Computing @ LIP and INCD

Jorge Gomes
LIP Distributed Computing and Digital Infrastructures group

Distributed Computing & Digital Infrastructures

Information and communications technology (ICT) services to support research, innovation, education, outreach and administrative activities at LIP.

Participation in national and international projects, initiatives and digital infrastructures for research.

Services leveraging LIP expertise for the benefit of the national research community, public sector and SMEs.

LIP ICT services

LIP web and **communication support services**

Scientific computing and **digital infrastructures**

Research, development and **Innovation**

Outline

A. **Computing infrastructure for HEP**

B. **Research and innovation projects**

C. **Portuguese Distributed Computing Infrastructure**

The challenge of **Computing for the LHC**

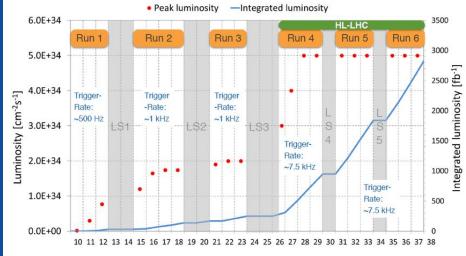LIP participates in the ATLAS and CMS experiments at the CERN Large Hadron Collider (LHC).

The LHC experiments produce unprecedented volumes of research data.

Novel approaches were required for distributed data simulation, processing, analysis and storage at global scale.
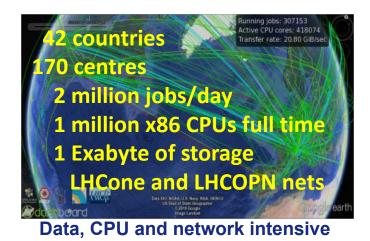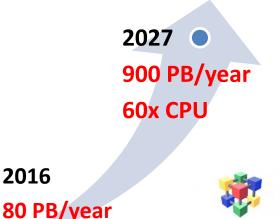
The challenge continues with the increase of luminosity.

Computing for the LHC - Worldwide LHC Computing Grid (WLCG)

Computing infrastructure for the participation in the CERN LHC.

Tier-2 > INCD > IBERGRID > EGI > WLCG > Experiments

Running jobs: 307153
Active CPU cores: 418074
Transfer rate: 20.80 GiB/sec

42 countries
170 centres
2 million jobs/day
1 million x86 CPUs full time
1 Exabyte of storage
LHCone and LHCOPN nets

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.
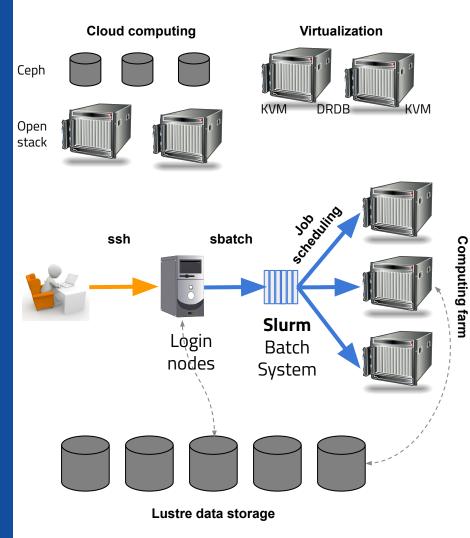
Computing and data
Local infrastructure

Research in High Energy Physics is compute, data and network intensive. The LIP scientific computing services are based on:

- Computing farm based on Slurm
- Data storage based on Lustre
- CVMFS for software sharing
- Virtualization KVM, LXC, Docker with DRDB
- Cloud compute based on Openstack
- Cloud storage based on Ceph

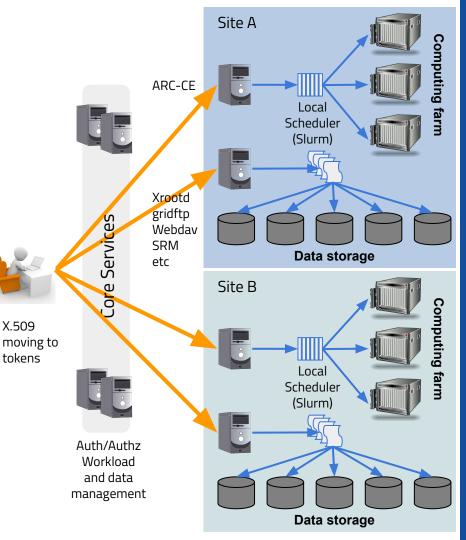These services make use of the National Distributed Computing Infrastructure (INCD).

