LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Advanced and Distributed Computing and Infrastructures

Jorge Gomes - Mario David
on behalf of the group

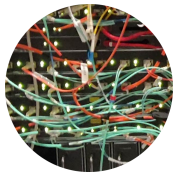
Distributed Computing & Digital Infrastructures

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**

Carlos Manuel	Technician	LIP Lisboa - Web development, design, events, multimedia, communication
Catarina Gonçalves	Communication	LIP Lisboa - Communication, dissemination, outreach
Hugo Gomes	Technician	LIP Lisboa - Web development, IT support, events, multimedia, communication
João Martins	Researcher	LIP Lisboa - Fabric management, storage systems, computing farms, virtualization
João Pina	Researcher	LIP Lisboa - WLCG Tier-2, software management, user support, farm
Jorge Gomes	Researcher	LIP Lisboa - Projects management, computing, development and networks
José Aparício	Engineer	LIP Lisboa - User support, datacenter, networks, notebooks, desktops
Mário David	Researcher	LIP Lisboa - Cloud computing, containers, quality assurance, development, GPUs
Miguel Viana	Engineer	LIP Minho - HPC, monitoring, HPC tools, software integration and validation
Nuno Dias	Researcher	LIP Lisboa - Security, networks, data protection, network services, desktops
Samuel Bernardo	Engineer	LIP Lisboa - Software development & quality, AAI, cloud, federated services
Zacarias Benta	Engineer	LIP Minho - HPC, computing farms, virtualization, user support
Catarina Ortigão	Administration	INCD - administrative and managerial support
César Ferreira	Engineer	INCD - HPC, computing farms, virtualization, user support
NEW CONTRACT	Researcher	INCD – PhD to be contracted in 2022
António Esteves	Researcher	University of Minho - application of machine/deep learning techniques
António Pina	Researcher	University of Minho - application performance analysis, parallel programming
José Rufino	Researcher	Polytechnic Institute of Bragança - parallelization strategies for GPU algorithms
Pedro Barbosa	MsC	Integration of local/remote accelerators in parallel and distributed applications with HMX
José Pinto	MsC	HpxTrace: application monitoring in HMX

Infrastructures

IT services
LIP

General IT services

- Network
- Virtualization platform
- Authentication & Authorization Infrastructure
- IT administrative support services
- Web, design and development

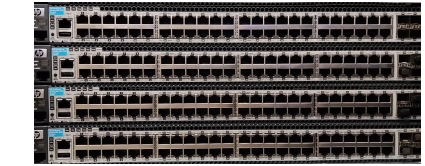
LIP Lisbon IT services

- Login servers
- Home directories
- Desktops
- Data backups
- Mail
- Multimedia
- Security
- Printers

New firewall (being deployed)

New mail (being installed)

New virtualization



Network improvements (ongoing)

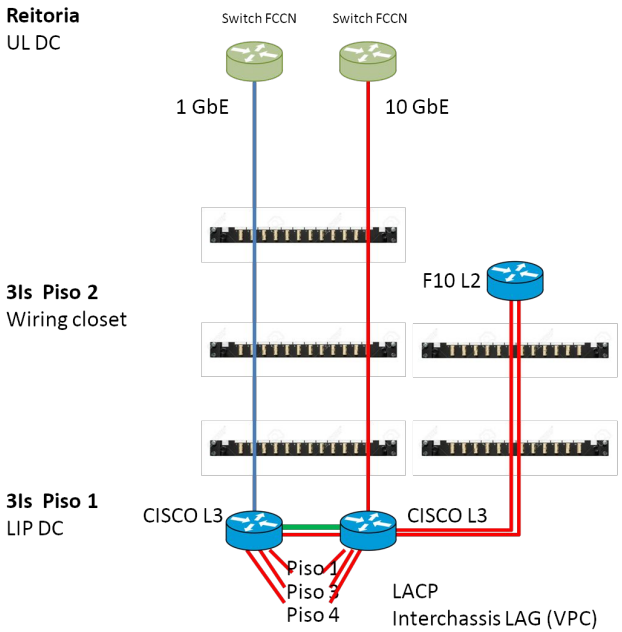
Datacenters
LIP

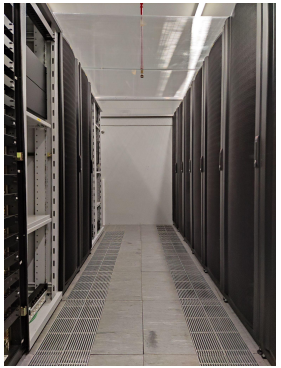
Reitoria-U.Lisboa
- Main LIP core and Lisboa services
- Login servers, homes, web, AAI, etc
- Plan consolidation at the DC @LIP-Lisboa

LIP-Lisboa (3Is)
- Redundant services
- Data backups
- SPAC group services
- Consolidate LIP Lisboa services

Physics-U.Coimbra
- To be reorganized
- Expand data backups capacity
- Consolidate LIP Coimbra services

