The Portuguese National Distributed Computing Infrastructure and Related Activities

Cofinanciado por:
COMPETE 2020
PORTUGAL 2020
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional
Lisb@20 20
FCT Fundação para a Ciência e a Tecnologia

National Distributed Computing Infrastructure

Services: scientific computing, data processing and other data oriented services

Target: scientific and academic community, infrastructures, R&I projects, SMEs

Promote: shared resources, advanced computing and data services for research

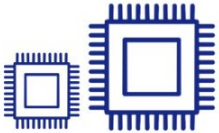
Interface: international digital infrastructures (EGI, IBERGRID, WLCG, EOSC)

Cloud Computing

cloud computing

HTC Computing

high throughput computing (GRID)



HPC Computing

high performance computing

More than 20 years of experience

2000 — LIP participation in R&D projects focused on grid computing.

2006 — The national grid initiative is established by the ministry. IBERGRID joins PT and ES centres. WLCG MoU is signed.

2008 — Portuguese main node for grid computing, joining LIP, LNEC and FCCN. Collaboration with U.Porto, U.Minho, U.Aveiro etc

2010 — Construction of a datacenter for the main node for grid computing at LNEC. Portuguese WLCG Tier-2 for ATLAS and CMS is established.

2010 — Portugal joins EGI. IBERGRID activities are boosted. PT LHC Tier-2 starts.

2013 — Creation of INCD is proposed to the national roadmap of research infrastructures (RNIE).

2014 — The proposal for the creation of INCD passes the scientific evaluation.

2015 — The legal entity “Associação INCD” is established.

2016 — The 1st project to fund INCD is submitted under the infrastructures roadmap.

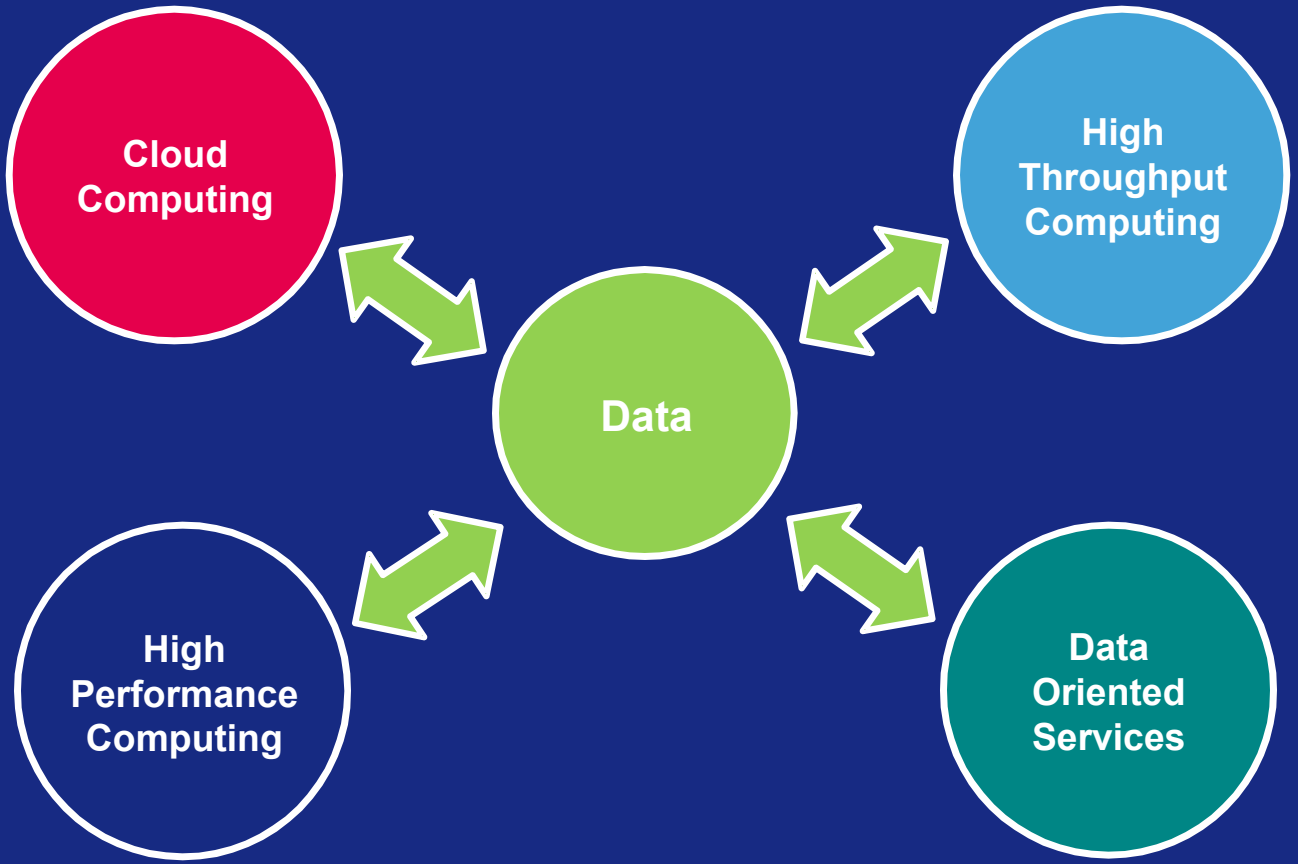
2017 — The project to establish INCD starts. Establishing the administrative and organizational basis for INCD.

2018 — First INCD collaborators are contracted. First international public tender for compute and storage.

2021 — Infrastructure project is extended until October 2022.

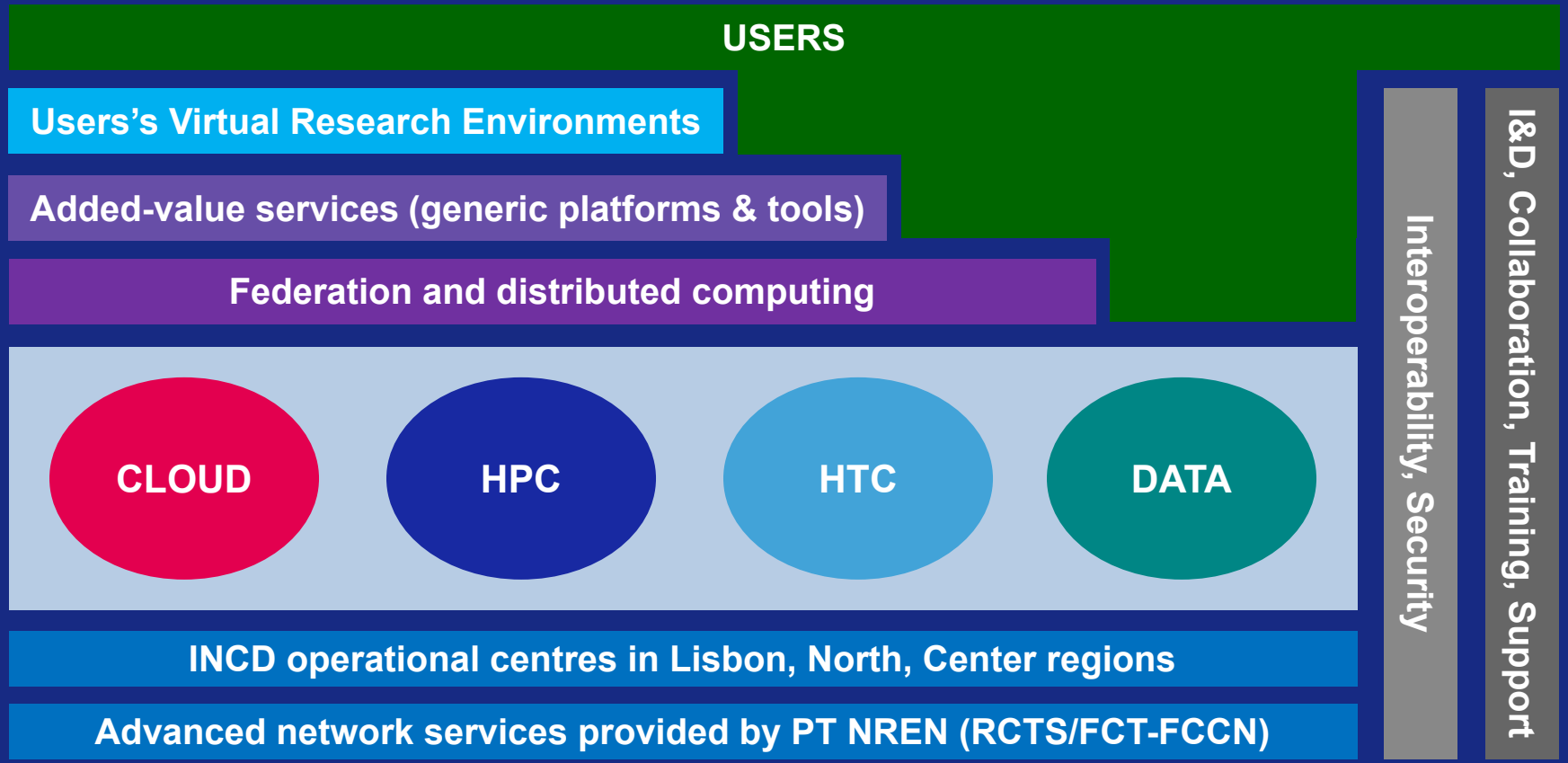
2022

Holistic approach to research requirements



Distributed Computing & Federation

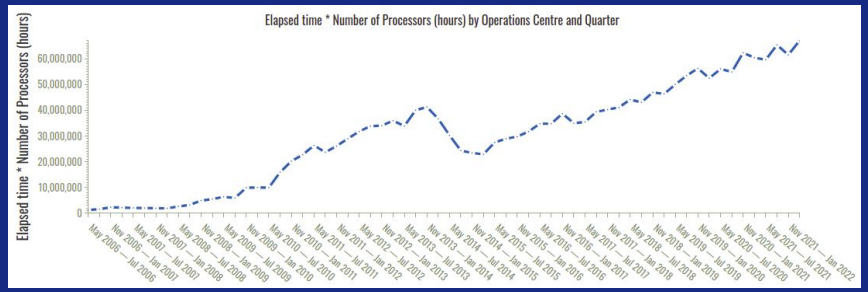
High level view



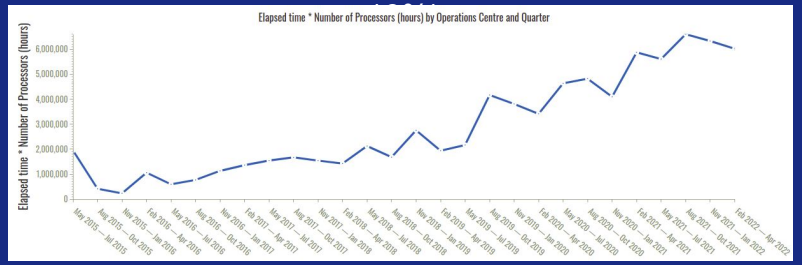
Federation in IBERGRID and EGI

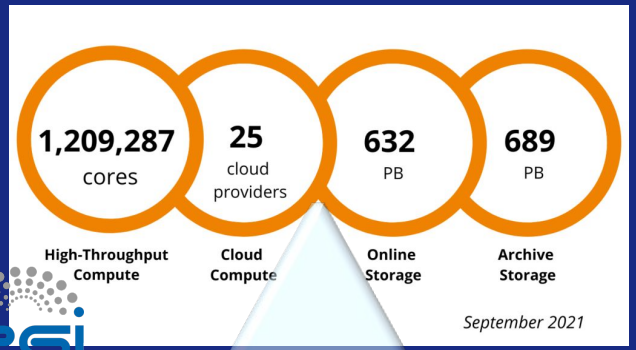
INCD is the main Portuguese infrastructure in EGI and IBERGRID

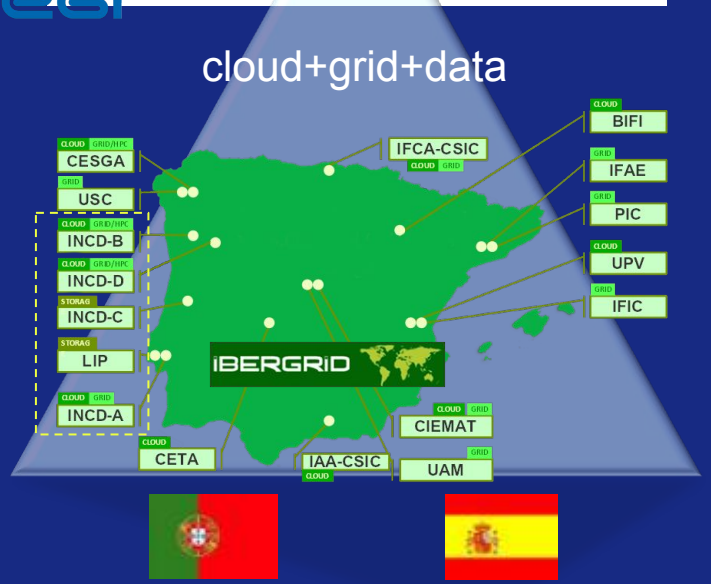
IBERGRID - 309 million jobs since 2006 (> 10%)



IBERGRID - 1 million instantiated VMs since 2015 (> 11%)







IBERGRID and EGI

Responsibilities and activities (LIP+INCD):

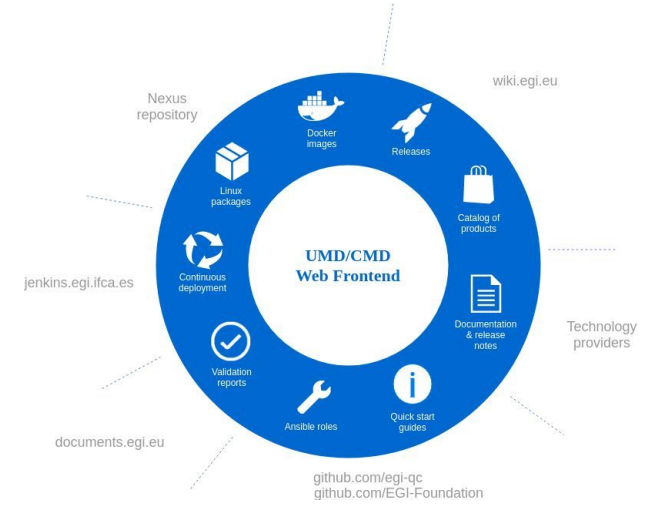
- Operations coordination at Iberian level and interface with EGI operations
- Software management for the EGI and IBERGRID federations
- National technical contact point
- Security contact for Portugal
- Support for user communities
- Developing and operating core services e.g. software repositories for the EGi federation
- Integration of thematic and/or user services

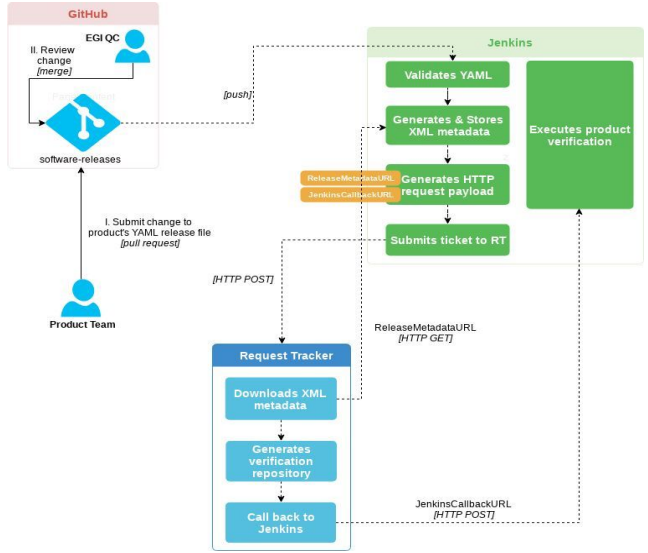
Software management and repositories for EGI

Quality assurance for the EGI middleware distributions for grid (UMD) and cloud (CMD).