Cloud computing

Ceph

Open stack

Virtualization

KVM   DRDB   KVM

ssh   sbatch   Job scheduling

Login nodes

Slurm Batch System

Computing farm

Lustre data storage

Computing and data
Going distributed

Accessing compute and data resources across datacenters and organizations is essential for research collaborations such as the LHC.

- Use remote computing capacity
- Access, transfer and manage research data

Federated distributed computing.

- Federated authentication and authorization
- Uniform protocols for resource access
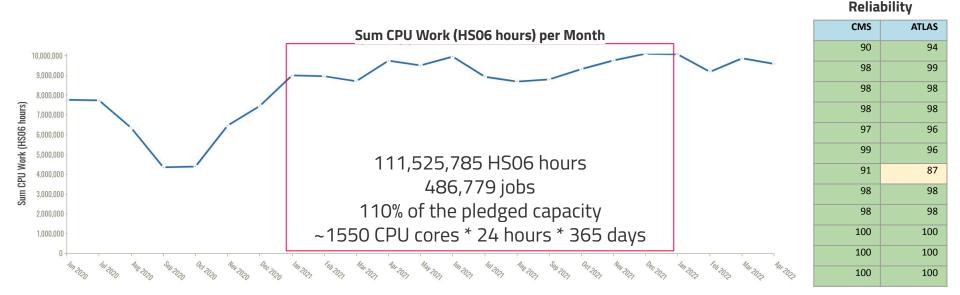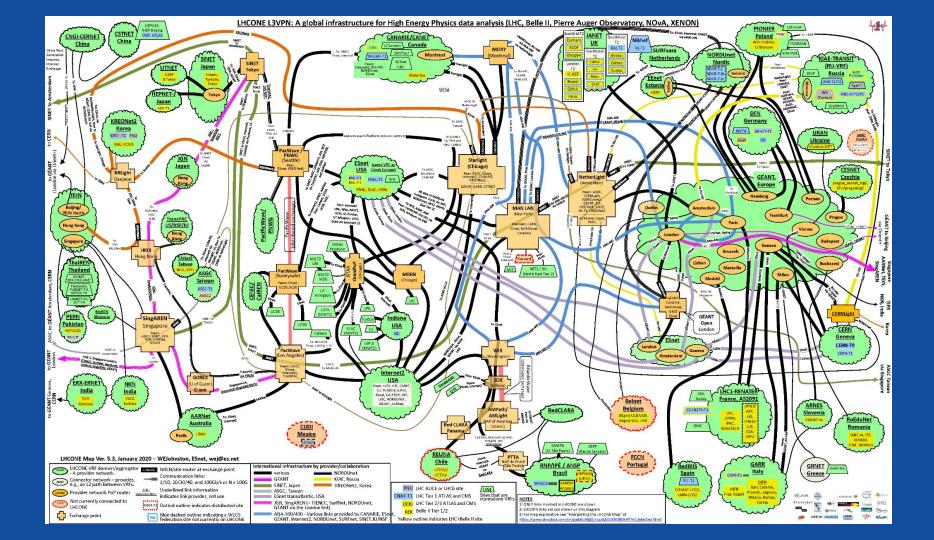- Meta-workload and data management

Computing for the CERN LHC - Portuguese WLCG Tier-2

Each country provides its share of compute and data capacity. LIP operates a Tier-2 computing facility for the ATLAS and CMS experiments at the CERN Large Hadron Collider.