LIP-U.Minho
- LIP minho cluster
- CPU/GPU platform machine learning

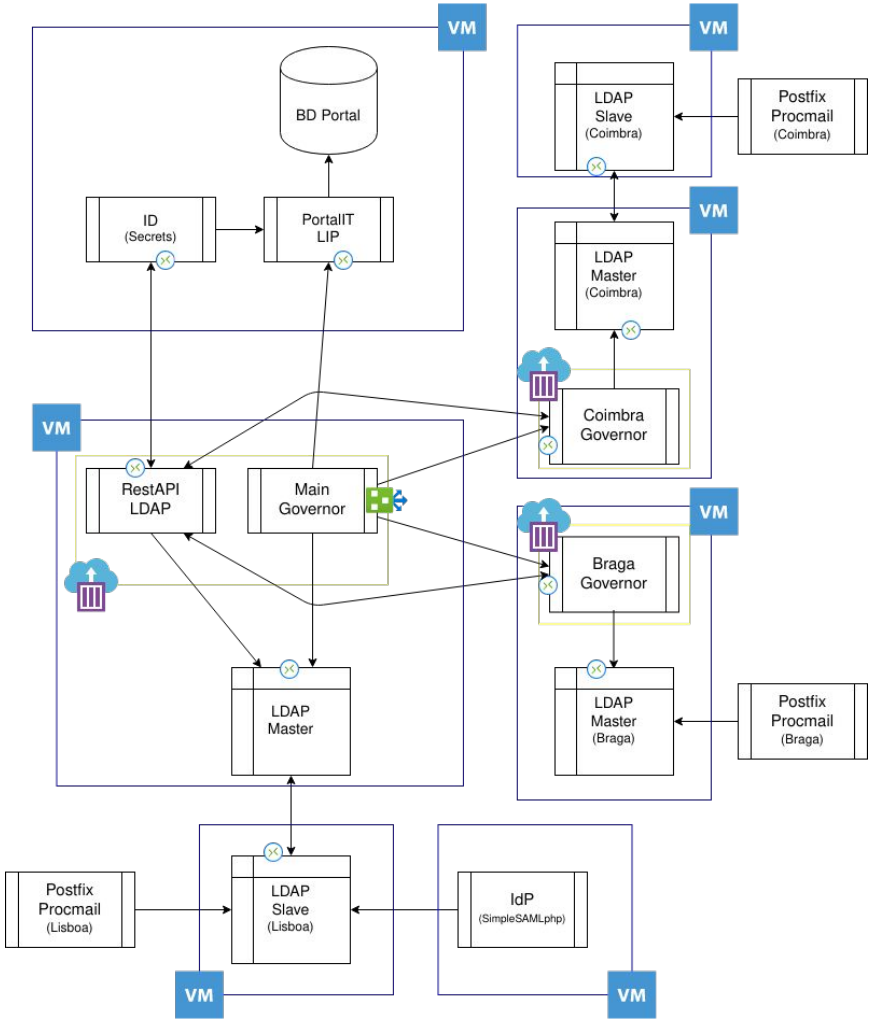




AAI
LIP

Authentication, Authorization Infrastructure

- LDAP authentication
  - (Lisbon + Braga) + Coimbra
- Federated identity
  - Identity Provider (IdP)
  - Attribute Provider (AP)
  - SAML + OIDC
  - RCTSaai + EDUGAIN
- IT portal
- Integrated with LIP projects DB
- Extending the AAI to the web apps



Computing and data **LIP**

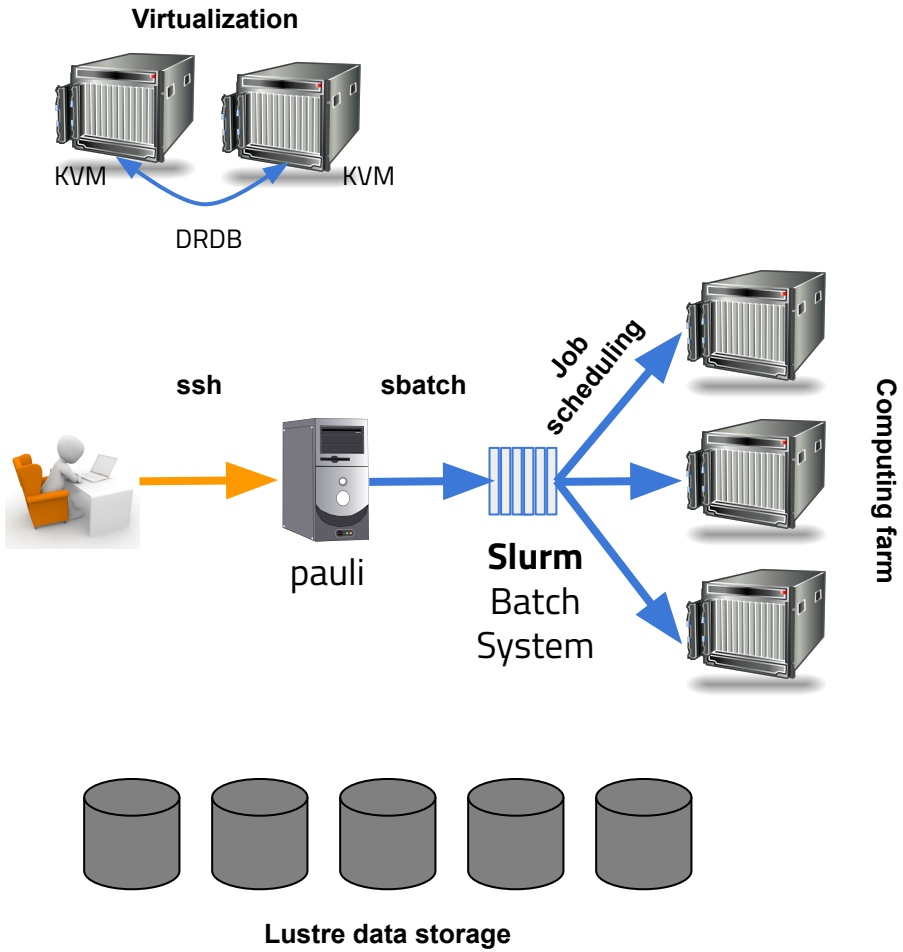
Computing and data services

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

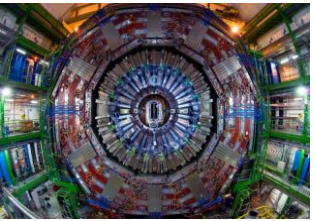
Access

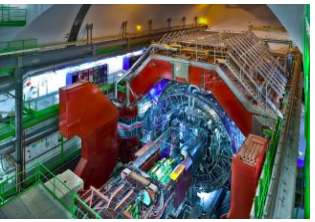
- Via front-end: pauli.ncg.ingrid.pt
- Open to all LIP users: helpdesk@lip.pt

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



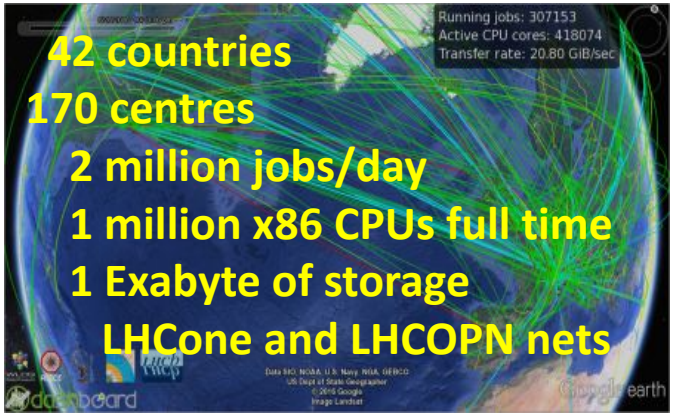
Computing for the LHC - Worldwide LHC Computing Grid (WLCG)





Computing infrastructure for the participation in the CERN LHC.