- **Production**
  - Manage the EGI software validation process.
  - Software validation of products to be released as part of CMD and UMD distributions.
  - Automated validation in isolation environments and piloting at selected sites.

- **Innovation**
  - New streamlined validation process.
  - New repositories with added capabilities.
  - New frontend.





INCD operational centers

Lisbon Region I
(UPGRADE ONGOING)
HPC / HTC / Cloud / Federation

North Region I
HTC / HPC

Center Region
(NEW TO BE DEPLOYED)
Tape storage expansion

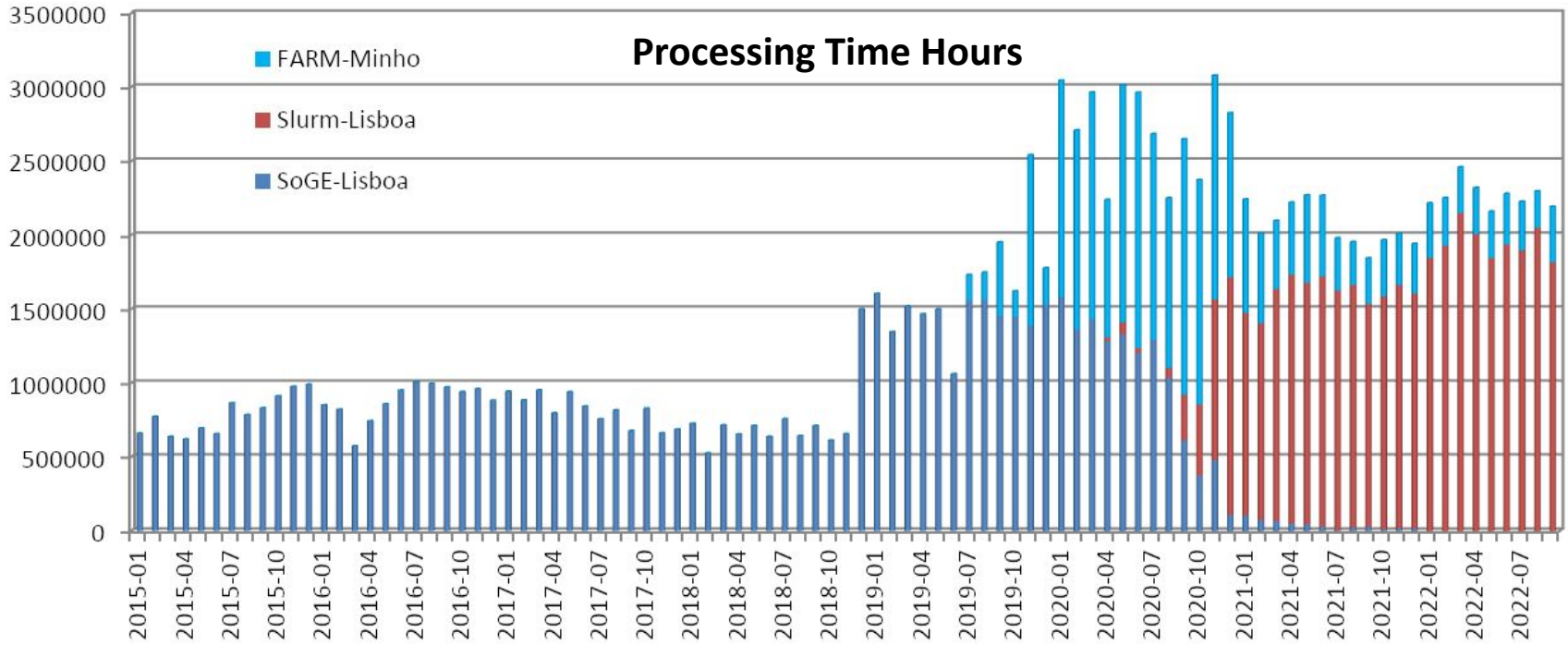
Lisbon Region II
Tape storage

North Region II
(NEW TO BE DEPLOYED)
HPC / HTC / Cloud / Federation

Main services

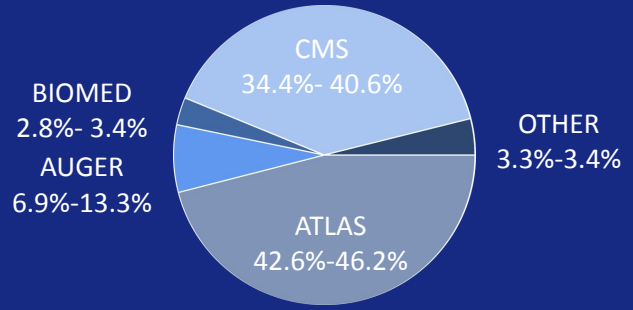
- **HPC**
  - Slurm batch system
  - Infiniband network
  - Lustre
  - CVMFS
  - Access to GPUs A100/V100/T4
  - Containers
- **HTC**
  - Slurm batch system
  - Ethernet network
  - Lustre
  - CVMFS
  - NFS
  - Access to GPUs A100/V100/T4
  - Containers
- **Cloud**
  - Openstack
  - Virtual networks
  - Ceph block storage
  - Ceph object storage
  - Dashboard and APIs
  - Access to GPUs V100/T4

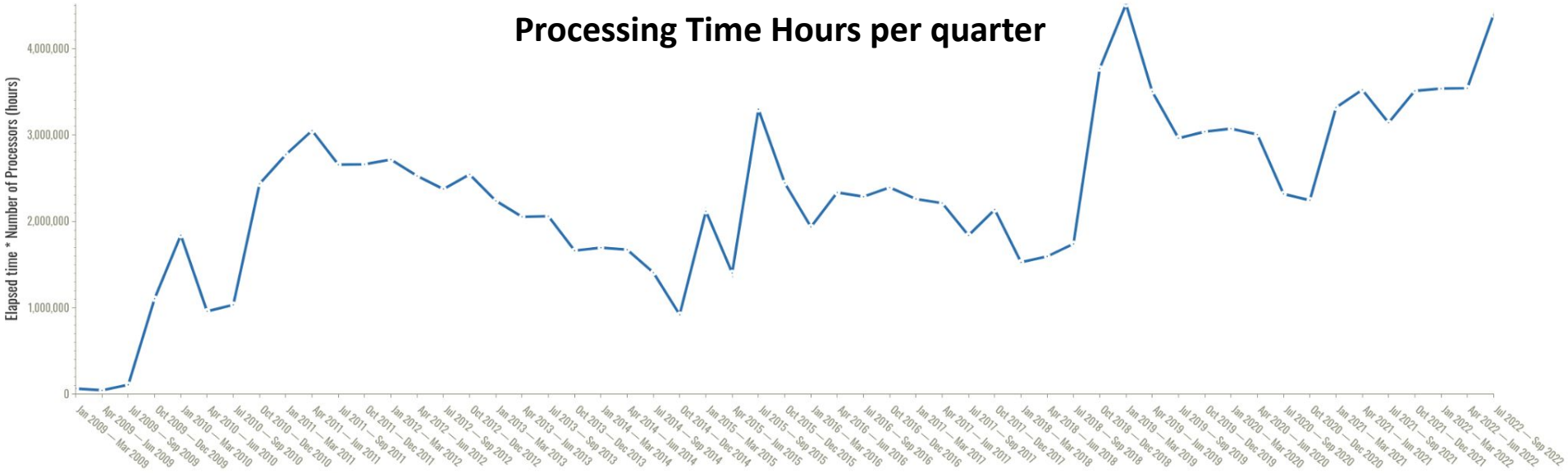
Batch clusters usage both HTC and HPC



Total usage in 2021 over 24.789.000 hours

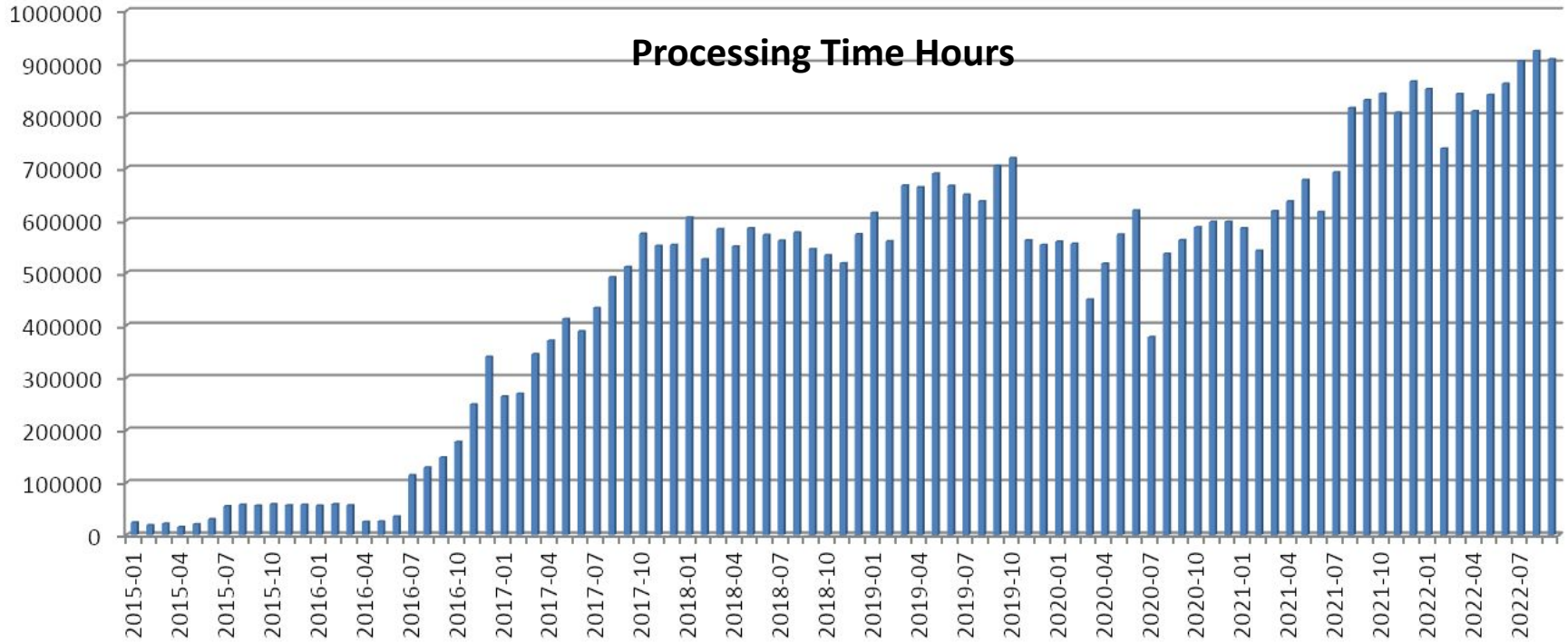
Distributed computing usage (grid)





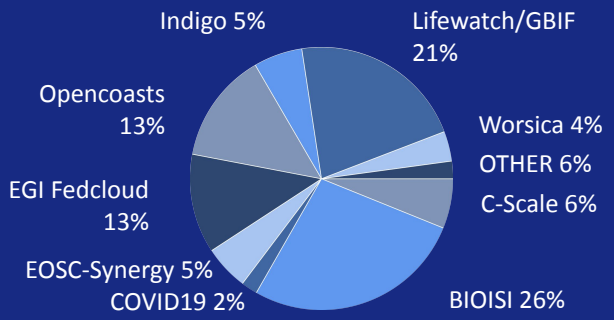
Total usage in 2021 over 13.500.265 hours