Sum CPU Work (HS06 hours) per Month

111,525,785 HS06 hours
486,779 jobs
110% of the pledged capacity
~1550 CPU cores * 24 hours * 365 days

Reliability

CMS	ATLAS
90	94
98	99
98	98
98	98
97	96
99	96
91	87
98	98
98	98
100	100
100	100
100	100

LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NOvA, XENON)

LHCONE Map Ver. 5.3, January 2020 – WEJohnston, ESnet, wej@es.net
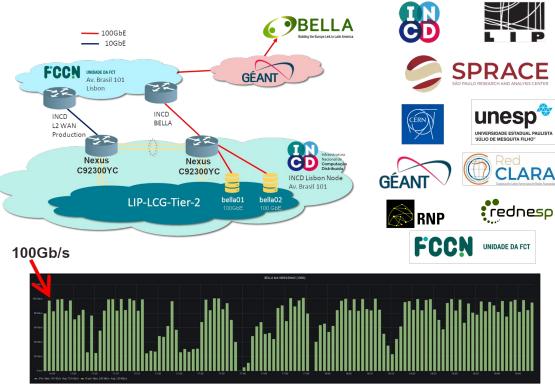
Experimenting with BELLA - on the EllaLink submarine cable

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

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

EuroCC
HPC Competence Center
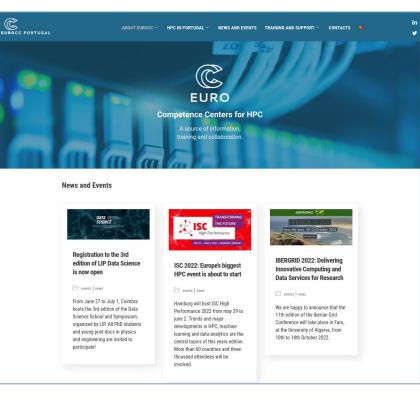
EuroHPC support project (2020-2022-2025)

High Performance Computing National
Competence Center in EuroCC.
- Boost European HPC knowledge
- European network of 33 national HPC
  competence centres (NCCs)
- Bridge the skills gaps and promote cooperation

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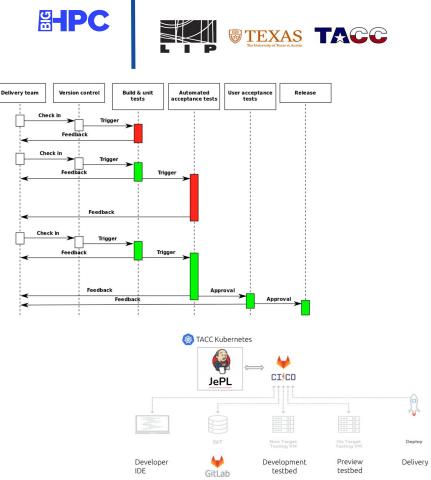- Access to expertise and knowledge

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

BigHPC
A framework for HPC

Portugal-UT-Austin project (2020-2023).

Simplify management of High Performance Computing infrastructures for BigData and parallel applications.

- Novel monitoring, virtualisation and software defined storage components
- Cope with HPC scale and heterogeneity
- Efficiently support different workloads
- Integrated with existing HPC stacks

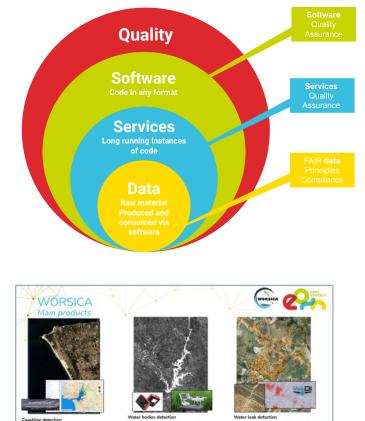LIP participation focused on quality assurance, DevOps, monitoring and containerisation.

EOSC-Synergy
Quality for EOSC

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. LIP participation on:

- Coordination of Quality activities for software, services and data.
- Quality baselines for software & services.
- Development of Jenkins Pipeline Library.
- Support to thematic services integration.
- Data repositories and computing services integration in EOSC.

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.

www.eosc-synergy.eu

EOSC-Synergy – quality work in a nutshell

Inspired by DevOps & Exploiting Continuous Integration and Continuous Delivery (CI/CD)

Provide a Platform to assess the Software, Services and Data Quality in EOSC

Based on Open Source tools & deployable "as a Service"

Enabling exploitation of quality criteria and FAIR principles

Validated with production-level Thematic Services

Earth Observ.

