

## Portugal advanced computing strategy: Iberia, Europe and the Atlantic

*Responding to the Challenge of Scalable Scientific Data Distribution and Processing in Europe*

May 2018

Europe should address the emerging challenge of **Scalable Scientific Data Distribution and Processing in Europe** by providing the appropriate funding streams to remove the obstacles that currently prevent an effective processing of scientific data, enabling exploitation of the full data potential and **harnessing the capacity of regional investments**.

This requires an innovative data and computing integrated infrastructure that responds to the problem of large scale distributed processing of research data, **unleashing the potential of collaborative science and innovation across Europe**.

Investments at the European level are necessary to provide services to access, process and compute research data in a scalable way at European level, ensuring that **research data is broadly exploitable as a public good**. Funding the synergy between Research Infrastructures (such as ESFRIs and large research collaborations) and scientific computing infrastructures will furthermore **avoid a “digital gap”** between EU countries with different readiness levels in R&D.

The impact of a strategy along these lines should maximize the return of European and national investments in publicly funded computing e-infrastructures at regional level by enabling shared open access via a **flexible federation of distributed interoperable computing resources**.

Actions:

- Address the need of processing research data close to the scientific computing facilities in a scalable way, by creating **economic incentives to the deployment, integration and operation of national and regional computing infrastructures federated at European level**.
- Establish **policies and incentives that promote the usage of these scientific distributed computing infrastructures** by the research communities (ex. funding co-development efforts, training and adoption of common interfaces, standards and best practices)
- Enable the **creation and provisioning of added-value services, based on advanced software technologies and data science** targeting researcher’s needs (scientific portals, processing workflows, virtual research environments, data exploration services) on top of the distributed scientific computing infrastructures;
- Ensure that researchers and entrepreneurs have access to geographically distributed and **specialized computing facilities** including accelerated computing and experimental technologies leading to innovative results and development of highly qualified digital skills;
- Supporting the implementation of testbeds in research performing organizations to act as pilots of **new services and protocols** to speed up scientific data transfer and processing, supporting the emergence of pan-EU networks of innovation clusters, widening the use of advanced computing and data analytics by high-tech SMEs.

## 1. An evolving capacity and collaborative agenda in Iberia

### 1.1 Grid computing: *IBERGRID since 2003*

IBERGRID is a state-of-the-art Iberian digital infrastructure deeply integrated in the global WorldwideLHC Computing Grid, which includes and links about one million CPUs. Ibergrid federates scientific computing datacenters in Portugal and Spain. It was created under the agreement of Scientific and Technological cooperation signed by Spain and Portugal in 2003. This joint strategy has allowed IBERGRID to support large scientific international collaborations, research projects, ESFRIs, etc., optimizing the usage of resources, reliability of services, and fostering at the same time regional synergies. The infrastructure has delivered more than 984 million processing hours since 2006.

R&D activities have been oriented to evolve the provision of services to support researchers requests by adapting new and emerging technologies, and providing more flexible access to resources. The infrastructure has evolved continuously to provide new services and capabilities addressing the user needs.

The strategy and actions outlined above are aimed at fostering this model across Europe building on top of existing initiatives such as EGI, EOSC and Géant. For further information see: <http://www.ibergrid.eu>

### 1.2 Iberian Advanced Computing Network: *IACnet since 2018*

Portugal and Spain are working together to build on IBERGRID and the increasing cooperation established through the wide use of the *Barcelona Supercomputer Centre (BSC)* throughout Spain and Portugal to further develop an “**Iberian Advanced Computing Network - IACnet**”. It should be aimed to better position Spanish and Portuguese scientific and business communities to face emerging opportunities and challenges towards a new generation of High Performance Computing Infrastructure in Europe. It also should **take into account the specificities of both computing approaches**, namely the distributed and the HPC paradigms.