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



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.

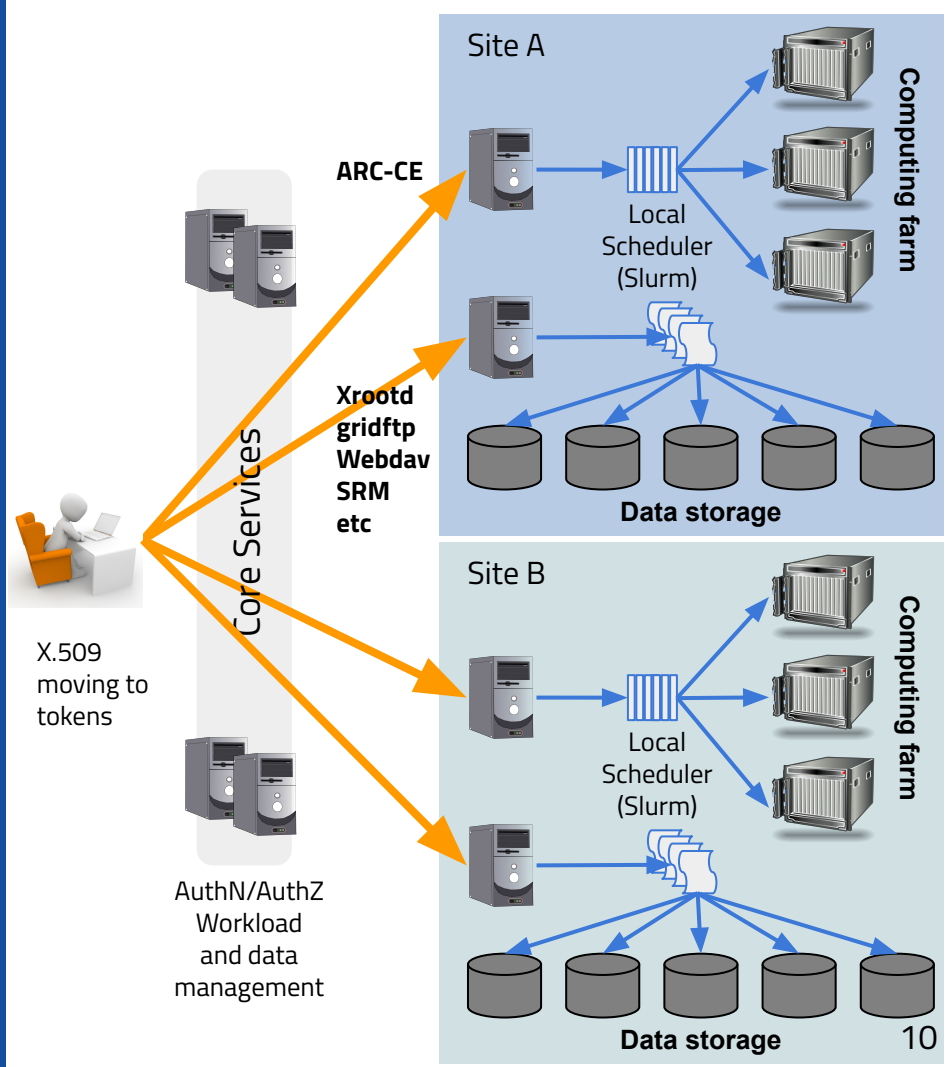
Computing and data
Going distributed

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

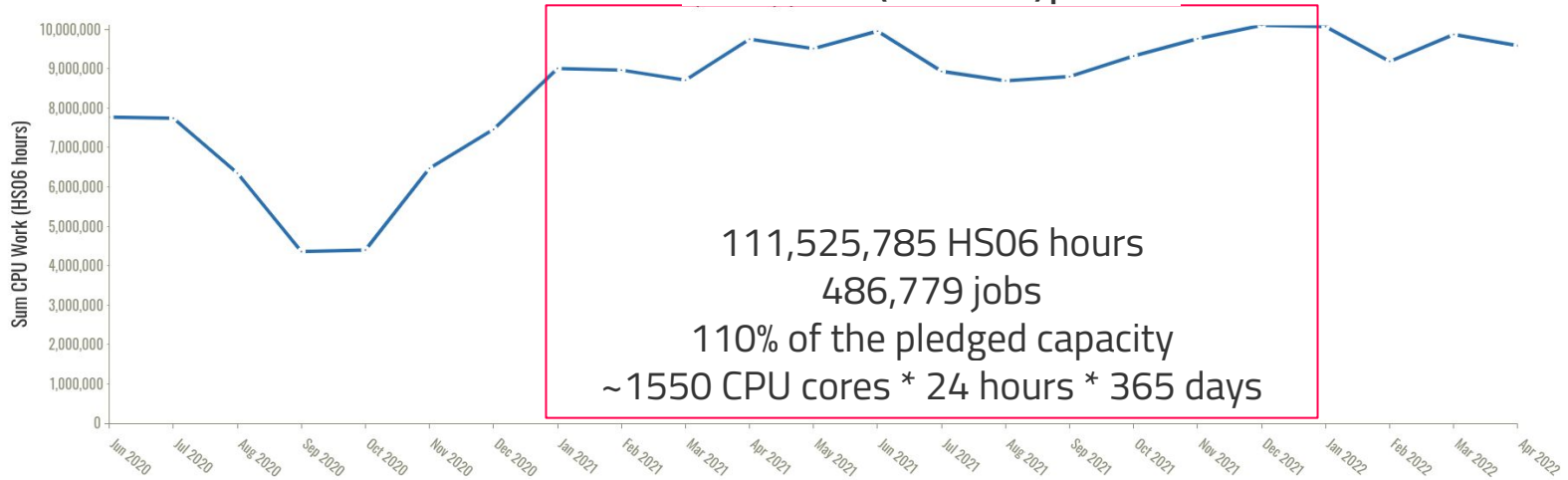
Federated distributed computing.