Biomedicine

Astrophysics

Environment

Establishing the foundations to reward Quality Achievements in EOSC

EOSC SYNERGY
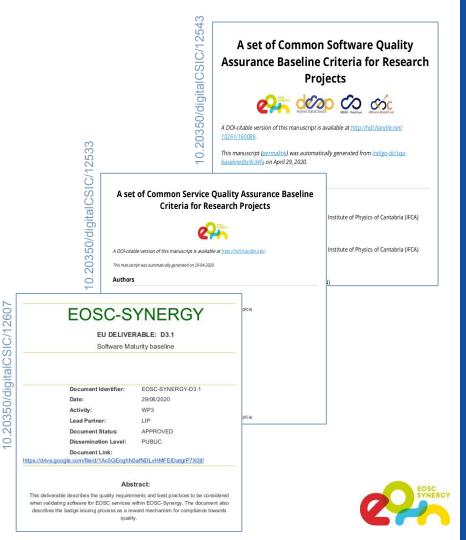
EOSC-Synergy
Quality baselines

Good practices aimed at improving *research* software and services. Set of quality conventions oriented to DevOps.

- Software baseline
  - Criteria meant to enhance the visibility, accessibility and distribution of source code.
  - Encouraging good coding practices to improve quality, reliability and security.
- Services baseline
  - Minimum set of principles for reliable and fit-for-purpose services such as web services, web applications, platforms etc.
  - Provides common coherent quality attributes aimed to ensure functional suitability and strengthening of the services reliability and stability.

The baselines are openly developed on github.

10.20350/digitalCSIC/12543

A set of Common Software Quality Assurance Baseline Criteria for Research Projects

A DOI-citable version of this manuscript is available at http://hdl.handle.net/10261/160086.

This manuscript (permalink) was automatically generated from indigo-dc/sqa-baseline@a9c34fa on April 29, 2020.

Institute of Physics of Cantabria (IFCA)

Institute of Physics of Cantabria (IFCA)

10.20350/digitalCSIC/12533

A set of Common Service Quality Assurance Baseline Criteria for Research Projects

A DOI-citable version of this manuscript is available at http://hdl.handle.net/.

This manuscript was automatically generated on 29-04-2020.

Authors

(IFCA)

(IFCA)

10.20350/digitalCSIC/12607

EOSC-SYNERGY

EU DELIVERABLE: D3.1

Software Maturity baseline

Document Identifier:	EOSC-SYNERGY-D3.1
Date:	29/06/2020
Activity:	WP3
Lead Partner:	LIP
Document Status:	APPROVED
Dissemination Level:	PUBLIC
Document Link:	

https://drive.google.com/file/d/1Ac5GEnghN3afNDLvHMFEiDatgrP7X0jf/

Abstract:

This deliverable describes the quality requirements and best practices to be considered when validating software for EOSC services within EOSC-Synergy. The document also describes the badge issuing process as a reward mechanism for compliance towards quality.

EOSC-Synergy
SQAaaS

Quality Assurance as-a-Service platform (SQAaaS)

- Enables the on-demand creation of CI/CD pipelines making quality verification and validation easily accessible to developers.
  - The **Pipeline as a Service** building block allows you to compose and test customized CI/CD pipelines in accordance with reference criteria.
  - The **Quality Assessment & Awarding** building block analyses, the level of compliance to the quality baselines.
- Integrates a wide range of quality verification tools that are made easily available through a friendly web interface.
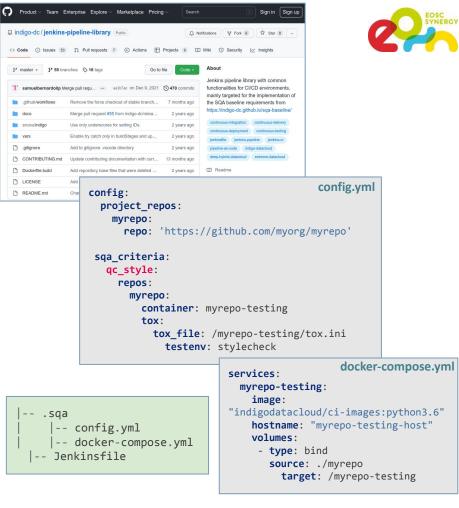
EOSC-Synergy
JePL

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.


```
config.yml
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
```

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

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

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

EOSC-Synergy
SQA as a Service

The SQAaaS will be provided as a cloud service.
Making adoption and usage easier.

- No need to deploy and setup the components, Jenkins, API, web, containers.
- No need to create the yaml configurations.
- No need to provide IT resources.
- No need to manage the platform.