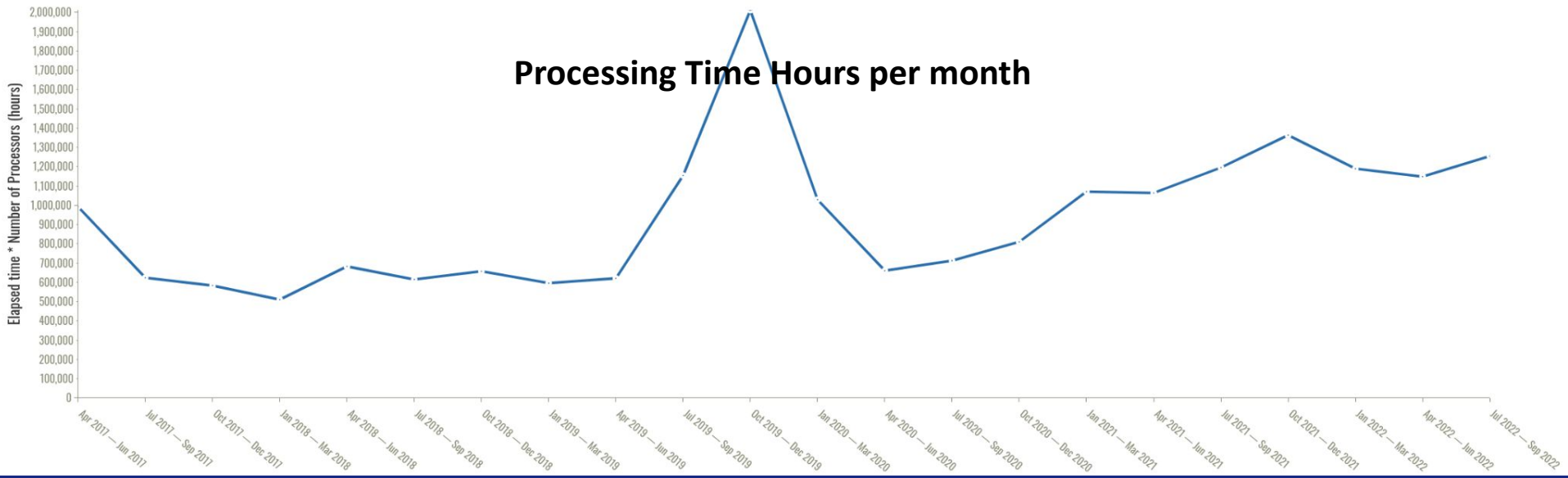
Cloud usage



Total usage in 2021 over 8.523.000 hours

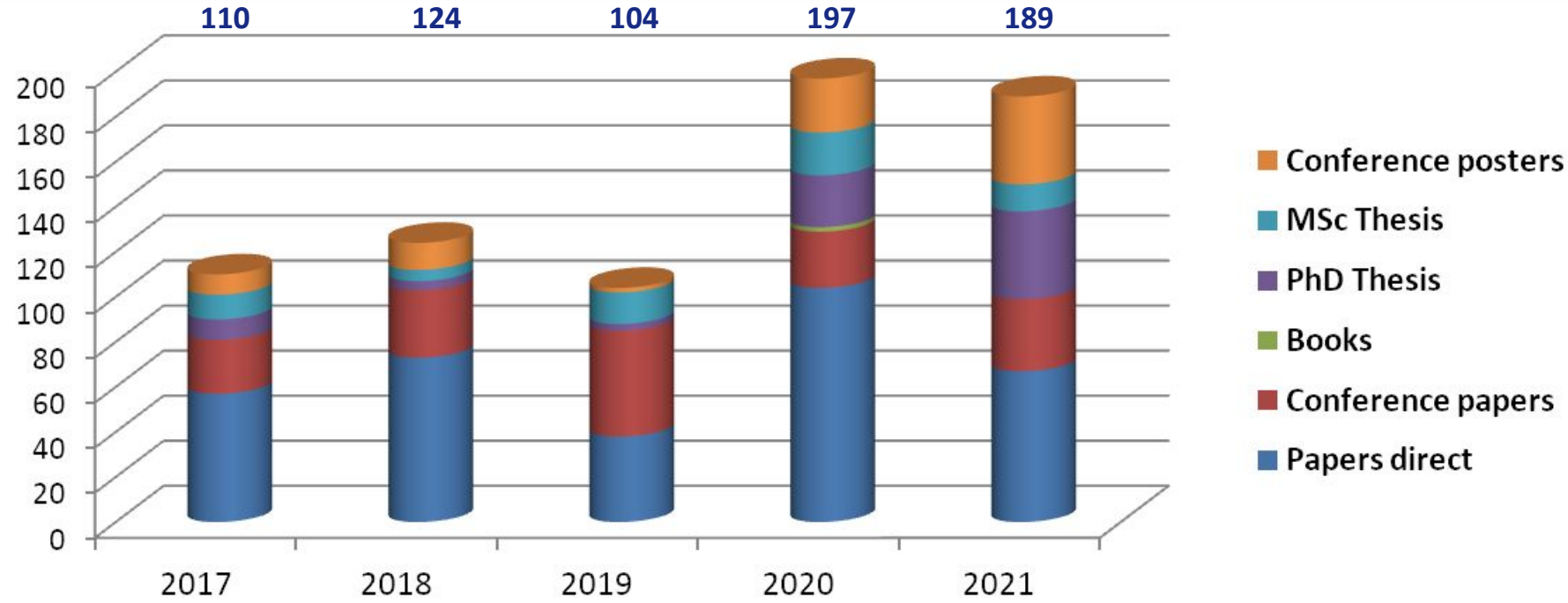
Federated cloud usage





2021 total usage 5,273,643 hours

Publications



Metrics

In 2021

○ Curated datasets:	2
○ Patents:	2
○ FCT and international funded projects supported:	68
○ CPCA supported projects:	50
○ Artificial intelligence in public administration projects:	2
○ Organizations, research units and infrastructures:	61
○ Publications (thesis, papers, posters, books):	189

Since 2017

○ Publications (thesis, papers, posters, books):	724

Some Technical Details

EC Projects & Related Activities

Other projects and initiatives

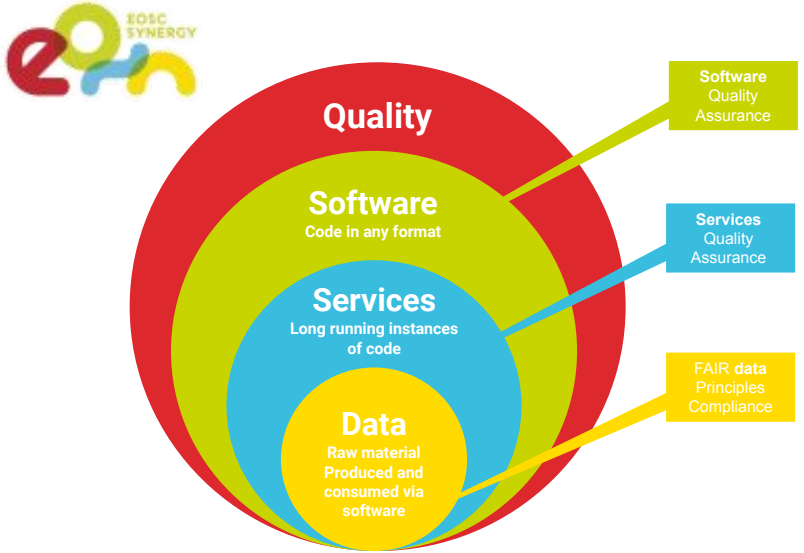
- **EOSC - European Open Science Cloud**
  - EOSC-synergy (2019-2022)
  - EGI-ACE (2021-2023)
  - C-Scale (2021-2023)
  - iMagine (2022-2025 new starts September)
- **EuroHPC**
  - EuroCC (20220-2022) indirectly through the INCD partners

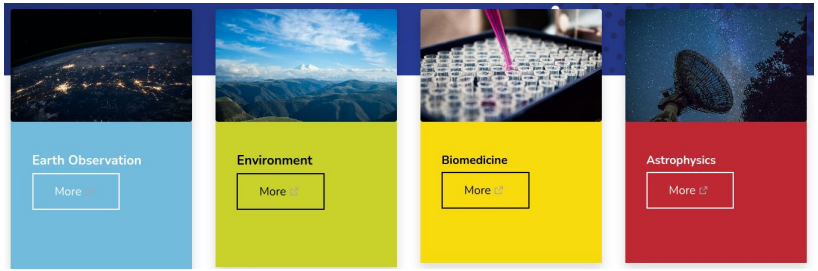
EOSC-Synergy

European Open Science Cloud (EOSC) project.

EOSC-Synergy (2019-2022) expanding national e-infrastructures and services in EOSC. Strong focus on quality, infrastructure, and thematic services. INCD participation with LIP, LNEC and FCT-FCCN on:

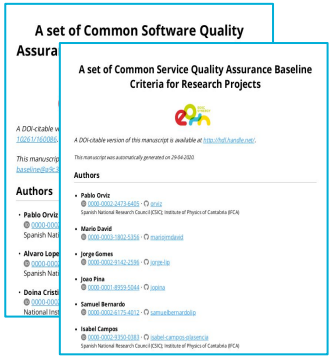
- Providing cloud services (INCD+LIP)
- Thematic services integration (LNEC)
- Support training activities (INCD+LIP)
- Data repositories (INCD+LIP)
- Computing services integration (INCD+LIP)
- Software and services quality assurance (LIP)
- Policy (FCT-FCCN)





EOSC-Synergy - SQAaaS

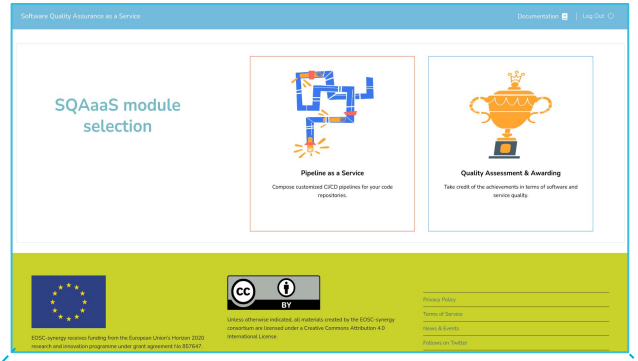
https://sqaaas.eosc-synergy.eu/



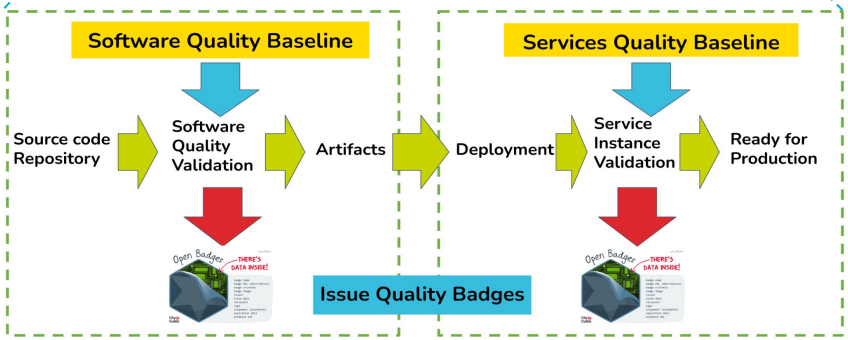
Quality Baselines Good Practices

IBERGRID partners

- LIP
- IFCA/CSIC
- UPV



SQAaaS Platform



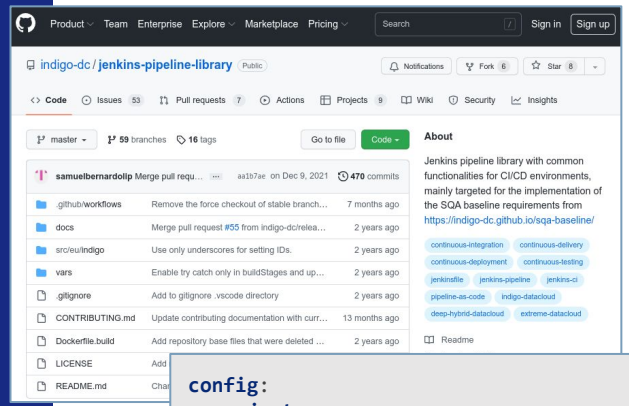
Quality Badges

Software Classification			
Quality Criteria	Quality Badges		
	Gold	Silver	Bronze
QC.Acc	Accessibility		
QC.Lic	Licensing		
QC.Sty	Code style		
QC.Met	Metadata		
QC.Uni	Unit tests		
QC.Doc	Documentation		
QC.Sec	Security		
QC.Wor	Workflows		
QC.Ver	Versioning		
QC.Man	Code management		
QC.Del	Delivery		

EOSC-Synergy - SQAaaS

Jenkins Pipeline Library (JePL)

- The library that powers the SQAaaS platform.
- Especially suitable for complex setups, you can use directly the JePL instead of the SQAaaS.
- Tech-savvy users tend to favor code over a graphical interface for the task of managing their CI/CD pipelines.
- JePL uses pipeline descriptions written in YAML.
- Just add JePL to your software repository and build your software or service quality assurance using YAML descriptions to benefit from the full set of features.
- JePL implements the software and service baselines maintained by EOSC-Synergy.




```yaml
config:
  project_repos:
    myrepo:
      repo: 'https://github.com/myorg/myrepo'

 sqa_criteria:
   qc_style:
     repos:
       myrepo:
         container: myrepo-testing
         tox:
           tox_file: /myrepo-testing/tox.ini
             testenv: stylecheck
```

```
|-- .sqa
|    |-- config.yml
|    |-- docker-compose.yml
  |-- Jenkinsfile
```

```yaml
services:
  myrepo-testing:
    image:
"indigodatacloud/ci-images:python3.6"
    hostname: "myrepo-testing-host"
    volumes:
      - type: bind
        source: ./myrepo
          target: /myrepo-testing
```

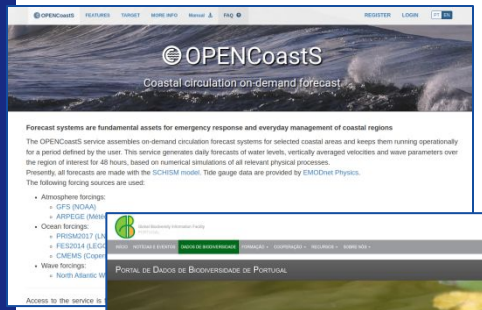
https://github.com/indigo-dc/jenkins-pipeline-library

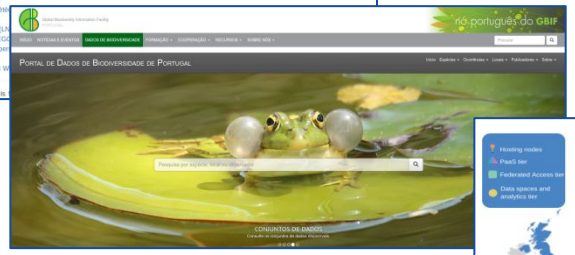
EGI-ACE

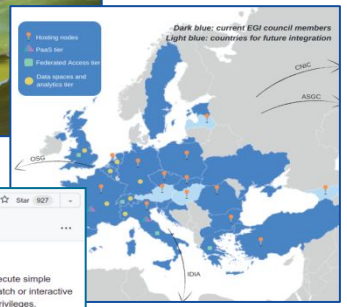
European Open Science Cloud (EOSC) project.