- 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.



Reliability 2021

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

Computing for the CERN LHC - Portuguese WLCG Tier-2

Country	Jun 2021 — Aug 2021	Sep 2021 — Nov 2021	Dec 2021 — Feb 2022	Mar 2022 — May 2022	Total	Percent
Switzerland	26,764,237	35,276,543	26,067,043	18,866,119	106,973,942	23.51%
United States of America	20,637,753	20,760,419	20,507,281	19,626,922	81,532,375	17.92%
United Kingdom	14,033,626	16,078,849	12,404,142	12,464,924	54,981,541	12.08%
Germany	11,562,916	13,095,282	12,966,196	10,837,334	48,461,728	10.65%
France	8,717,336	8,913,951	8,703,193	10,423,810	36,758,290	8.08%
Italy	8,100,539	6,455,174	5,717,242	4,705,788	24,978,743	5.49%
Canada	5,904,419	5,172,082	5,098,775	4,609,148	20,784,424	4.57%
Russia	2,950,712	2,882,827	2,437,252	2,186,181	10,456,972	2.3%
Sweden	2,292,220	2,868,302	2,068,587	1,887,089	9,116,198	2%
Netherlands	1,944,436	2,201,182	2,278,185	1,974,558	8,398,361	1.85%
Spain	1,940,425	1,936,920	1,560,490	1,714,132	7,151,967	1.57%
Japan	1,748,091	2,025,907	1,019,980	2,248,549	7,042,527	1.55%
Romania	2,162,672	1,475,544	1,321,955	1,772,525	6,732,696	1.48%
Slovenia	1,793,162	1,375,932	1,556,628	1,075,689	5,801,411	1.27%
Czech Republic	1,415,636	1,261,344	1,091,499	1,971,795	5,740,274	1.26%
Poland	1,091,459	1,041,839	939,762	811,997	3,885,057	0.85%
Israel	753,324	854,959	789,549	513,781	2,911,613	0.64%
Taiwan	201,400	550,642	763,961	558,910	2,074,913	0.46%
Slovakia	471,766	585,137	429,234	422,778	1,908,915	0.42%
Brazil	562,115	394,468	431,629	442,359	1,830,571	0.4%
China	193,662	404,078	423,120	464,822	1,485,682	0.33%
South Korea	337,074	397,914	385,407	331,185	1,451,580	0.32%
India	636,713	350,943	208,102	84,463	1,280,221	0.28%
Belgium	176,973	113,108	110,417	77,933	478,431	0.11%
Chile	112,564	100,863	94,227	176,019	483,673	0.11%
Portugal	136,705	103,577	123,007	75,394	438,683	0.1%
Hungary	89,711	84,452	115,213	121,121	410,497	0.09%
Australia	111,510	0	22,662	189,645	323,817	0.07%
Estonia	65,598	85,094	103,960	70,553	325,205	0.07%
Azerbaijan	30,165	67,238	63,840	94,365	255,608	0.06%
Total	117,092,713	127,089,974	109,983,811	100,913,789	455,080,287	

Computing and data **Upgrades**

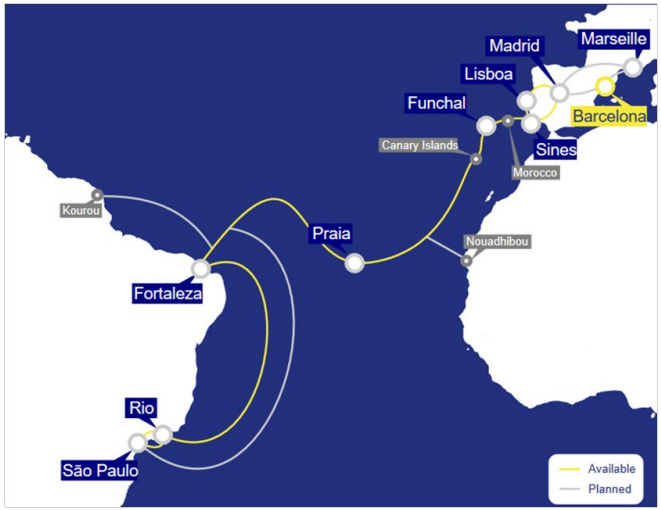
Submitted a proposal for a small capacity upgrade to the CERN fund.

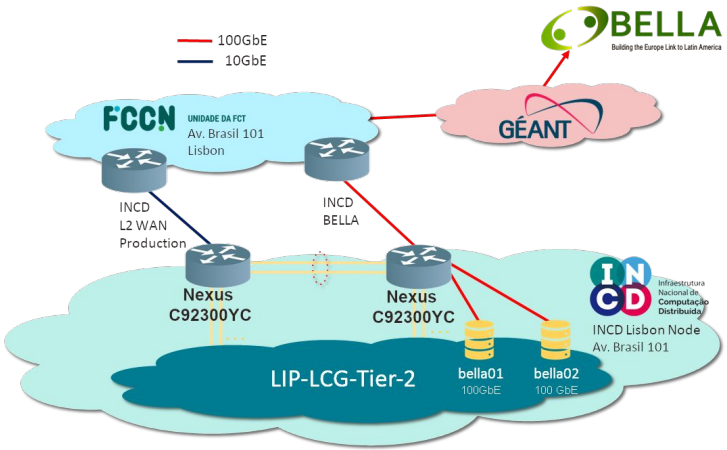
- Mainly for storage to address to most pressing needs.
- Requested ~ 130.000€
- Got 29.000€
- Acquired 576TB

ATLAS trigger farm offered equipment.