Basis for EOSC quality assessment
- Can issue digital badges to reward and highlight the quality achievements.
- Based on OpenBadges specification.
- Produce detailed quality reports.

https://www.eosc-synergy.eu/technical-areas/quality/

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.
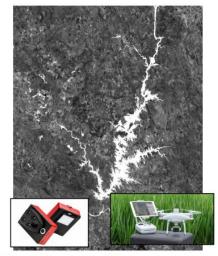
EOSC-Synergy – **Water mOnitoRing SentInel Cloud platform**

- 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
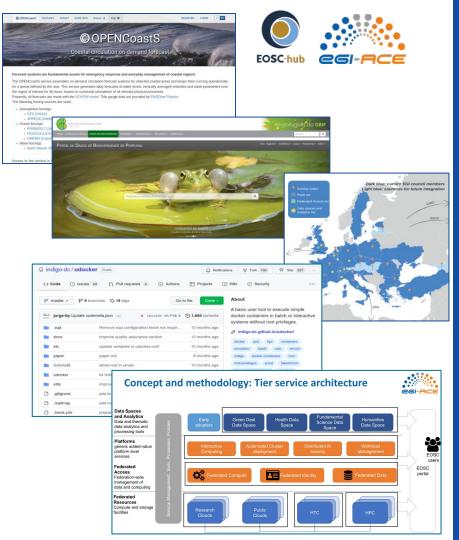
EGI-ACE
Computing for EOSC

European Open Science Cloud project.

EGI-ACE (2021-2023) Advanced Computing for EOSC, with LIP work focused on

- HPC integration, software management, thematic services.

- Provisioning and support of cloud services.

- Software management for EGI.

- Implement EGI software repositories.

EGI-ACE
Coastal forecasts

I N C D

Coastal forecasts on-demand. Accurate and timely predictions on water conditions.

- 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

OPENCoastS
Coastal circulation on-demand forecast
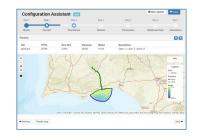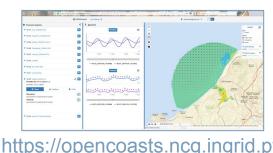
onweek.org/agenda/

Summary

The OPENCoastS service assembles on-demand circulation forecast systems for selected coastal areas and keeps them running operationally for a period defined by the user. This service generates daily forecasts of water levels and vertically averaged velocities over the region of interest for 48 hours, based on numerical simulations of the relevant physical processes.

https://opencoasts.ncg.ingrid.pt

eGI-ACE

EGI-ACE
Biodiversity

Integration of Iberian biodiversity information in the Global Biodiversity Information Facility.

- Joining the Spanish and Portuguese GBIF nodes
- Integration of GBIF nodes in EOSC
- Enabling combined analysis of data
- Implement resiliency for both national nodes

Collaboration with the GBIF nodes of Portugal (ISA) and Spain.

EGI-ACE
udocker

UDOCKER
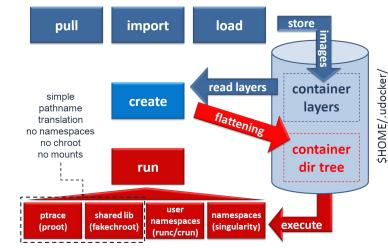
User tool to execute docker containers in user space. Developed at LIP:

- Fully user space.

- No root privileges required to use or install.

- Does not require compilation.

- Download and execution of docker containers by non-privileged users.

- Suitable for Linux batch systems and interactive clusters managed by other entities such as grid infrastructures.

- Does not require Linux namespaces.

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

EGI-ACE
Software quality

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.
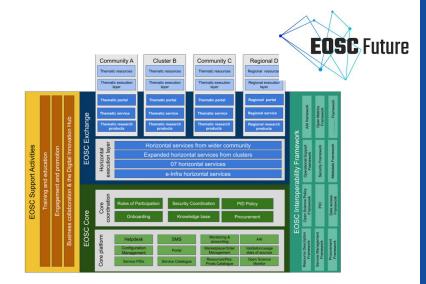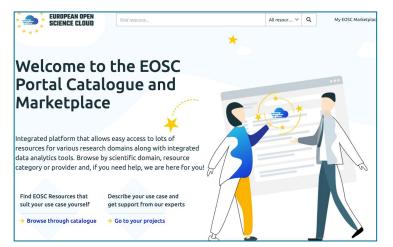 - New repositories with added capabilities.
  - New frontend.

