www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

The RI ecosystem and data

Cyberinfrastructures (HPC) and data

Fabrizio Gagliardi
Barcelona Supercomputing Center




Disclaimer:

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

These slides reflect only my personal opinion, as a veteran of EU projects and large data and computing RIs

- While at CERN (1975-2005), leader of large data and distributed computing infrastructures such as European Data Grid, EGEE

Now at BSC (2013 -) mostly involved in HPC RIs such as EuroHPC and RDA

Questions to us

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

- What learnings can you share with regard to effective linkage of cyberinfrastructure (HPC) and data to facilitate engagement with trusted, quality data?
- How can RIs and other parts of the research ecosystem better work together?
- What are potential policy implications of this?

BSC

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

- Spanish national supercomputing centre and Severo Ochoa Centre of Excellence in Spain
- Hosting member of PRACE research infrastructure and coordinator of the Spanish Supercomputing Network (Red Española de Supercomputación)
- Internationally recognized research centre employing over 650 members of staff with portfolio of 75 competitively funded projects, mostly from the EU

BSC as an example

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

- BSC is both a HPC service provider and data centre for a number of communities (external and internal)
- Data are both the results of simulations on the HPC platforms and acquired scientific/industrial data
- Complementarity is obvious and data locality essential for many critical applications
- Most of the data are stored locally with high efficiency and trusted, quality data but also high cost
- Data are also accessible and acquired from distributed Ris (Elixir, Copernicus…)

Issues

- Open Access is not uncontrolled access and free for everybody
- Industrial collaborations often involve exclusive access and strict data protection
- Often scientific data have no value outside their community
- Of course data produced with public funding need to be made available to other public entities but…
- No indiscriminate access or public domain…
- Who pays for the data production should own the data…
- Otherwise soon all scientific data in the world will belong to Google…

Sociology and politics

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

- HPC (cyberinfrastructures) and distributed data computing are converging from a pure technology point of view:
  - A supercomputer today is logically very similar to a computer cluster
  - Commodity Technology is predominant
  - Big Data everywhere
- Historically HPC derives from big and expensive mainframes hosted in glass rooms
- While Grids, Clouds and RIs derive from distributed user communities e.g. W-LCG for HEP (Particle Physics community)
- Often competing for the same public funding resources…
- E.g. EuroHPC and EOSC should clearly be synergetic, this is not really the case today… or not enough…
- Major challenge for the funding agencies and public authorities to make these still separate communities to converge and cooperate
- The final end users need to see a coherent environment

www.bsc.es

Thank you!

For further information please contact
fabrizio.gagliardi@bsc.es