- Server enclosures
- Network switches

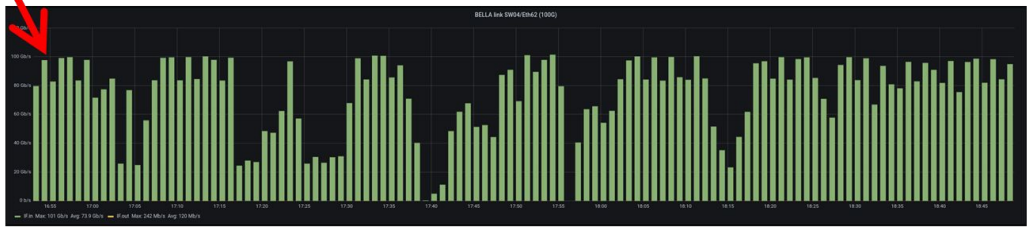
Experimenting with BELLA - on the EllaLink submarine cable





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

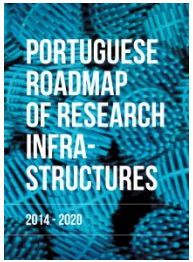
100Gb/s

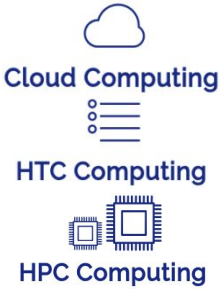


Using 2x INCD servers and 2x SPRACE servers (with 100GbE interfaces)

Infrastructures

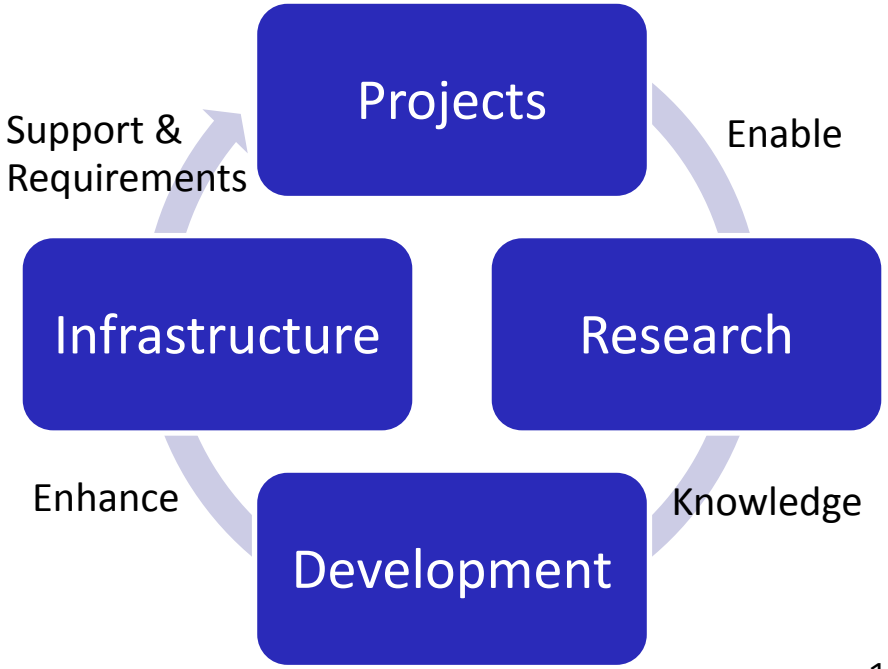
Infraestrutura Nacional de Computação Distribuída





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.



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

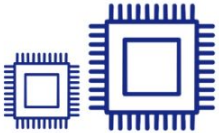
Interface: international digital infrastructures (EGI, IBERGRID, WLCG, EOSC)

Cloud Computing

cloud computing

HTC Computing

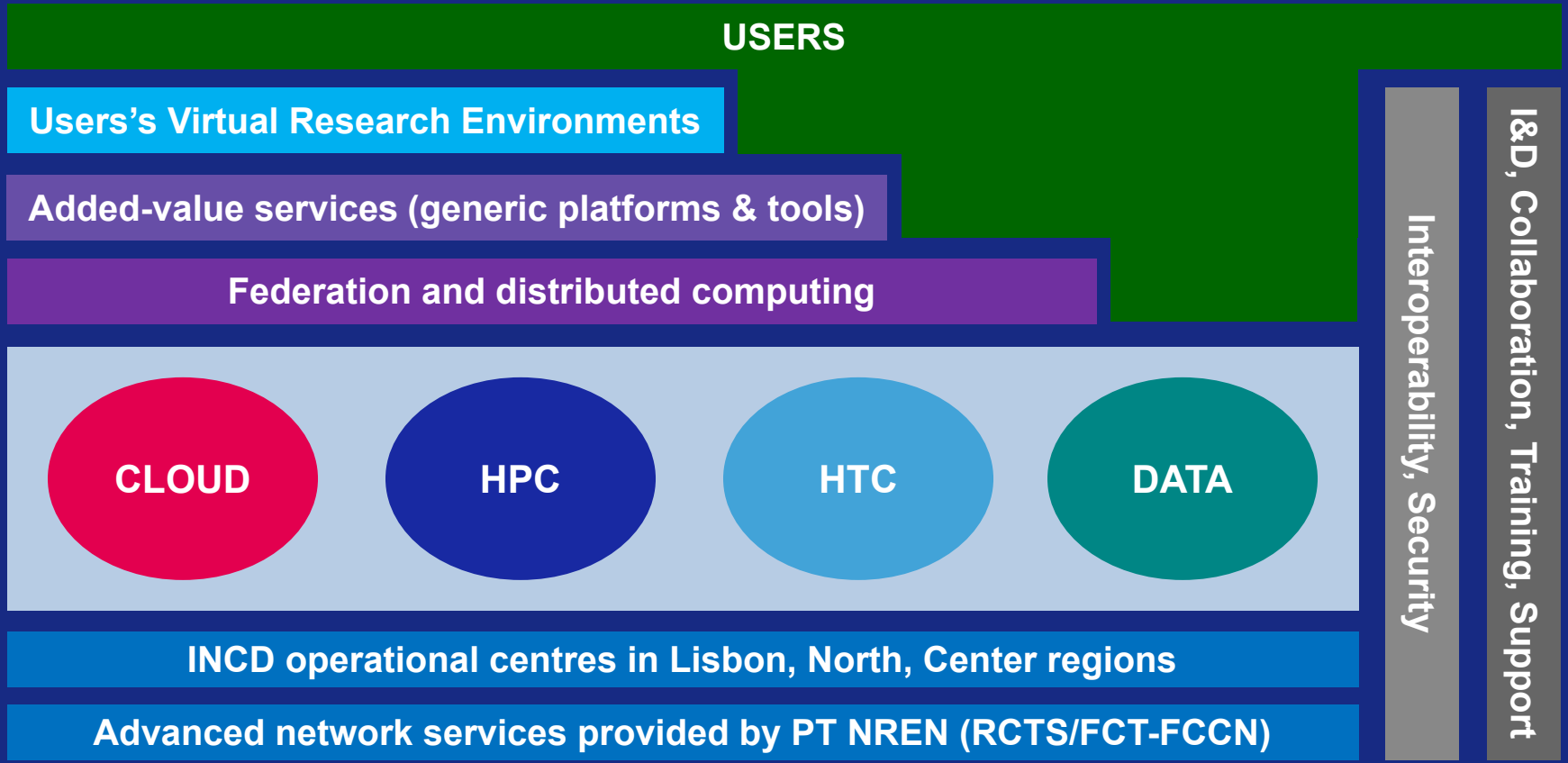
high throughput computing (GRID)