EOSC-Future
IT services for EOSC

European Open Science Cloud project.

EOSC-Future (2021-2023) publish, find and reuse data, tools and services for research.

LIP participation in:

- IT service management for EOSC
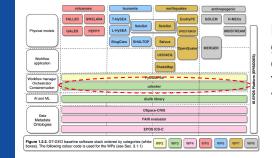- Managing the EOSC services ecosystem.
- Training

New projects (2022-2025)
Looking ahead

New projects starting in September focused on digitalTwins and Artificial Intelligence.

- **DT-GEO** (INFRA-2021-TECH-01)
- **InterTwin** (INFRA-2021-TECH-01)
- **AI4EOSC** (INFRA-2021-EOSC-01)
- **iMagine** (INFRA-2021-SERV-01)
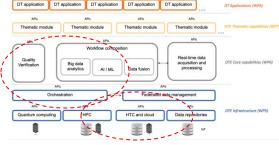
Leveraging experience and IPR from previous projects:
- Quality assurance
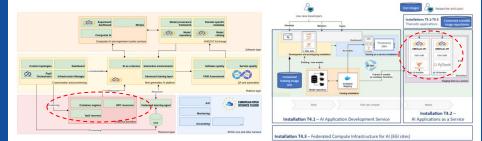- SQA-as-a-Service
- udocker
- DEEP-as-a-Service

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**) and ocean related (**iMagine**)

Infraestrutura Nacional de Computação Distribuída

INCD

- Spin-off of the LIP distributed computing and digital infrastructures group activities.

- Research infrastructure in the FCT roadmap of research infrastructures.

- INCD itself is a legal entity, private non-profit association of LIP, FCT and LNEC.

- Provide compute and data oriented services to PT researchers.

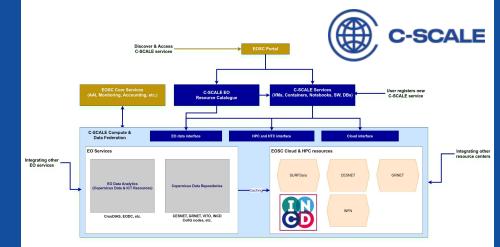- LIP coordinates the technical activities and presides the management board.

Projects

Enable

Support & Requirements

Collaborative Research

Infrastructure

Knowledge

Enhance
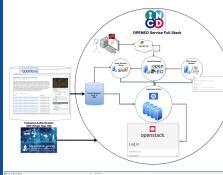
Development

C-Scale
Copernicus

European Open Science Cloud project.

C-Scale (2021-2023) aims to federate European EO infrastructure services:

- Copernicus DIAS and others.
- Capitalise on the European Open Science Cloud (EOSC) capacity and capabilities.
- Support Copernicus research and operations with large and easily accessible European computing environments.

LIP supporting EO use cases and INCD providing cloud computing resources.

Aqua Monitor detects how the Earth's surface water has changed during the last 30 years.

Changes are detected in real-time using satellite imagery for any place on Earth.

Porting the application from the Google Earth Engine platform to the open C-SCALE infrastructure, providing an interactive (zoomable) map that displays land use changes (wet vs dry).

Relies on the top-of-atmosphere reflectance images from Landsat 4,5, 7, and 8 and will be extended to use Sentinel-2 MSI Level-1C data.
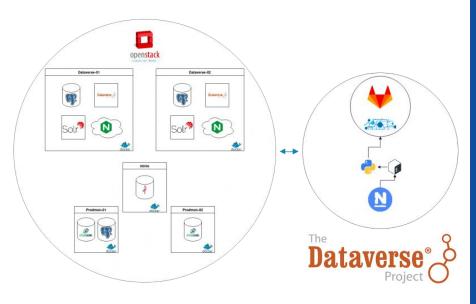
Research data
Repositories

FCT contract, performed under INCD activities.

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, RCTSaai.

- Integration with permanent identifiers.