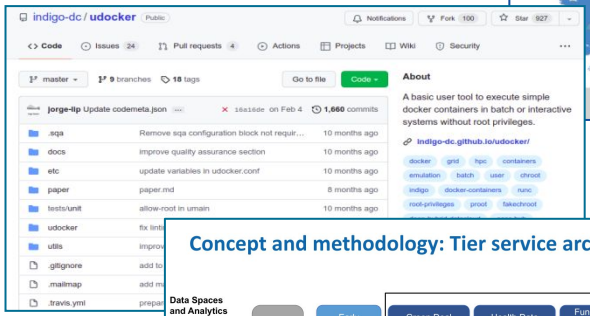
EGI-ACE (2021-2023) Advanced Computing for EOSC.
INCD participation on:

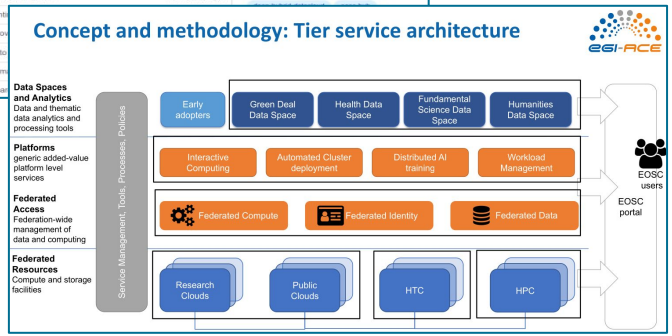
- Provisioning of cloud services.
- HPC integration in EOSC.
- Hosting of software repositories for EGI.
- Integration of Iberian biodiversity data repositories in collaboration with the GBIF national node.



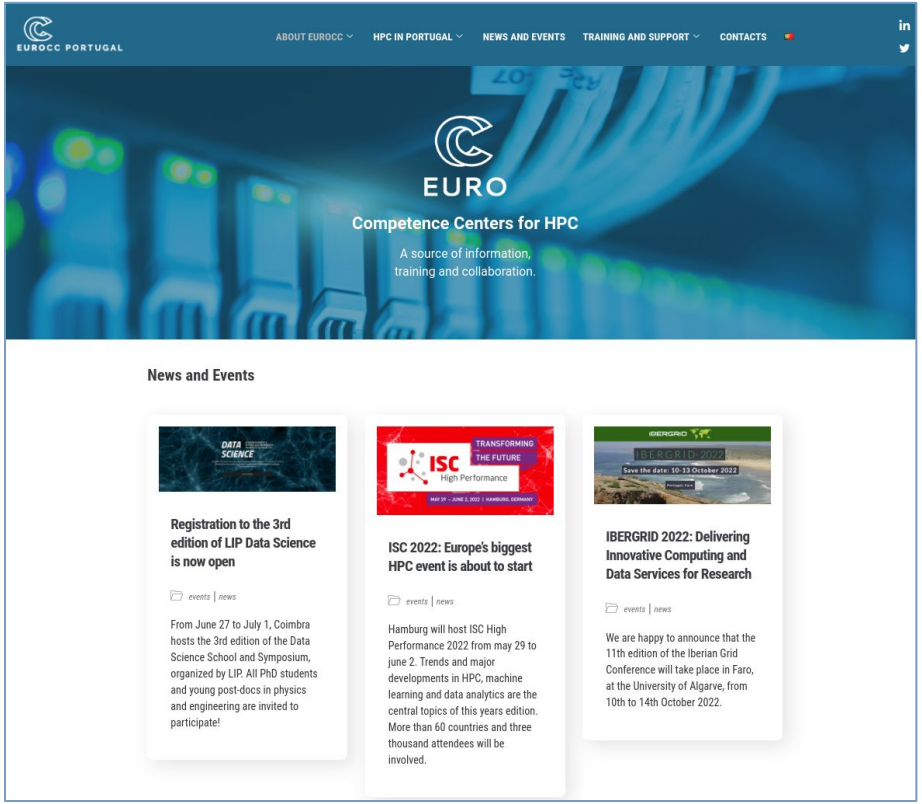








EuroCC - National Competence Centers



High Performance Computing National Competence Center in EuroCC

- Boost European High Performance Computing (HPC) knowledge
- European network of 33 national HPC competence centres (NCCs)
- Bridge the existing HPC skills gaps while promoting cooperation

INCD is not a partner but collaborates in:

- Technology transfer
- Training and skills development
- Awareness and collaboration
- Collaboration with industry
- Access to expertise and knowledge

https://eurocc.fccn.pt / contacto@eurocc.fccn.pt

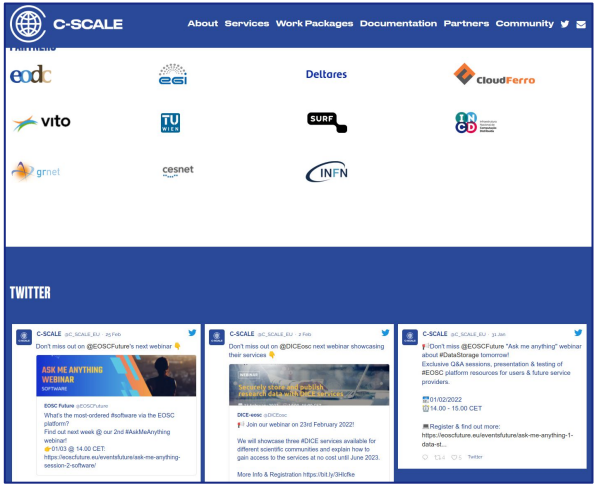
C-Scale: Copernicus - eoSC AnaLytics Engine

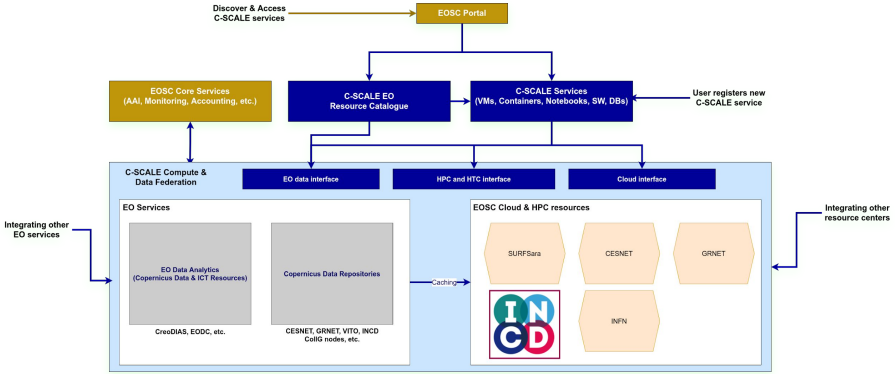
European Open Science Cloud (EOSC) project.

C-Scale (2021-2023) federate European Earth Observation infrastructure services such as Copernicus DIAS and others. Support Copernicus research and operations with large and easily accessible European computing environments.

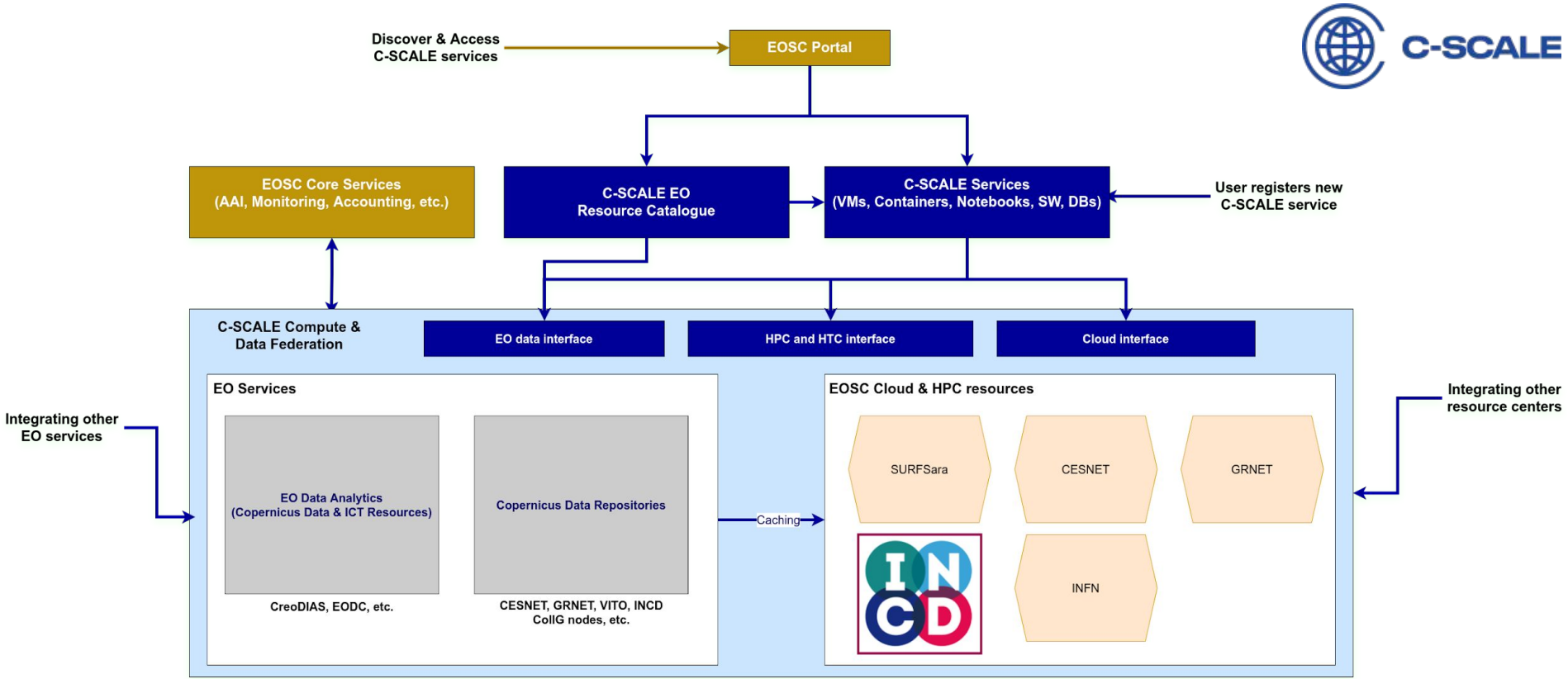
INCD participation:

- Providing cloud services.
- Support to thematic services integration.
- Data access and federation.

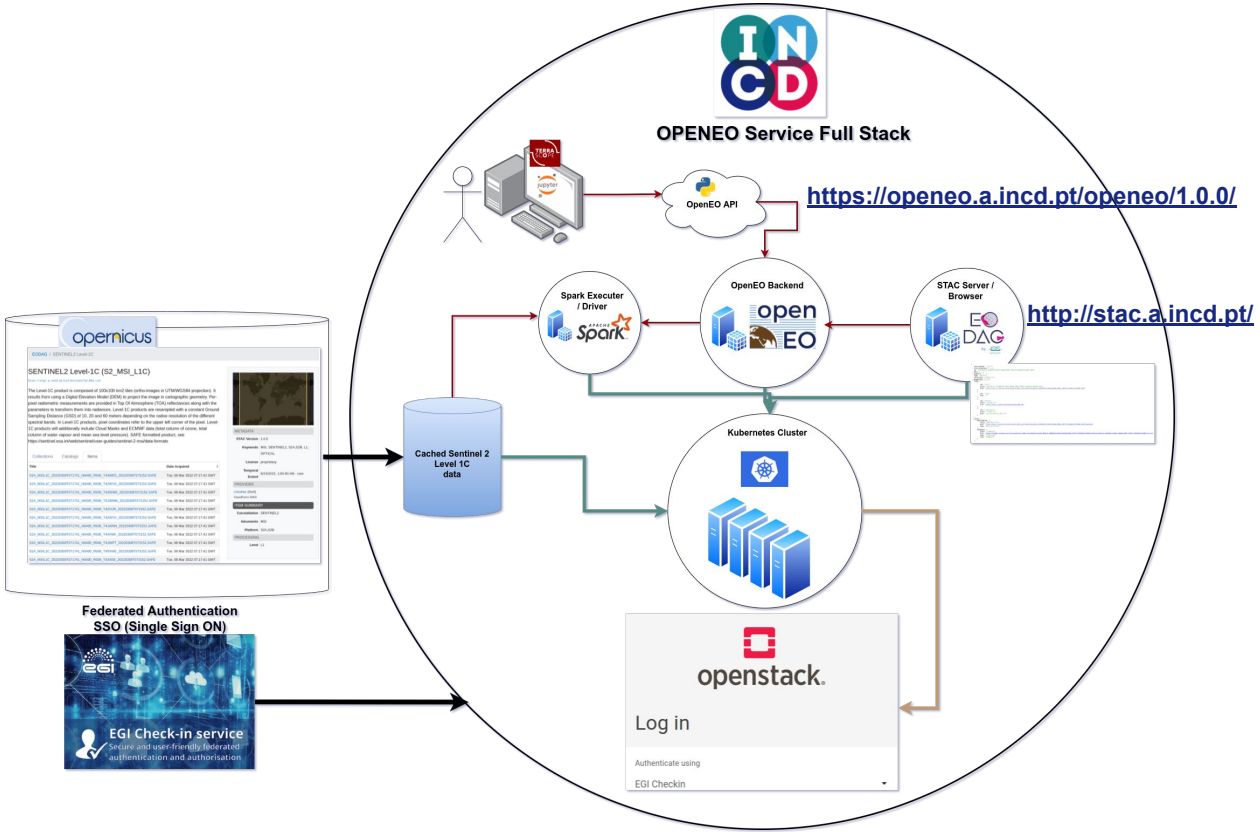




C-Scale - High level architecture



C-Scale: Deployment at INCD



Serving and caching Copernicus Sentinel images to serve EO applications.

Deployed at INCD exploiting the EGI fedcloud.

