

# The BioData.pt Compute Platform

Jorge S. Oliveira<sup>1,2</sup>, Diogo Pratas<sup>1,3</sup>

On behalf of the Compute Platform

1 - BioData.pt | ELIXIR-PT, 2 - Instituto Gulbenkian de Ciência, 3 - University of Aveiro



## Summary

The BioData.pt Compute Platform provides a full stack of computational resources to the Portuguese life science research community, including a dedicated user support. The BioData.pt computational infrastructure, implemented over a hybrid cloud and high performance computing resources, supports access, usage and storage of digital objects and is coordinated by INESC-ID.

This infrastructure was maintained and expanded in 2021 and includes CCMAR, IGC and Técnico ULisboa as cloud service providers, and constitutes the backbone of most of BioData.pt's activities.

Additionally, as a member of the Galaxy Pulsar Network, the BioData.pt continues to contribute with computing capacity to Galaxy Europe, enabling researchers to use the Pulsar Network resources.

## Hardware Resources



**336 CPUs**  
**2.4 RAM (TB)**  
**79.2 Storage (TB)**  
**25 Instances**



**400 vCPUs**  
**5.1 RAM (TB)**  
**112.6 Storage (TB)**  
**19 Instances**

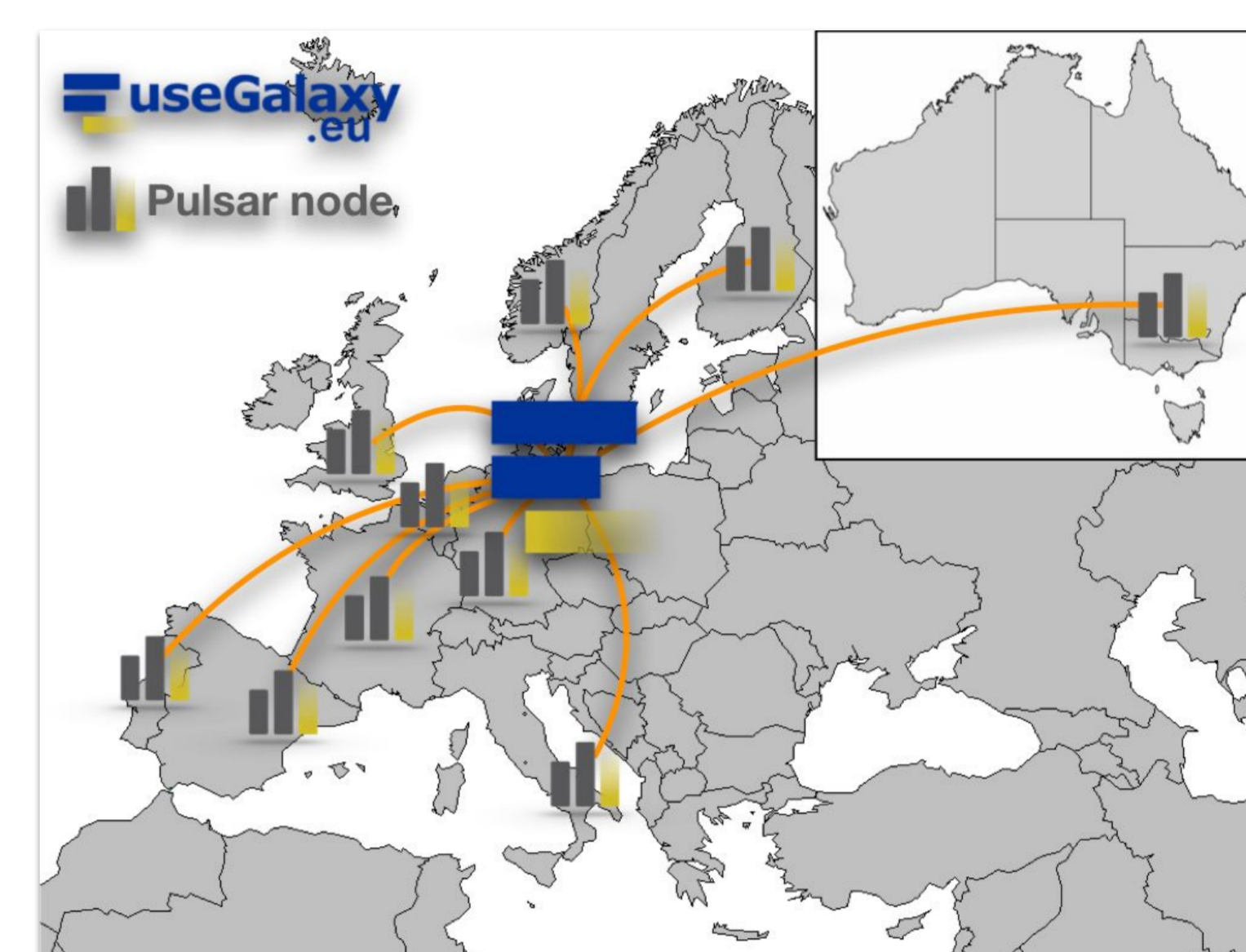


**1368 CPUs**  
**6.5 RAM (TB)**  
**1656 Storage (TB)**  
**52 Instances**

## 1 Main Goals

- Identifying the main National resources and infrastructures.
- Ensuring the operation of ELIXIR-PT | BioData.pt services by providing and maintaining a distributed compute infrastructure.
- Supporting the community in the creation of new services and tools.
- Link the information to ELIXIR at European level.

## Pulsar Network



## Software resources

# 2



### OpenStack-based Cloud Service

- We provide custom Virtual Machines for our associates to support their service, and also support their maintenance.

### Virtual Machine Images

- Bioinformatics or Galaxy Docker Images & Genomics Virtual Lab images.

### Data Compression

- Availability of data compression tools with general and specific purpose to diminish the storage with loss of information.
- Identification of recommended data compression tools for specific research purposes, for example genomics, proteomics, etc.

### Data Encryption

- Availability of data encryption tools for data transport for external applications when sensitive data is used.



## 3 Future Work

- Participation in the project "**Genomics Data Infrastructure**" (GDI), submitted and approved under the call "DIGITAL-2021-CLOUD-AI-01-FEI-DS-GENOMICS — Federated European infrastructure for genomics data".
- **Training** (Together with Training Platform)
  - Perform landscape analysis of community specific needs
  - Develop and deliver training (online and F2F)
- Regular **dissemination** of available infrastructures and software tools.
- Improvement of current web resources, like the **Data Management Portal** with integration of better functionalities and ontologies.



## Final Remarks

The BioData.pt Compute Platform is connected to all its activities. Its correct operation ensures the smoothness of the activities and guarantees our reputation.

It has been challenging to keep the quality standards at a high level without funding for appropriate human resources and maintenance of the infrastructure, hopefully it will get better with GDI.

## Contacts

Name: Jorge Oliveira

E-mail: cto@biodata.pt

Website: www.biodata.pt



@BioData\_PT



@BioData.PT



@BioData-PT



GULBENKIAN  
CIÊNCIA



TÉCNICO  
LISBOA

CCMAR  
Centro de Ciências do Mar