By bringing together the various Spanish and Portuguese advanced computing networks into a totally collaborative network, a major step is being developed towards the advancement of the digital agendas in the two countries and to support processes of digital transformation of our economies and societies.

In parallel, Spain and Portugal will strengthen a collaborative position to face the current discussion in Europe about the evolution of EuroHPC. A flexible management structure for EuroHPC should allow every participating Member State to have a fair voice and where common EU resources are apportioned according to the merit of the projects, so that Spain and Portugal, together, can act with an increasing relevance at an European level.

## **2. EuroHPC supporting a pan-European Computing and Data Innovation ecosystem**

As per the Rome EuroHPC declaration, *the European HPC Infrastructure should include a truly inclusive pan-European dimension leading to a world class Data and Computing Ecosystem of technologies and services (e.g. using cloud paradigms) supporting science, innovation and development of skills.*

In April 2016, as part of its Digitising European Industry strategy, the European Commission adopted the Communication on the **European Cloud Initiative** (ECI). The Communication calls on the European Commission (EC) and Member States (MS) to create a leading **European HPC and Big Data ecosystem**, underpinned by a world-class high performance computing, data and network infrastructure. In the same communication the Commission proposes the development of the **European Science Cloud**.

This is a prerequisite for European competitiveness and the basis for provisioning advanced digital services across borders. The ECI initiative was endorsed by the European Council in June 2016 and further adopted by the European Parliament in February 2017.

On 23 March 2017, during the 60th Anniversary celebrations of the Treaty of Rome, seven Member States (MS) signed **the EuroHPC declaration**: France, Germany, Italy, Luxembourg, the Netherlands, Portugal and Spain. Other MS were invited to sign.

The **EuroHPC declaration and the ECI initiative** call for the creation of a leading-class integrated supercomputing infrastructure as an essential component to exploit the data revolution in Europe and contribute to global growth. The underlying vision is to **establish in Europe an integrated world-class HPC and Big Data (BD) ecosystem** covering all value chain segments (hardware, software, applications, interconnections, skills, services, etc.) based on European leadership in HPC, Cloud and Big Data technologies.

Investing on such ecosystem will upgrade Europe's scientific capabilities and engineering skills and know-how in converging **HPC, Big Data and Cloud Computing technologies**, applications and services. It will permit to strengthen the European technology supply in these areas and secure Europe's own independent HPC technology and system supply.

It will enable Europe to develop the next generation supercomputing systems and acquiring leadership-class systems based on European know-how and technologies, closing the chain from research and development to the delivery of exascale HPC systems.

Making such systems widely accessible and available will stimulate the development and use of the best HPC and data intensive codes in today's and future's most innovative scientific and industrial applications. It will help providing innovative, usable and competitive solutions that satisfy the demands of users from science, industry and notably SMEs, and the public sector; and, it will **underpin the European Open Science Cloud** and contribute to the digitisation of European industry, science and public services.

### **3. Transatlantic cooperation in advanced computing: *MACC and AIR data Net since 2018***

#### **3.1 Building on UT Austin-Portugal Cooperation to build MACC**

Ten years after the first cooperation agreement signed between Portugal and the University of Texas at Austin (as signed in March 2007), the Portuguese Government, through the Minister of Science, Technology and Higher Education and the Science and Technology Foundation (FCT), and the University of Texas at Austin agreed on conditions to expand and strengthen a long-term collaboration focused on advanced and high performance computing, exploring mutual interests in science and technology oriented towards social well-being, job creation, economic growth and quality of life.

It is under this context that the expansion of the UT Austin Portugal International Program Partnership for the period 2018-2022 included the transfer of **twenty racks of the state-of-the-art STAMPEDE 1** computing platform from the Texas Advanced Computing Center (TACC) to FCT, which were decided to be used to launch in Portugal a new center at the University of Minho in a way totally open to research groups throughout Portugal.