The **Dataverse**® Project

National Distributed Computing Infrastructure

Services: scientific computing, data processing and data storage

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

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

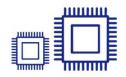Interface: international digital infrastructures (EGI, IBERGRID, WLCG, EOSC)

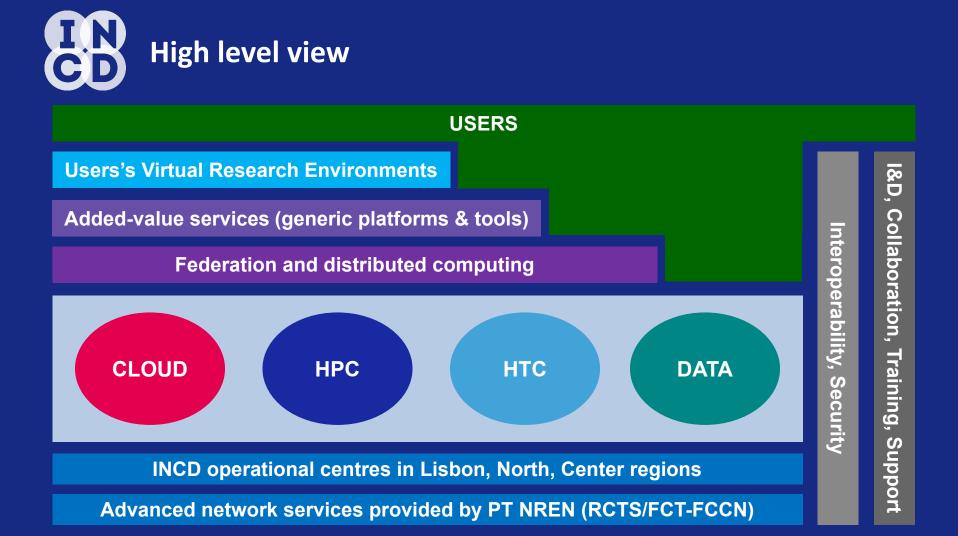**Cloud Computing**
cloud computing

HTC Computing
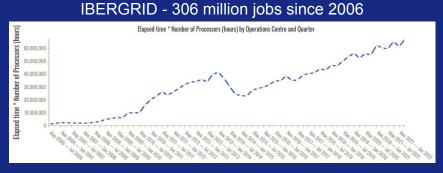high throughput computing (GRID)

HPC Computing
high performance computing

High level view

USERS

Users's Virtual Research Environments

Added-value services (generic platforms & tools)

Federation and distributed computing

CLOUD

HPC

HTC

DATA

INCD operational centres in Lisbon, North, Center regions

Advanced network services provided by PT NREN (RCTS/FCT-FCCN)

Interoperability, Security

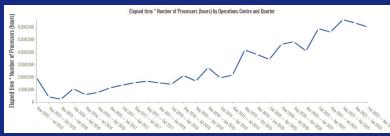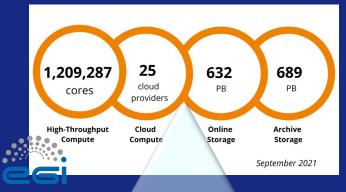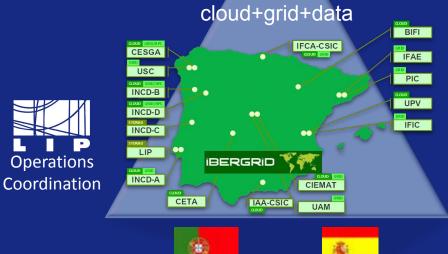I&D, Collaboration, Training, Support

Federation in EGI and IBERGRID

- **EGI** is a federation that joins national and regional infrastructures.
- **IBERGRID** is an Iberian federation.

IBERGRID - 306 million jobs since 2006

IBERGRID - 1 million instantiated VMs since 2015

1,209,287 cores	25 cloud providers	632 PB	689 PB
High-Throughput Compute	Cloud Compute	Online Storage	Archive Storage

September 2021

cloud+grid+data

LIP
Operations
Coordination

INCD operational centers

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

North Region I
HTC / HPC

Center Region
(NEW TO BE DEPLOYED)
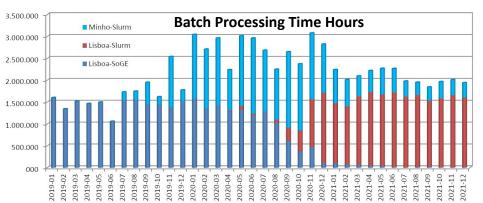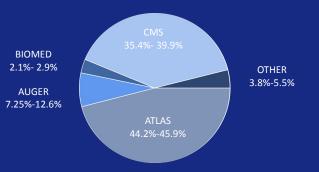Tape storage expansion