HPC Computing

high performance computing

High level view



INCD operational centers

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

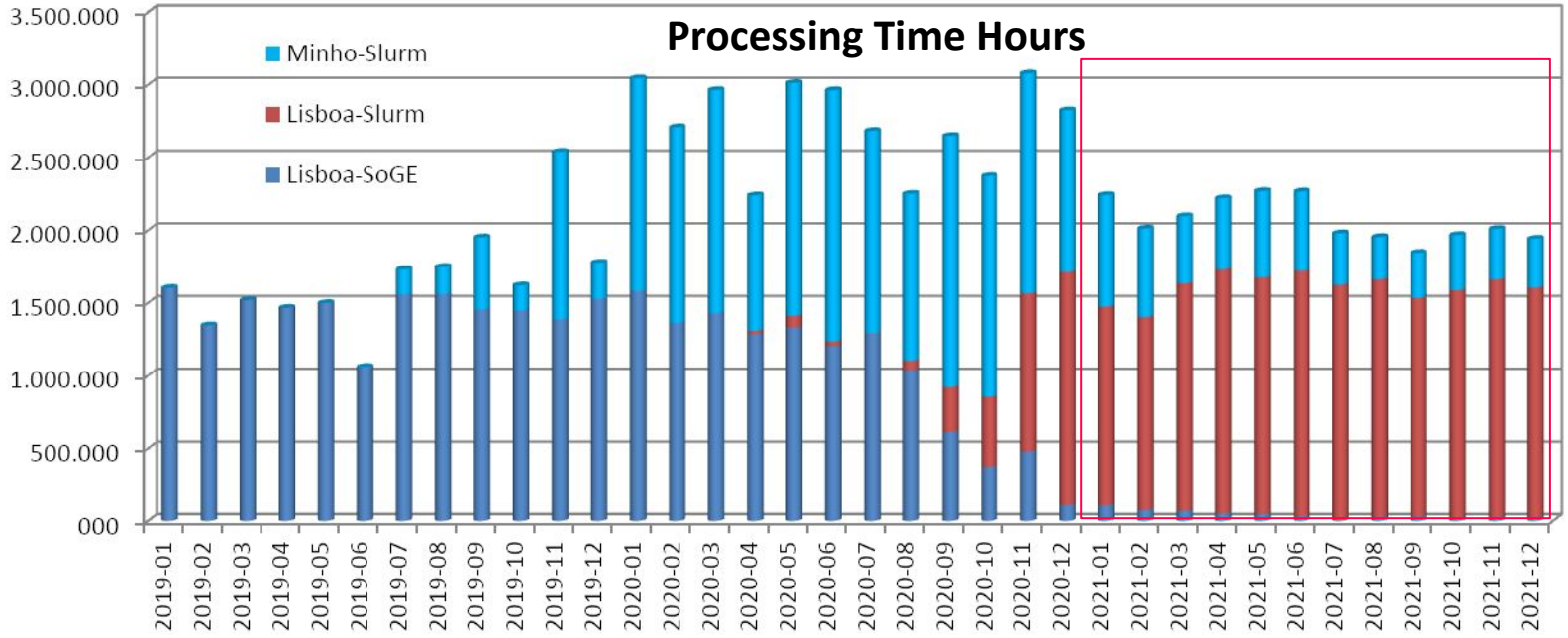
North Region I
BOB @ Riba de Ave
HTC / HPC

Center Region
Physics @ UC
(NEW TO BE DEPLOYED)
Tape storage expansion

Lisbon Region II
LIP @ 3Is
Tape storage

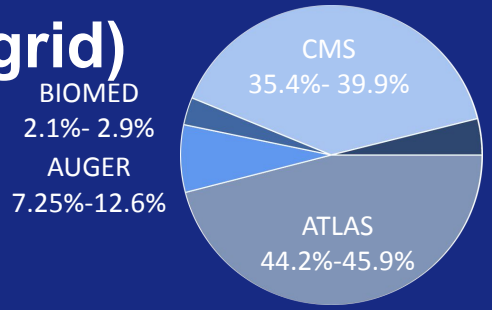
North Region II
UTAD @ Vila Real
(NEW TO BE DEPLOYED)
HPC / HTC / Cloud / Federation