Using openEO, STAC Server with Kubernetes on Openstack

C-SCALE

iMagine

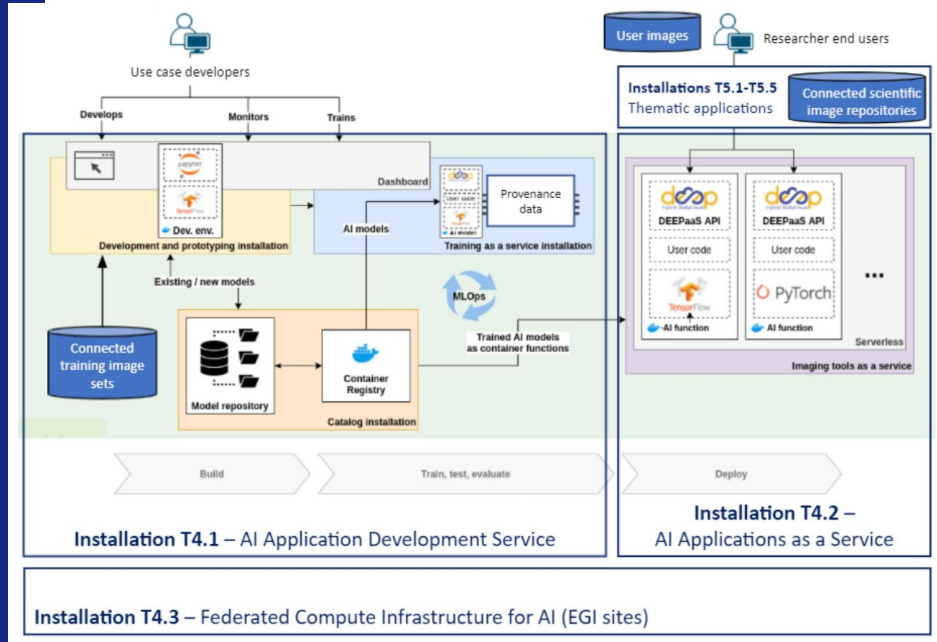
New project to start in September 2022

iMagine (2022-2025) Imaging data and services for aquatic science.

- Imaging data and services for aquatic science
- Ocean warming, and acidification
- Litter and oil spills monitoring of water surfaces
- Marine biomass estimation and preservation through real-time monitoring
- Coastal ecosystems, and beach-related human activities monitoring and analysis

INCD participation on:

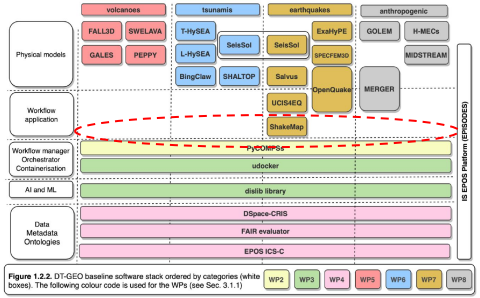
- Federated Compute Infrastructure for AI
- INCD partners also involved in other tasks



Other projects by the partners

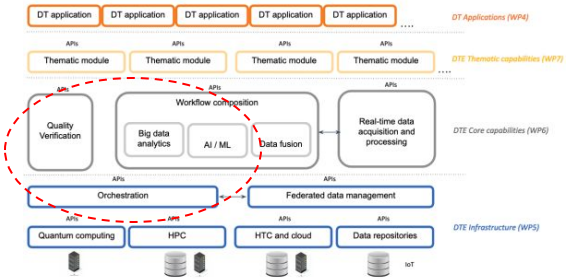
- New projects started in September focused on digitalTwins and Artificial Intelligence:
  - DT-GEO (INFRA-2021-TECH-01)
  - InterTwin (INFRA-2021-TECH-01)
  - AI4EOSC (INFRA-2021-EOSC-01)

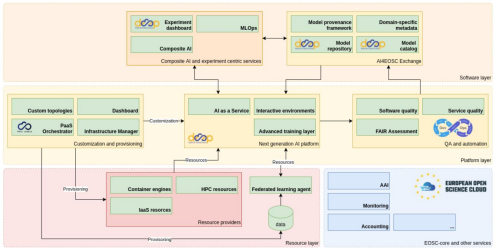
- Leveraging resources and capabilities:
  - Quality assurance
  - SQA-as-a-Service and JePL
  - udocker



DT-GEO, Digital Twin of geophysical extremes dealing with geohazards from earthquakes, volcanoes, and tsunamis.

InterTwin, to develop a common approach to the implementation of DTs applicable across the whole spectrum of scientific disciplines e.g. HEP, astronomy, climate, environment.

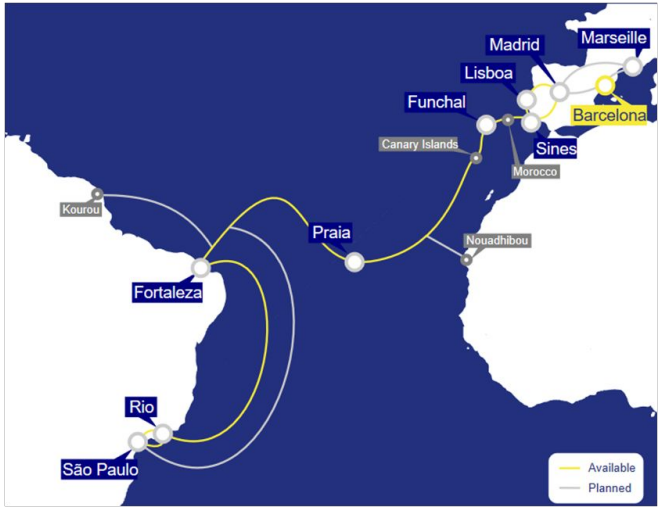


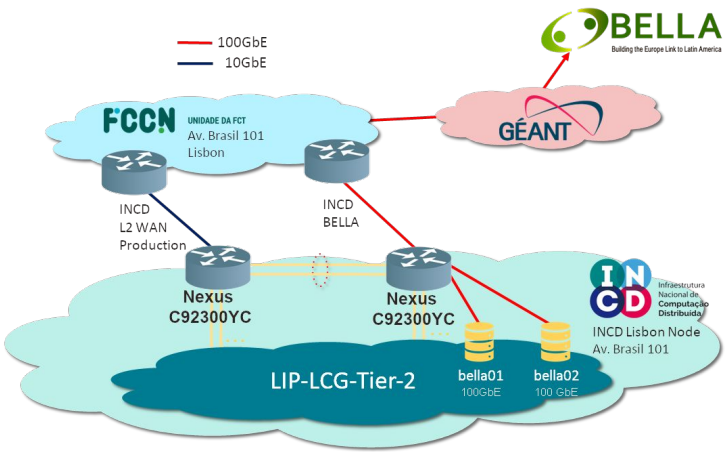


Advanced services for Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) models and applications in the EOSC.
Both generic (**AI4EOSC**)

Other infrastructure Related Activities

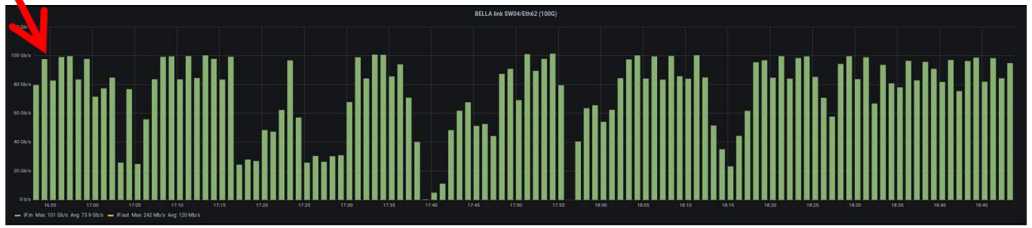
Experimenting with the ELLA transatlantic cable





1 172.16.203.254 (172.16.203.254) 0.382 ms
2 194.210.4.169 (194.210.4.169) 1.162 ms
3 Router30.Lisboa.fccn.pt (194.210.6.108) 0.562 ms
4 Router1.Lisboa.fccn.pt (194.210.6.103) 0.646 ms
5 fccn.mx2.lis.pt.geant.net (62.40.124.97) 0.495 ms
6 redclara-gw.lis.pt.geant.net (62.40.127.151) 62.728 ms
7 for-sao.redclara.net (200.0.204.7) 106.989 ms
8 sprace01.redclara.net (200.0.207.116) **106.452 ms** !X

100Gb/s

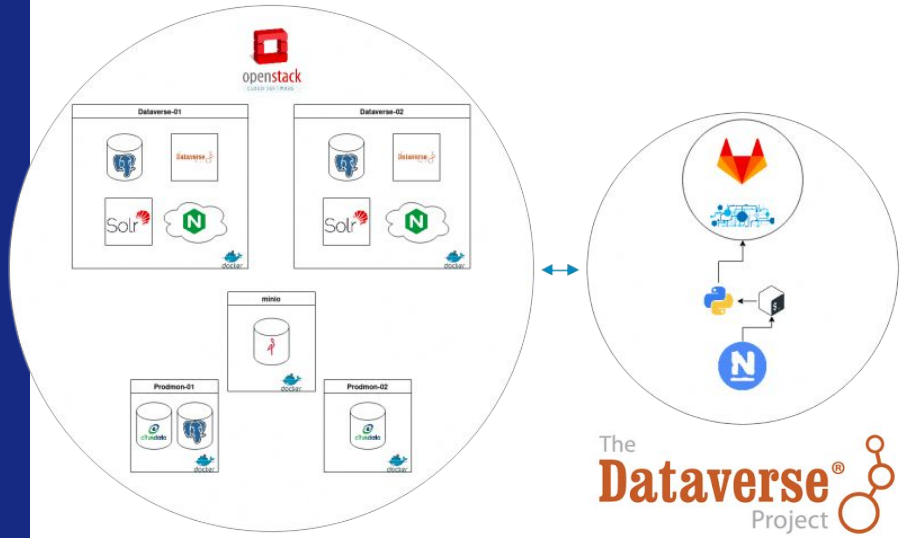


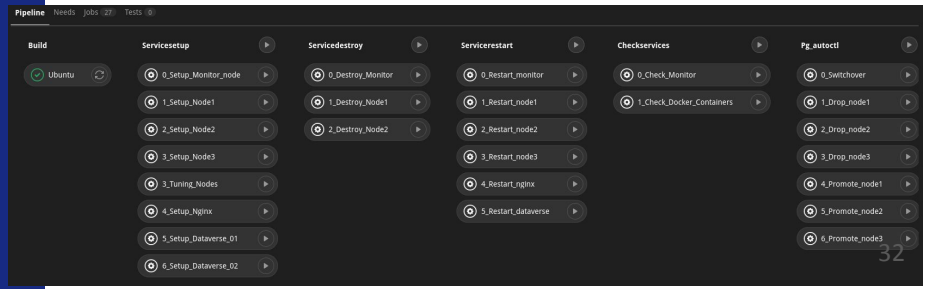
Using 2x INCD servers and 2x SPRACE servers (with 100GbE interfaces)

Open data

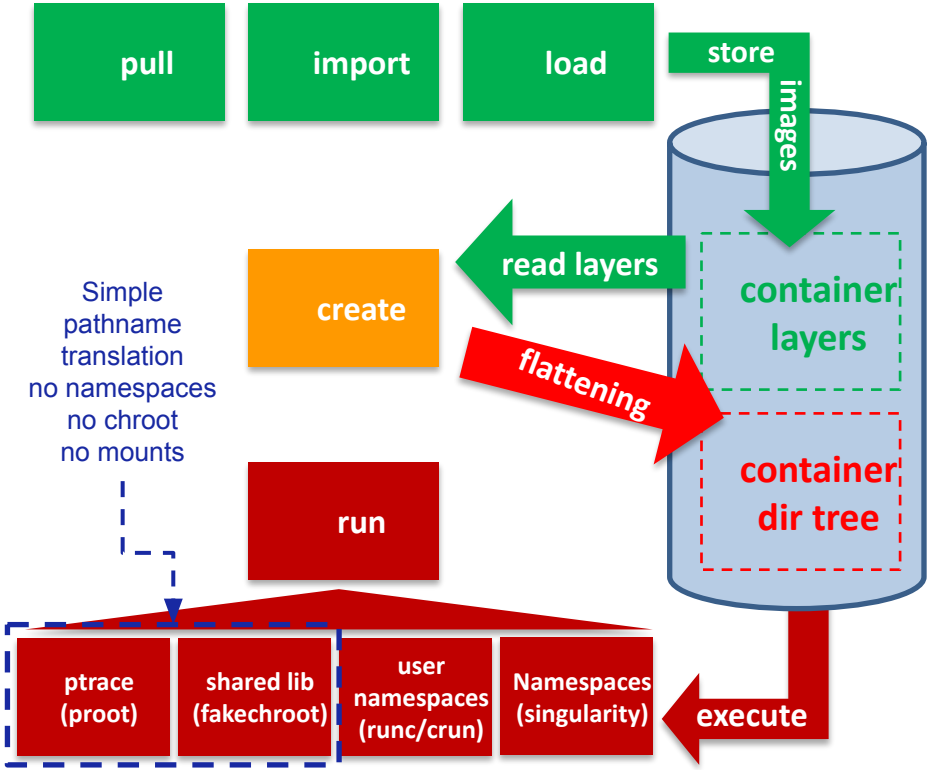
Feasibility study for a national catchall data repository aligned with open science and FAIR data principles.