Lisbon Region II
Tape storage

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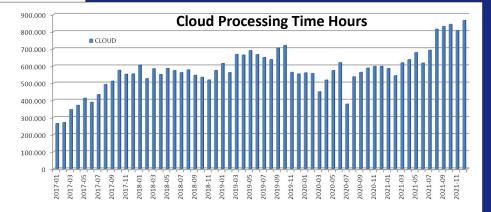HPC / HTC / Cloud / Federation

Usage HTC, HPC and cloud

Batch Processing Time Hours
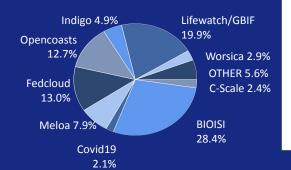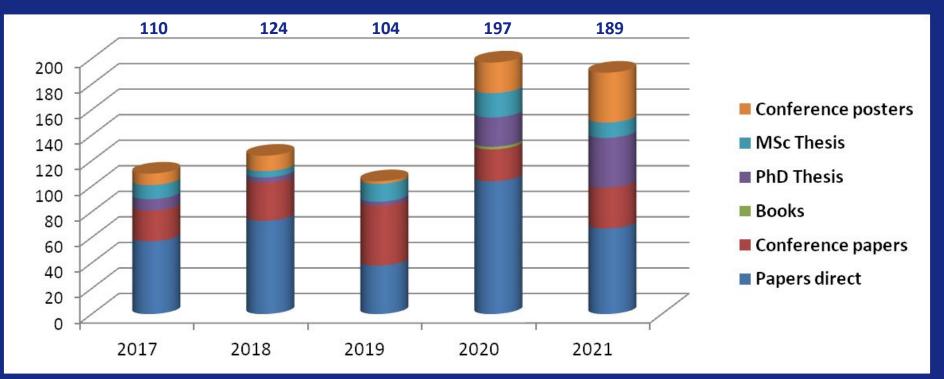
Cloud usage in 2021 over 8.523.000 hours

Pie chart: CMS 35.4%- 39.9%, BIOMED 2.1%- 2.9%, AUGER 7.25%-12.6%, ATLAS 44.2%-45.9%, OTHER 3.8%-5.5%

Batch usage in 2021 over 24.789.000 hours

Pie chart: Indigo 4.9%, Opencoasts 12.7%, Fedcloud 13.0%, Meloa 7.9%, Covid19 2.1%, BIOISI 28.4%, C-Scale 2.4%, OTHER 5.6%, Worsica 2.9%, Lifewatch/GBIF 19.9%

Cloud Processing Time Hours

Publications

Other metrics

- Curated datasets: 2
- Patents: 2

- FCT and international funded projects supported: 68
- CPCA supported projects: 50
- Artificial intelligence in public administration projects: 2
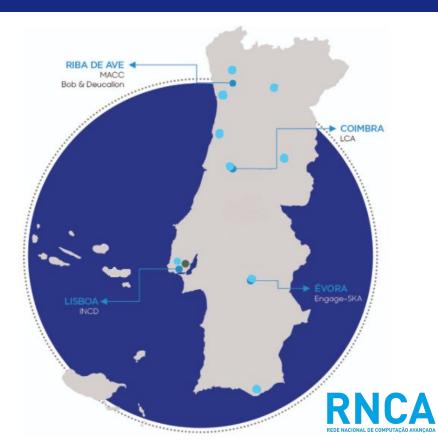- organizations, research units and infrastructures: 62

IBERGRID 2022

Faro, University of Algarve 10th to 13th October 2022

- Cooperation between Iberian research communities
- Research applications in advanced Digital Infrastructures
- Development of innovative software services
- R&D for computing services, networking, and data-driven science
- Quality of software, services and data
- Design and implementation of Digital Twins
- Enabling and fostering Open Science adoption in EOSC

- European Open Science Cloud day
- Training and tutorials

Call for abstracts is open !