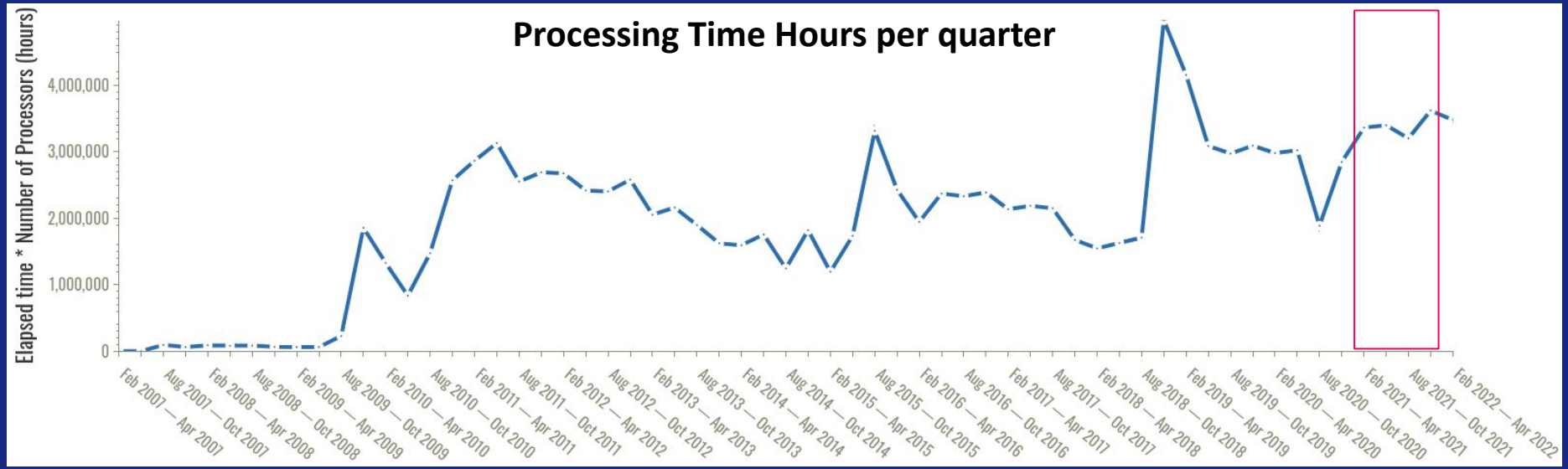
Batch clusters usage both HTC and HPC



Total usage in 2021 over 24.789.000 hours

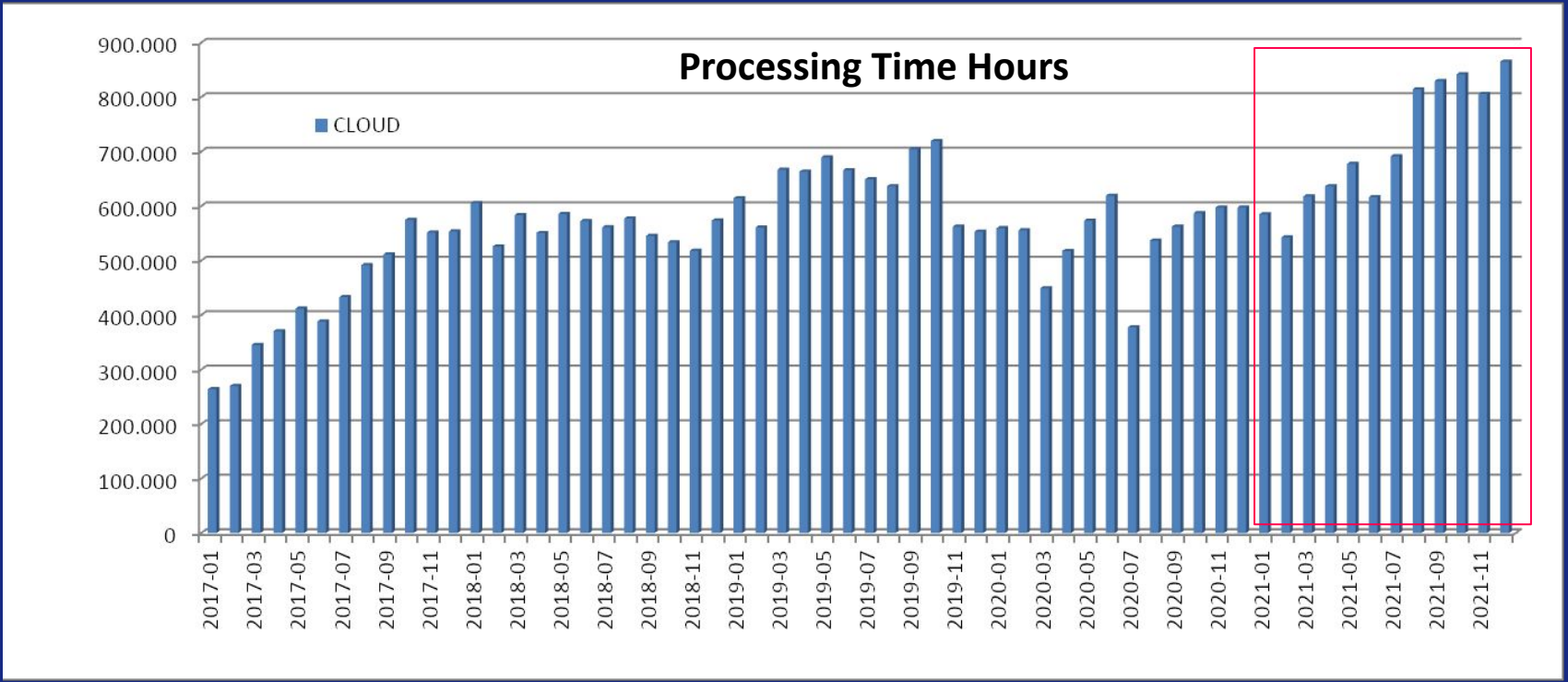
Distributed computing usage (grid)





