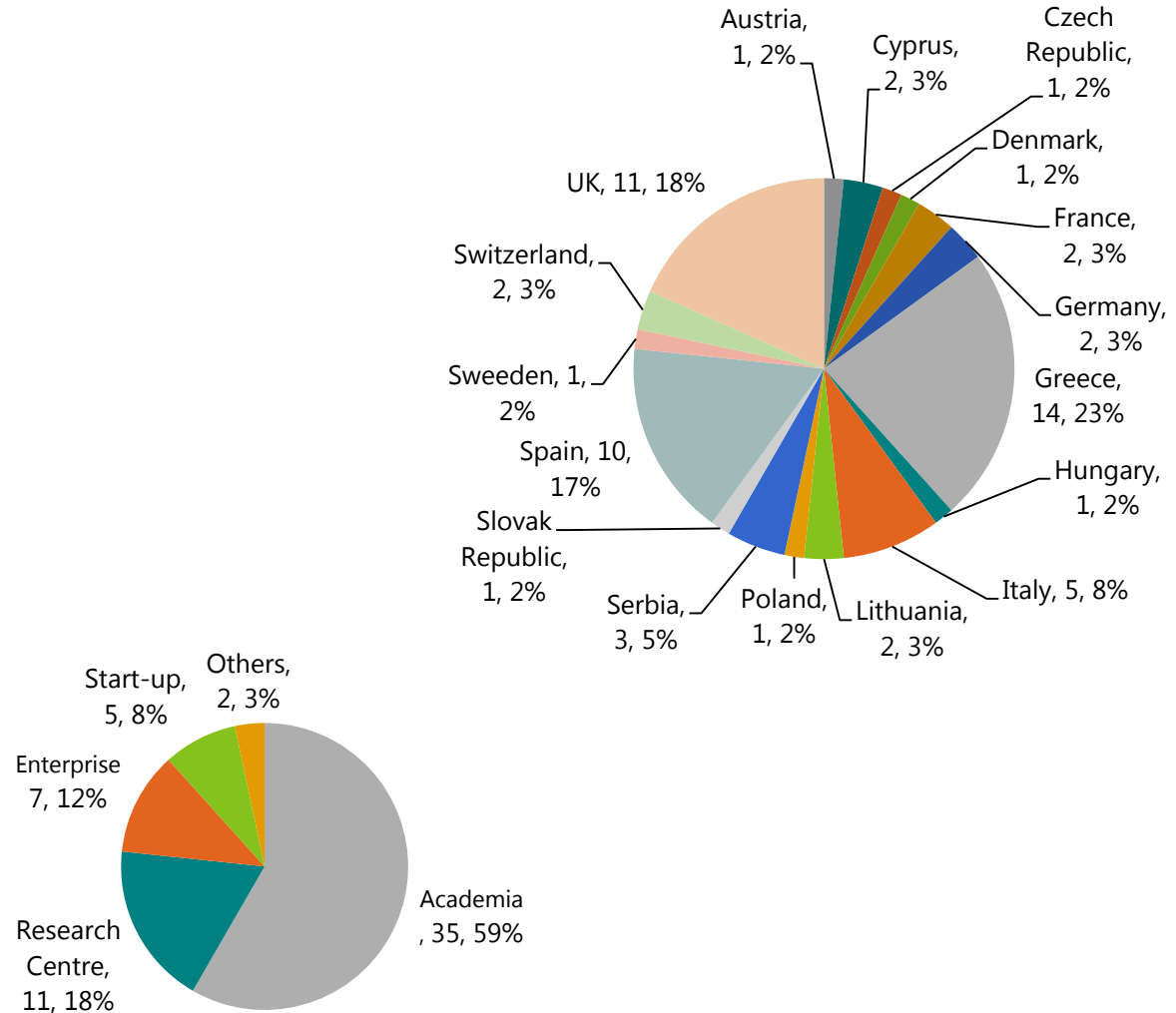
- **60 Submissions from 18 different scientific areas**

- Main areas are ICT, Molecular, Cellular and Genetic Biology and Earth Sciences



Results of VENUS-C Open Call

- 17 Different countries
 - Main areas are ICT, Molecular, Cellular and Genetic Biology and Earth Sciences.
- Mainly from Academia and Research centres but also from some start-ups



Extending VENUS-C to LATAM: **EUBrazilOpenBio**

- **EUBrazilOpenBio** – Open Data and Cloud Computing e-Infrastructure for Biodiversity
 - Proposal submitted under the Objective FP7-ICT-2011-EU-Brazil Research and Development cooperation.
 - A twenty-four month work plan to **deploy an e-Infrastructure of open access resources supporting the needs of the biodiversity scientific community**



OpenBio Consortium



European Partners

- Barcelona Supercomputing Center – Centro Nacional de Supercomputación (BSC).
- Consiglio Nazionale delle Ricerche (CNR).
- Trust IT Services Ltd (TRUS-IT).
- Universitat Politècnica de València (UPVLC).
- Species 2000 (Sp2000).



Brazilian Partners

- Reference Center on Environmental Information (CRIA).
- Recife Center for Advanced Studies and Systems (CESAR).
- Fluminense Federal University (UFF).
- National Education and Research Network (RNP).

MSR in advisory role

OpenBio Activities

Use Case 1: Integration between Regional & Global Taxonomies

A pilot study to analyse and cross-map all or part of the regional plant catalogue of the Brazilian Catalogue of Life (over 30,000 species of flowering plants) with the global plant catalogue served by Sp2000 (with up to 150,000 species of plants)

Analysis of the pilot studies to make requirements and specifications with components and APIs available in consortium from VENUS-C, d4Science, OpenModeller and other ones

Use Case 2: Data usability and the use of ecological niche modelling

A pilot study to concentrate on the distribution of plant species through the use of the tool called openModeller. Models will exploit the compute and storage resources of VENUS-C providers

Testing

To operate and support a production quality computational infrastructure by the integration of existing research infrastructures to promote open accessibility to bio, life science and climatology data and its processing

Software Platform Specification and Development

Re-engineering and adaptation for pilot studies to work according design and specifications components and infrastructure accessible for OpenBio.

Conclusions

- Cloud computing is becoming a powerful enabler for computational science by removing traditional obstacles and lowering the access cost to massive computing and data processing
- Easy of use also for non CS scientists, affordable also for remote institutes in less developed countries
- Good platform for preservation and sharing of scientific data
- Availability of public funding as demonstrated in the recent EU-Brazil joint RfP
- MRC available to advice and support experimental access to this technology
- Happy to talk to interested scientists in LATAM



Thank you for your kind attention
and interest