This has allowed to launch the **Minho Advanced Computing Center, MACC**, centered on advanced computing activities and emerging forms of collaboration between research groups throughout Portugal with TACC (Texas Advanced Computing Center of the University of Texas at Austin), among other internationally reference centers in advanced computing.

The objectives of the **Minho Advanced Computing Center, MACC**, are as follows:

- a) To reinforce traditional High Performance Computing research areas and enable the emergence of new research domains heavily dependent on High Performance Computing and High Throughput Computing for digital modelling and data science;
- b) To promote and accelerate the structuring and coordination of the national scientific and technological critical mass in advanced computing as a means to draw new scientific and economic activity;
- c) To increase the potential for the creation of economic activity in the domains of data intelligence and other advanced computing areas through scientific driven spin-offs and the attraction of research laboratories from reference international ICT companies.

#### **3.2 Atlantic Interactions: building the “AIR data Net”**

The design of the Atlantic International Research Centre - *AIR Centre* - has made significant progress over the last two years in terms of its goals to promote an integrative approach to space, climate change and energy, earth and ocean science in the Atlantic, fostered by emerging methods of data science, and to be continuously promoted in close alignment with the United Nations’ Sustainable Development Goals (SDGs).

By May 2018, with the organization of the “**Third High-level Industry-Science-Government Dialogue on Atlantic Interactions**”, held in Praia, **Cabo Verde, 7-8 May 2018**, the progress achieved in establishing the AIR Centre with a sound and scientifically relevant R&D agenda on “Atlantic Interactions” considered the definition of a preliminary Scientific Program and the preliminary identification of the first technical initiatives of the AIR Centre. This has been possible together with the setting up of a non-profit

association to promote the AIR Centre, in a way to provide, promote and foster the bases of a truly international scientific agenda, organization and workforce.

The initial Scientific Program for the AIR Centre considers the following topics, without prejudicing their re-formulation or the inclusion of other pertinent topics to be identified:

- a) **Marine Resources and Biodiversity:** Promote Sustainable Fisheries, offshore aquaculture and ecosystem valorization;
- b) **Healthy and Clean Ocean:** Observing, modelling and monitoring oceans and coastal areas for a better management of the Atlantic resources and the protection of related marine and coastal ecosystems to avoid significant adverse impacts.
- c) **Systems integration from outer space to Deep Ocean:** development, integration and use of sensors, devices and systems;
- d) **Mitigation and Adaptation to Climate Change:** including resilience of cities and coastal areas, disaster risk reduction and regional planning;
- e) **Sustainable Energy Systems:** including but not limited to marine renewable energy and the study of a transatlantic Ultra-High Voltage electrical connection between the Iberian Peninsula, Africa and Brazil.
- f) **Data science, artificial intelligence and learning systems for ocean, atmosphere dynamics and climate issues:** matching data producers and user needs.

In the meantime, a Cross-cutting Work Plan of initiatives is supporting the development of the Scientific Program of the AIR Centre in close cooperation with all relevant parts, to be approved in the **Canary islands in November 2018** and aligned with the Belem Statement (July 2017) signed between the European Commission, Brazil and South Africa and the associated concerted and support actions. Those initiatives may include, among others, developing the **"AIR data Net"** to implement common standards for data sharing, using existing supranational e-infrastructures to test them, including a large collaborative platform involving the Texas Advanced Computing center (TACC), the Barcelona Supercomputer Centre (BSC), the Minho Advanced Computing Center (MACC), and the LCC at the Federal University of Rio de Janeiro (UFRJ), among others;

The objectives of the **AIR data Net** are as follows:

- a) To enable and support a very large capacity federated data infrastructure providing ready-to-use data focused on the scientific domains of the AIR Center;
- b) To provide a one-stop shop data management facility for Open Data storage, processing and retrieval facility, a FAIR Data & Metadata Catalogue, Persistent IDs and versioned mid & long-term storage;
- c) To provide state-of-the-art presentation and visualization big data services and tools.