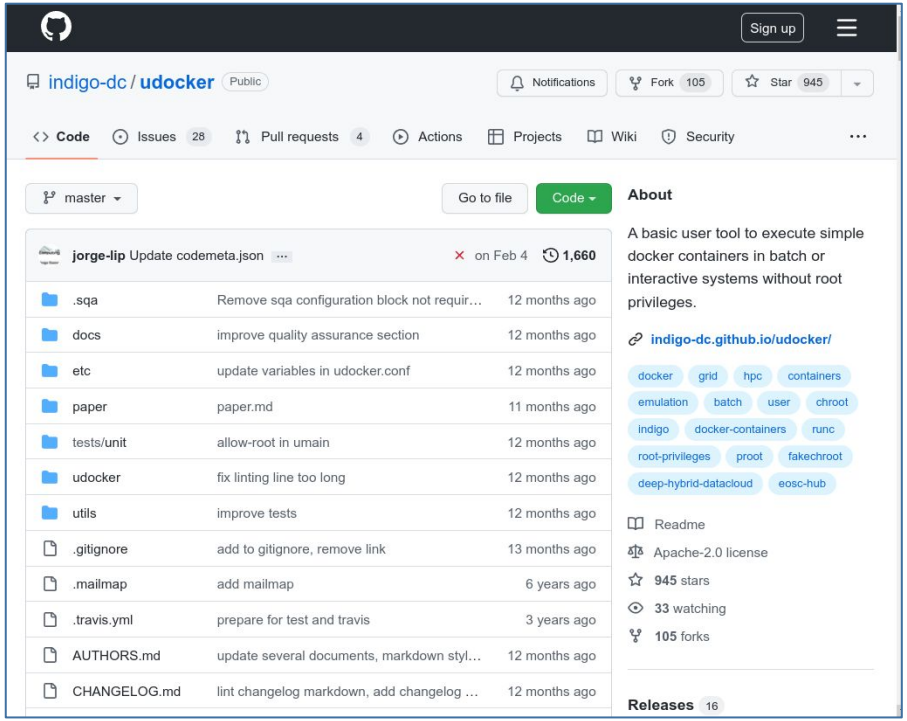
- Leverage EOSC-Synergy work on thematic data repositories and FAIR quality indicators.
- Productization and automation of Dataverse based data repositories.
- Address resiliency, redundancy and data recovery aspects.
- Integration with federated identity e.g. Ciência ID, ORCID.
- Integration with permanent identifiers.
- Integration with Ciência Vitae.





udocker - containers for end-users





https://github.com/indigo-dc/udocker

Advanced Computing Network - RNCA

RNCA Operational Centers

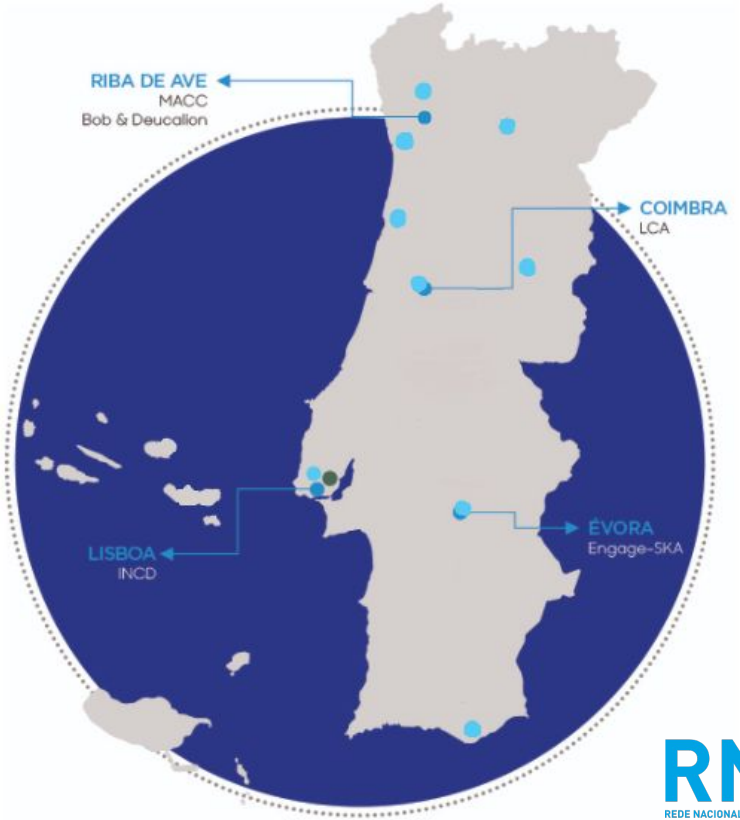
Infraestrutura Nacional de Computação Distribuída

LCA Laboratory for Advanced Computing

Minho Advanced Computing Center

UNIVERSIDADE DE ÉVORA

FCCN UNIDADE DA FCT
Tecnologia para o Conhecimento



RNCA
REDE NACIONAL DE COMPUTAÇÃO AVANÇADA

Advanced Computing Network - RNCA

RNCA aims to guarantee the conditions for the production of new knowledge and active participation in international R&D networks and programmes.

Calls supported by the operational centers.

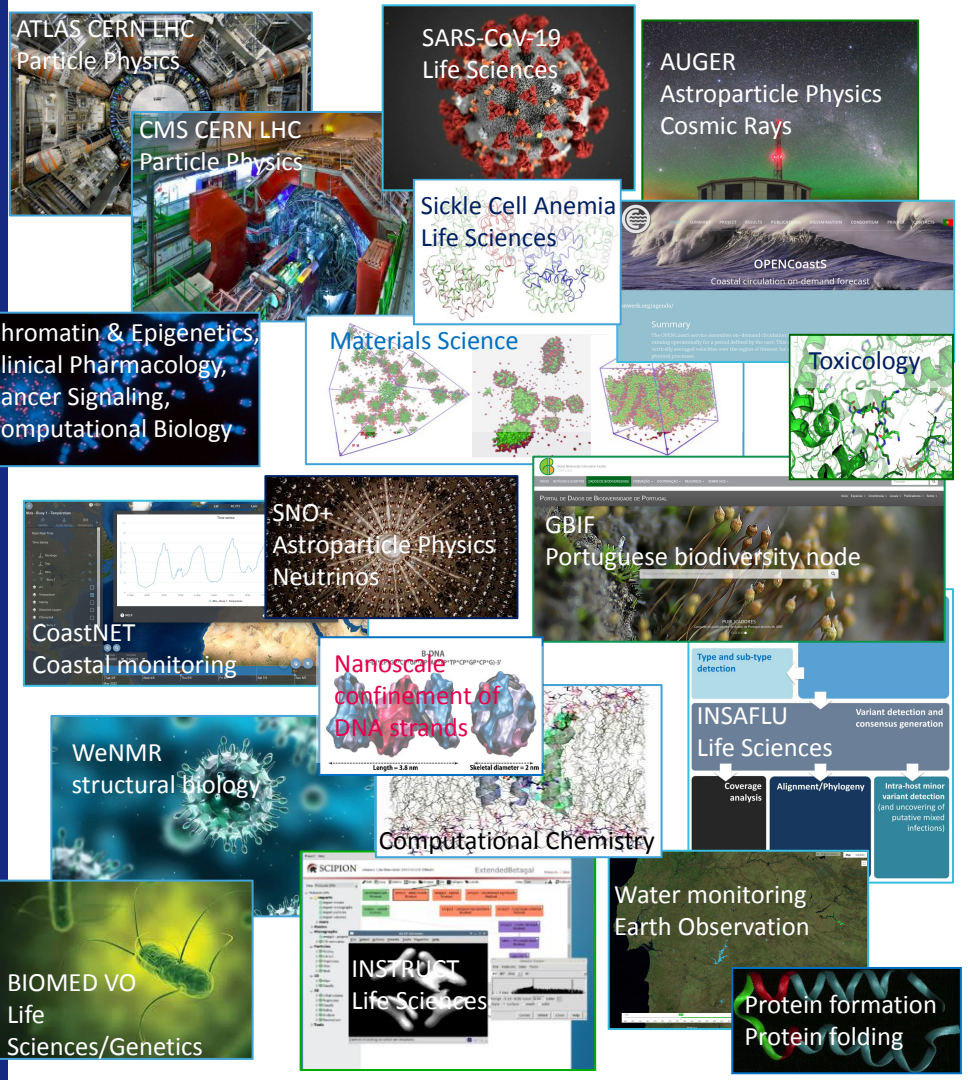
- 1st CPCA
- 2nd CPCA
- Data science and AI in the public administration
- Working towards the 3rd CPCA

RNCA
REDE NACIONAL DE COMPUTAÇÃO AVANÇADA

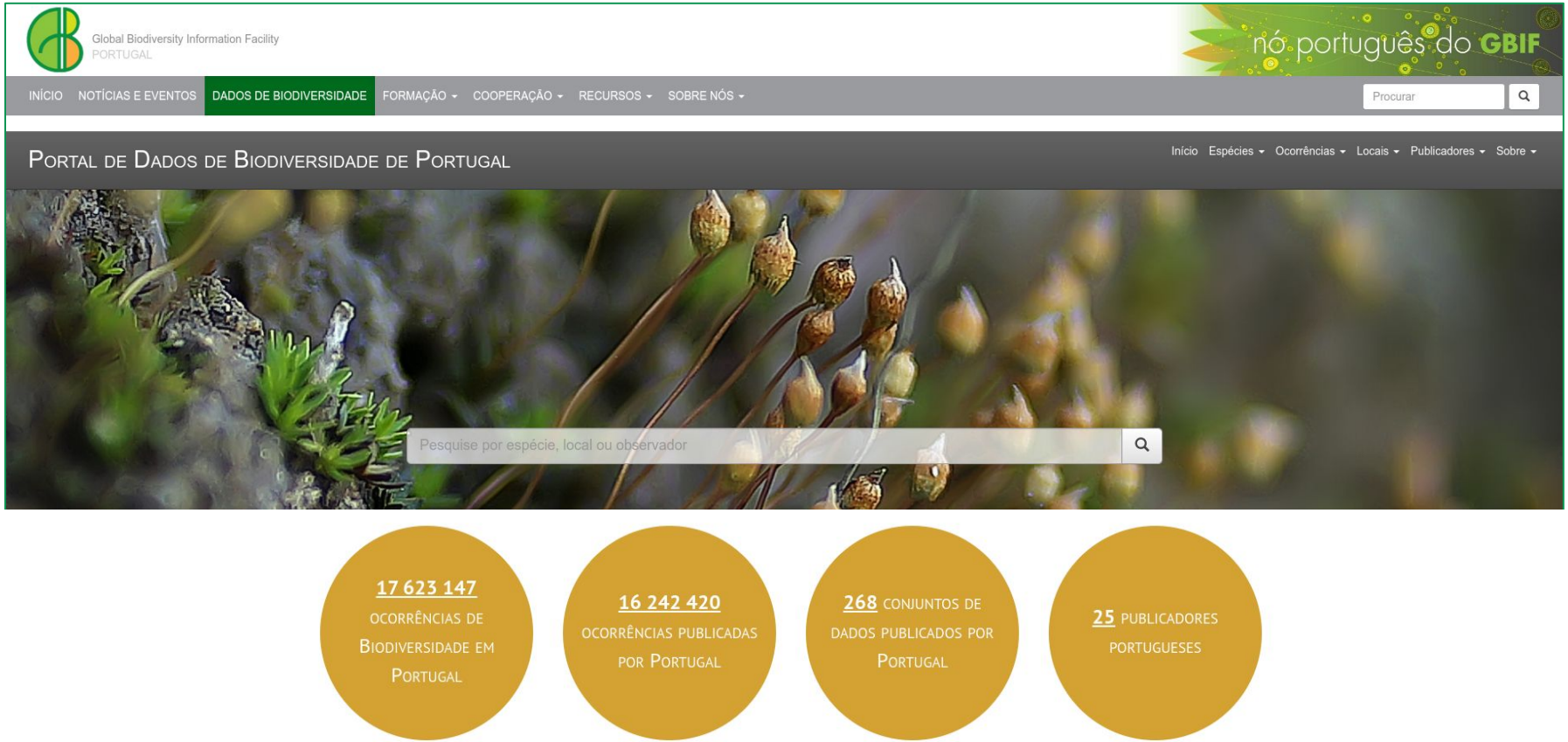
Usage Examples

Users

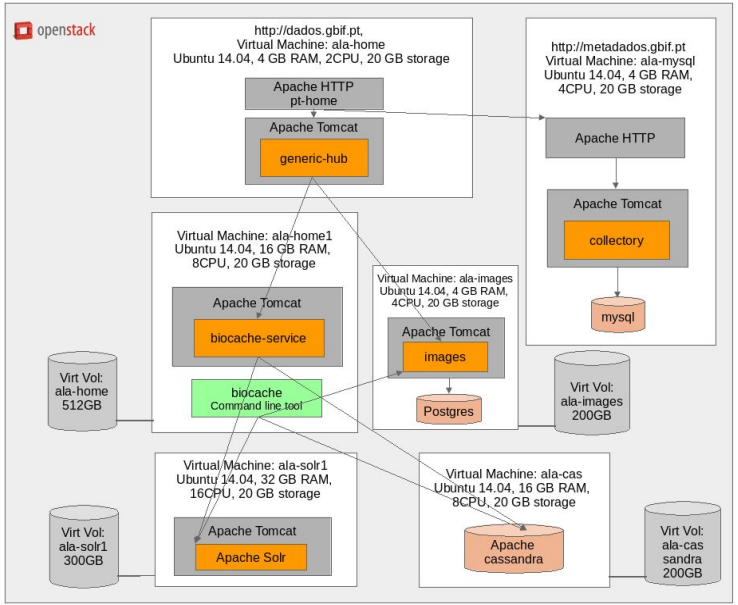
Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação (ARDITI)
Biomolecular SIMulations Research Group (BioSIM)
Business Research Unit (BRU) / ISCTE
Research Center for Biosciences & Health Technologies (CBIOS) / U.Lusofona
Concepção e Desenvolvimento de Sistemas de Informação (CDSI) / ISCTE
Centro de Estudos de Doenças Crónicas (CEDOC)
Centro de Engenharia e Tecnologia Naval e Oceânica (CENTEC)
Center for Astrophysics and Gravitation (CENTRA)
Centro de Ciências do Mar e do Ambiente (MARE)
Centro de Ecologia Evolução e Alterações Ambientais (cE3c)
Centro de Física do Porto (CFP)
Centro de Investigação e Desenvolvimento em Matemática e Aplicações (CIMA)
Centro de Investigação em Sistemas Electromecatrónicos (CISE)
Centro de Física da Universidade de Coimbra (CFISUC)
Centro de Física Teórica de Partículas (CFTP)
Fundação Champalimaud
Instituto de Materiais de Aveiro (CICECO)
Centro de Neurociências e Biologia Celular (CNBC)
Departamento de Engenharia Informática / FEUP / U.Porto
Departamento de Engenharia Informática / UC
Departamento de Biologia / U.Minho
Departamento de Física / U.Coimbra
Departamento de Informática / U.Coimbra
Faculdade de Ciências / U.Porto
Faculdade de Farmácia / U.Lisboa
Faculdade de Medicina / U.Lisboa
Fundação para a Ciência e a Tecnologia - Unidade FCCN (FCT-FCCN)
Nova School of Science and Technology (FCT/UNL)
Institute for Bioengineering and Biosciences (IBB)
Instituto de Biologia Experimental e Tecnológica (IBET)
Instituto Dom Luiz (IDL)
Instituto Gulbenkian de Ciência (IGC)
Instituto de Geografia e Ordenamento do Território (IGOT)
Instituto de Medicina Molecular (IMM)
Instituto de Ciência e Inovação em Engenharia (INEGI)
Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento (INESC-ID)
Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC-TEC)
Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA)
Instituto de Ciências Sociais da Universidade de Lisboa (ICS)
Instituto de Saúde Ambiental (ISAMB)
Instituto de Telecomunicações (IT)
Instituto de Plasmas e Fusão Nuclear (IPFN)
Instituto Superior de Agronomia (ISA)
Instituto Superior de Educação e Ciências (ISEC)
Instituto Superior de Engenharia de Lisboa (ISEL)
Instituto Universitário de Ciências Psicológicas, Sociais e da Vida (ISPA)
Instituto Superior Técnico (IST)
Instituto de Tecnologia Química e Biológica António Xavier (ITQB)
Computer Science and Engineering Research Centre (LASIGE)
Laboratório de Instrumentação e Física Experimental de Partículas (LIP)
Laboratório Nacional de Engenharia Civil (LNEC)
Nova School of Business and Economics (Nova SBE)
Research Center for Risks and Sustainability in Construction (RISCO)
Sociedade Portuguesa de Botânica (SPBOTANICA)
Universidade Aberta (U.Aberta)
Universidade do Algarve (UALG)
Applied Molecular Biosciences Unit (UCIBIO)
Portuguese node of the Global Biodiversity Network (GBIF)
Portuguese Coastal Monitoring Network (CoastNET)
Portuguese Biological Data Network (BIODATA.PT)
National Facility for Genome Sequencing and Analysis (GenomePortugal)



Global Biodiversity Network - GBIF



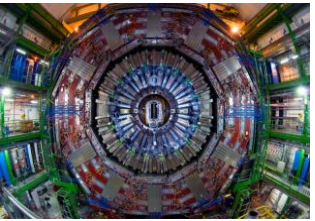
Global Biodiversity Network - GBIF

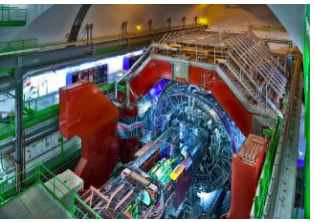


https://dados.gbif.pt

https://datos.gbif.es

Computing for the CERN LHC - Portuguese Tier-2

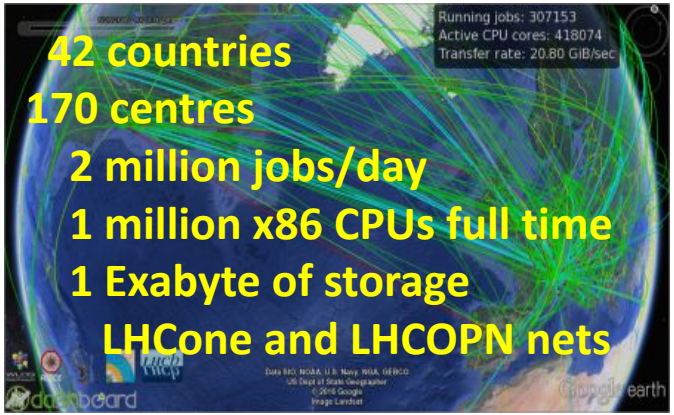




Computing infrastructure for the participation in the CERN LHC.

INCD HTC/CLOUD + EGI/IBERGRID services.

WLCG specific middleware, policies and requirements.



Data, CPU and network intensive

2027

900 PB/year

60x CPU

2016

80 PB/year

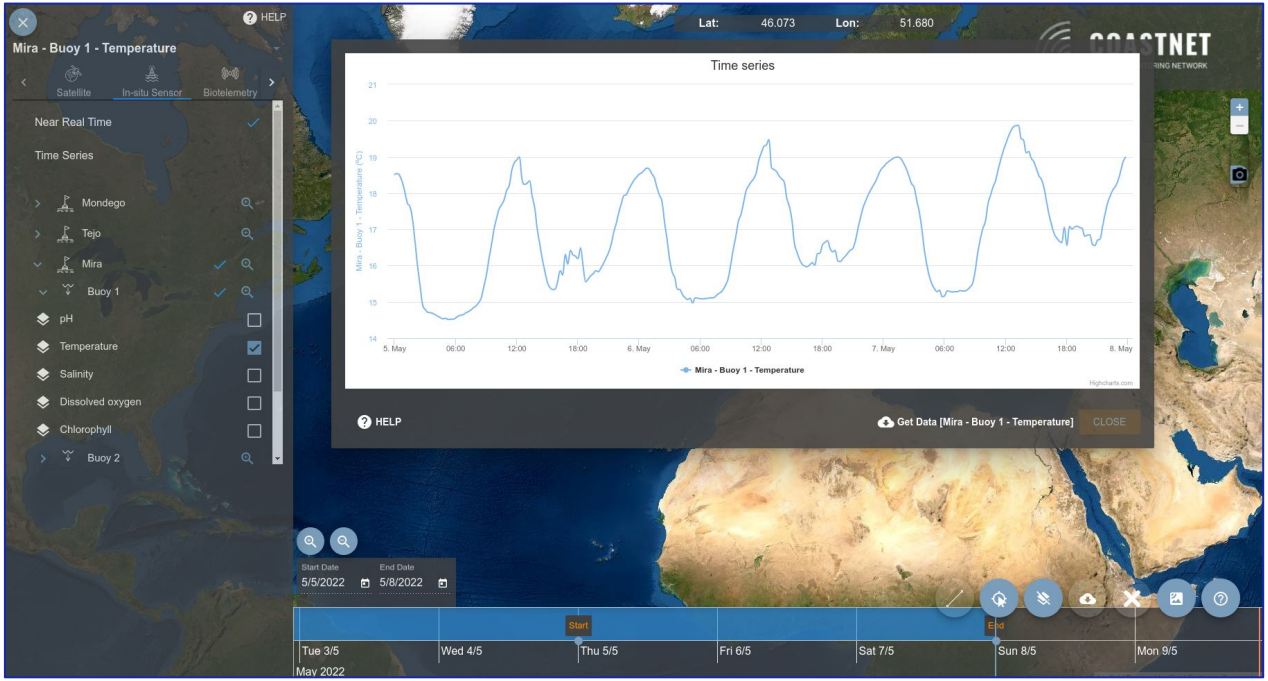
WLCG
Worldwide LHC Computing Grid

Contribution to the global simulation, reconstruction and analysis in the context of the CERN LHC computing MoU.

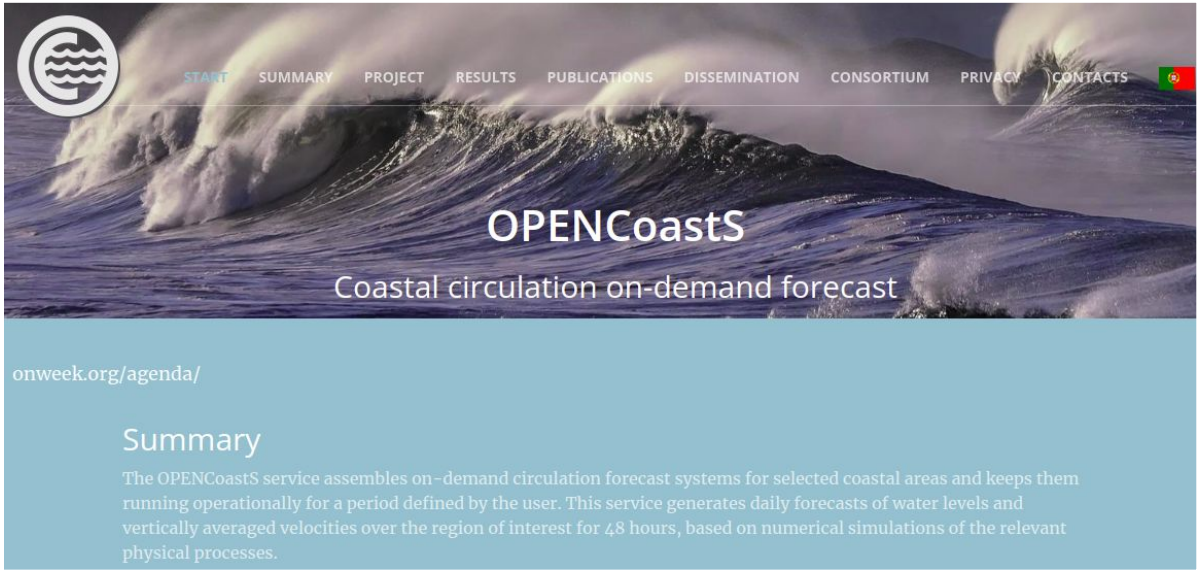
Coastal Monitoring Network - CoastNET

The CoastNet RI assesses and monitors environmental conditions from selected Portuguese coastal systems, ultimately contributing to the sustainable exploitation of resources and biodiversity conservation.

COASTNET
COASTAL MONITORING NETWORK



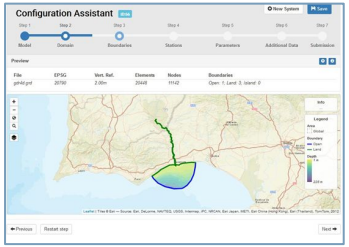
OPENCoastS: Coastal circulation on-demand forecast

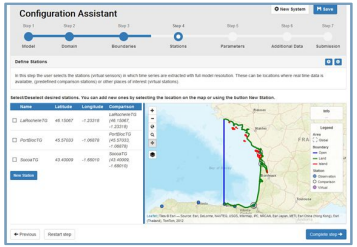


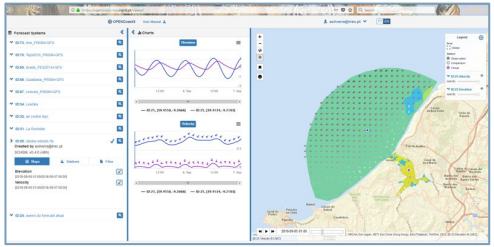
Coastal forecasts on-demand:

- Accurate and timely predictions on water conditions on-demand
- Water levels and velocities, wave characteristics
- Forecasts using the SCHISM model
- Atmosphere forcings:
  - GFS (NOAA)
  - ARPEGE (MétéoFrance)
- Ocean forcings:
  - PRISM2017 (LNEC)
  - FES2014 (LEGOS)
  - CMEMS (Copernicus)
- Wave forcings:
  - North Atlantic WW3

https://opencoasts.ncg.ingrid.pt





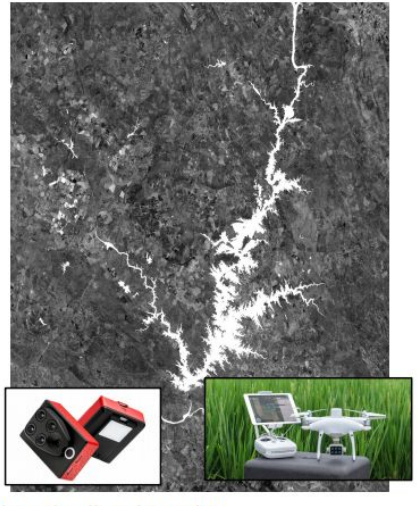


Water mOnitoRing SentInel Cloud platform - WORSICA

WORSICA
Main products

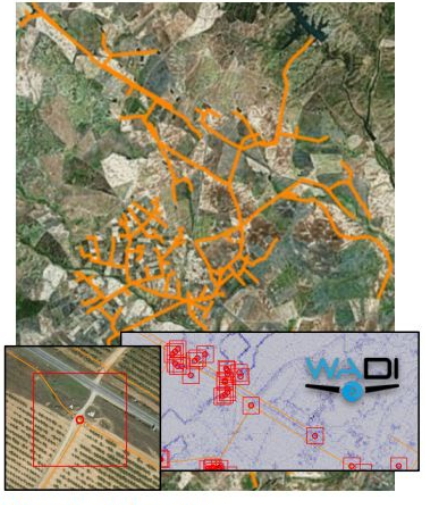
Coastline detection

Detection of water-land interface and possible calculation of the Digital Elevation Model for each line using the EOSC-hub OPENCoastS service and Fassoni et al. (2021) methodology.



Water bodies detection

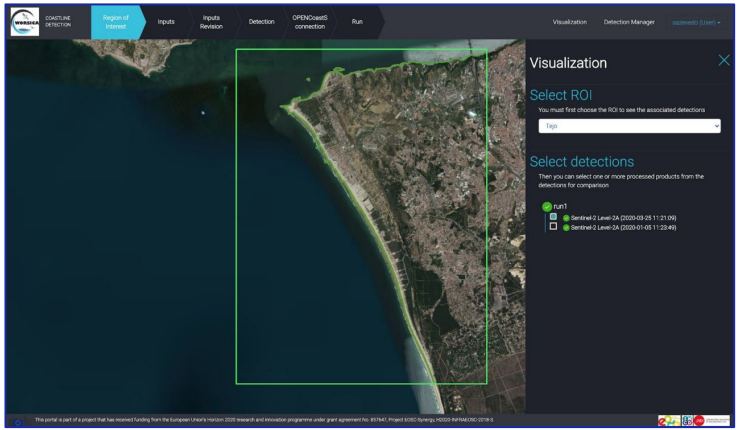
Determination of water indexes to detect water bodies in inland areas (lagoons, reservoirs, etc.), using satellite and drone-based imagery.



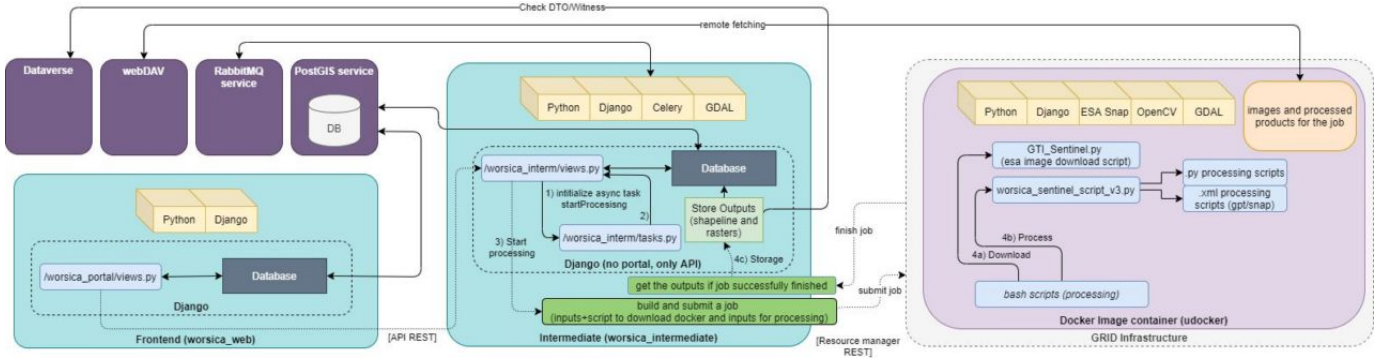
Water leak detection

Take advantage of the work developed in H2020-WADI project (with "low resolution" images from sentinel-2) and try to improve it using Pleiades and drone-based imagery.

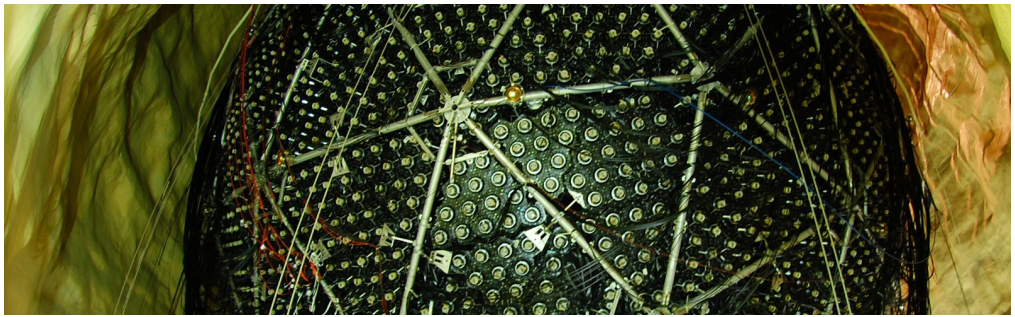
Water mOnitoRing SentInel Cloud platform - WORSICA



- Integrates remote sensing and in-situ data for the determination of water presence in coastal and inland areas. Uses EOSC and EGI.
- Applicable to a range of purposes from the determination of flooded areas (from rainfall, storms, hurricanes or tsunamis) to the detection of large water leaks in major water distribution networks.
- https://worsica.incd.pt



Neutrino Physics



The LIP neutrino physics research group uses the INCD services for simulation and data analysis.

The group participates in SNO+ and joined the DUNE collaboration, that aims to measure one of the missing parameters of neutrino oscillations, the "CP violation phase".

Neutrinos are puzzling elementary particles with neutral electric charge and tiny mass, are among the most abundant particles in the Universe, a billion times more than the matter particles that make up stars and galaxies.

However, they interact with matter very rarely, and as such they are very difficult to detect and study.

The SNO+ detector is an acrylic sphere with 12m diameter containing 1000t of heavy water installed at 2Km depth.

Space Radiation

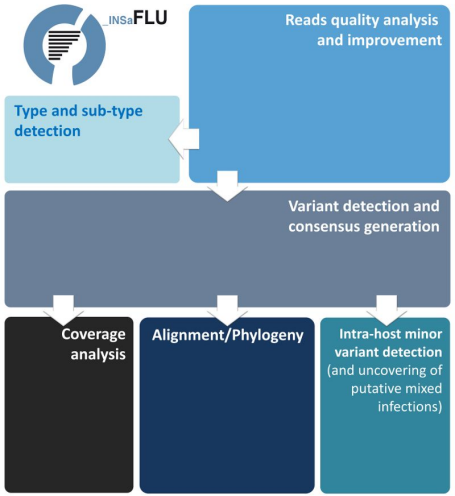
Environment analysis & Modelling
Development and improvement of radiation belt models; development of radiation environment models for specific locations in the solar system;

Radiation Effects Analysis tools Development of tools enabling radiation shielding and effects calculations in the electronic components of satellites and space missions.

Radiation Hardness Assurance Investigation of the effects of radiation on new types of electronic components and in mission and satellite dependent environments.

National Health Institute and INSaFLU

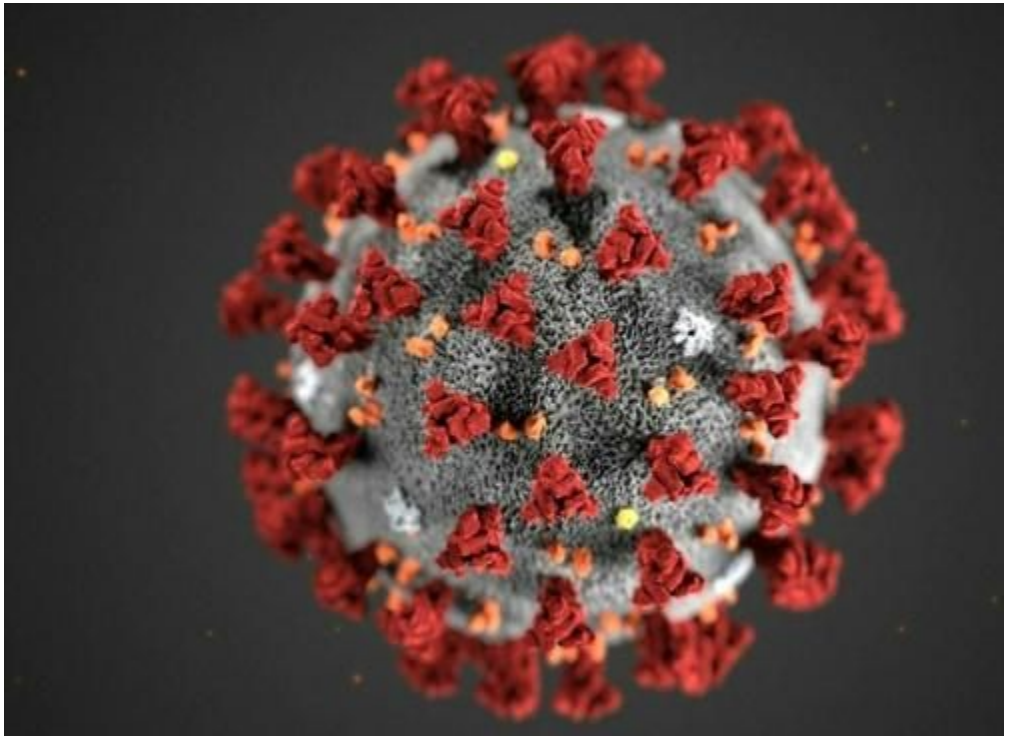
INSaFLU (“INSide the FLU”) is a bioinformatics suite that deals with primary data (reads) towards the automatic generation of the output data that are actually the core first-line “genetic requests” for effective and timely **viral influenza** and **SARS-CoV-2 laboratory surveillance** (e.g., type and sub-type, gene and whole-genome consensus sequences, variants annotation, alignments and phylogenetic trees). Data integration is continuously scalable, fitting the need for a **real-time epidemiological surveillance during the flu and COVID-19 epidemics**.



Acknowledgments

We thank the Infraestrutura Nacional de Computação Distribuída (INCD) for providing computational resources for testing the INSaFLU platform. INCD was funded by FCT and FEDER under the project 22153-01/SAICT/2016

Fighting Covid-19 - UCIBIO



Selection of potential molecules to neutralize Covid-19, using GPU computing at INCD.

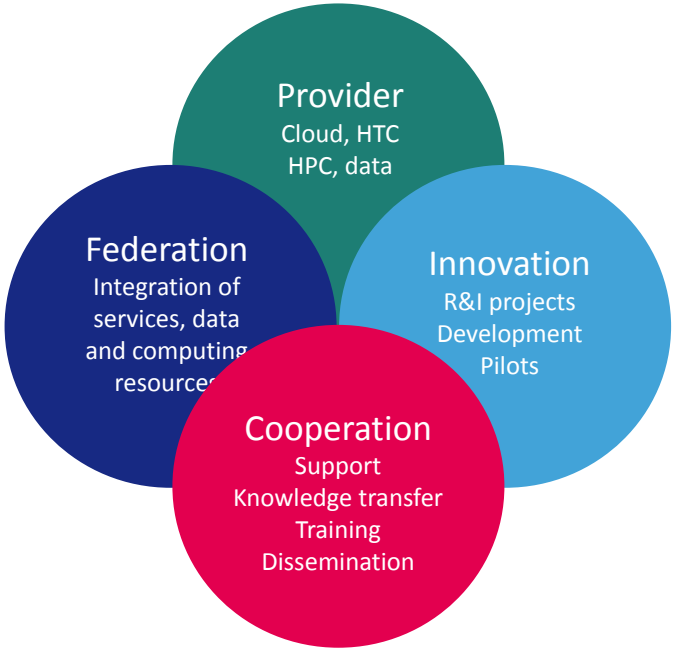
In-silico identification of potential molecules later confirmed in-vitro.

Enabled registration of two patents.

UCIBIO
Applied Molecular Biosciences Unit

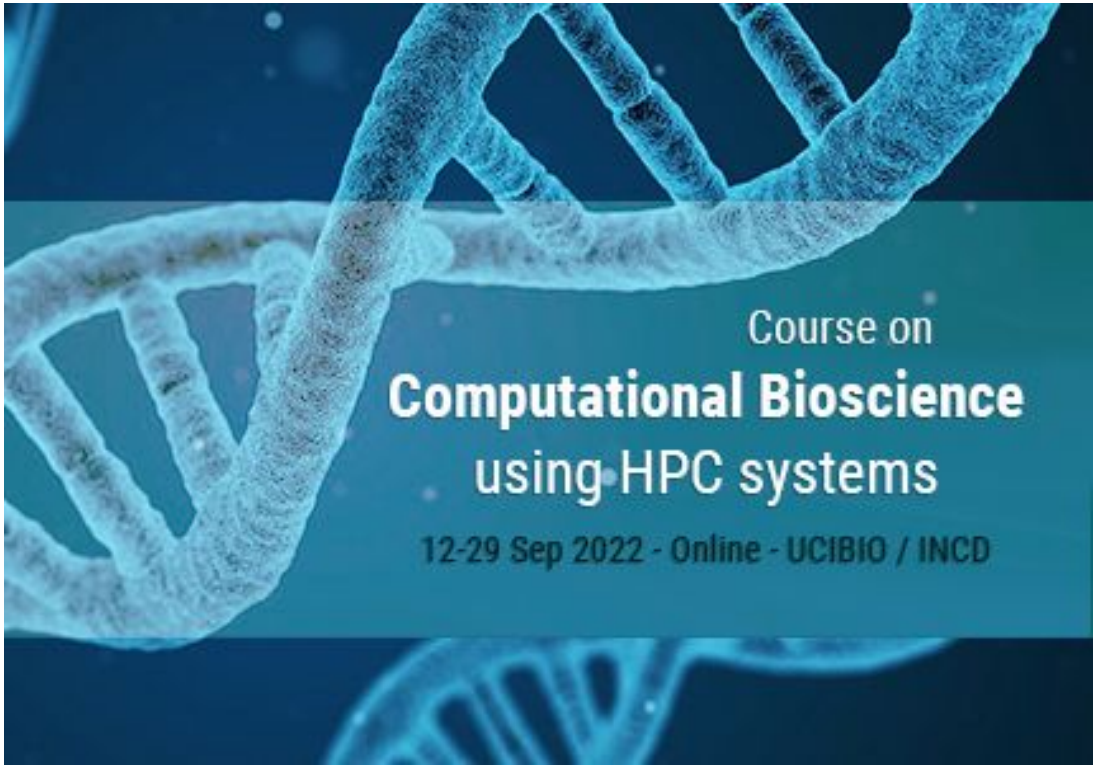
Cooperation beyond usage

- Infrastructures
  - CoastNET
  - BioData
  - Porbiota, LifeWatch, GBIF
- Academic institutions
  - UTAD - Collaborating in the housing of INCD in the North
  - ISA - Management of the GBIF national node
  - ISEC - Collaboration in the management of the local HPC cluster
  - ISEL and ISCTE - Support to MSc training
  - U.Porto, U.Aveiro, U.Minho - Provided LoS for INCD
  - UCIBIO - Joint training actions
- Operational Centres
  - Co-located with LIP at U.Minho, U.Coimbra, U.Lisboa
  - Co-located with FCCN and LNEC
- Other organisations
  - INSA - Support to INSAFLU



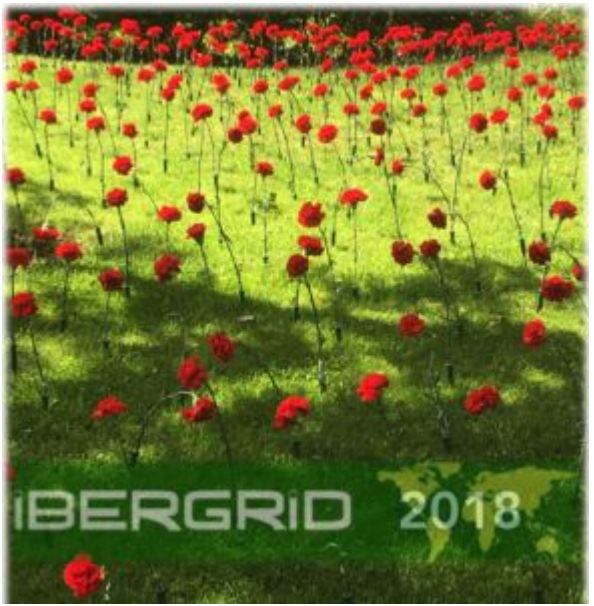
Training on Computational Bioscience

EUROCC PORTUGAL



Sep 29
Curso UCIBIO (Module 6 - Hybrid Quantum Mechanics/Molecular Mechanics)
Sep 28
Curso UCIBIO (Module 5 - Molecular Dynamics)
Sep 27
Curso UCIBIO (Module 4 - Structure-based Virtual Screening)
Sep 14
Curso UCIBIO (Module 3 - Phylogenomics)
Sep 13
Curso UCIBIO (Module 2 - Transcriptome Assembly)
Sep 12
Curso UCIBIO (Module 1 - High-throughput Sequencing Data)

IBERGRID EVENTS



https://www.incd.pt

helpdesk@incd.pt

Thank you

Cofinanciado por:
COMPETE 2020
PORTUGAL 2020
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional
Lisb@20 20
FCT Fundação para a Ciência e a Tecnologia