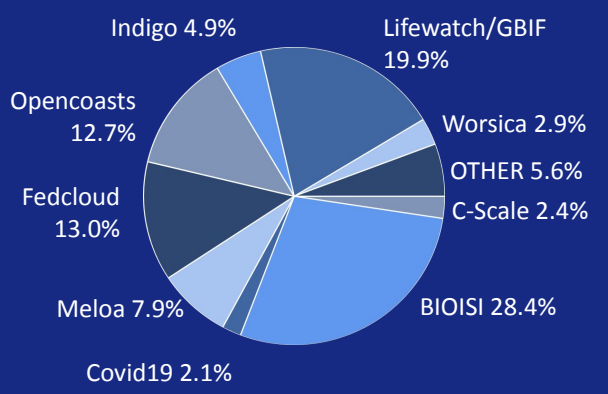
Total usage in 2021 over 13.500.265 hours

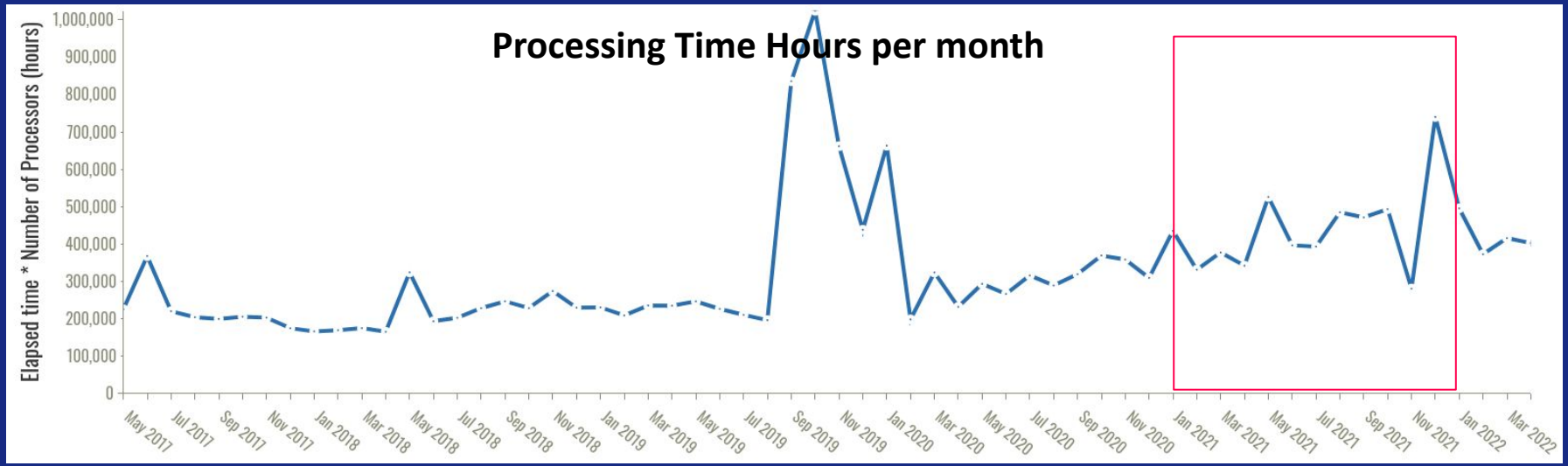
Cloud usage



Total usage in 2021 over 8.523.000 hours

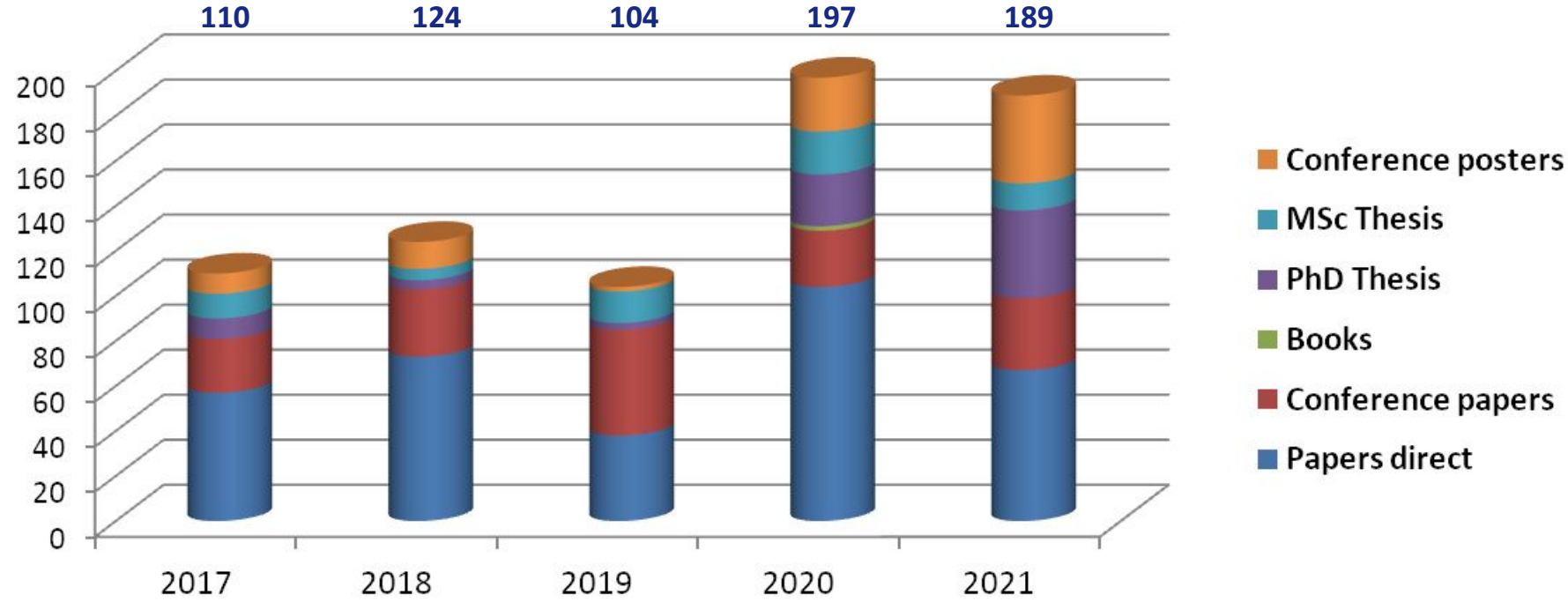
Federated cloud





2021 total usage 5,273,643 hours

Publications



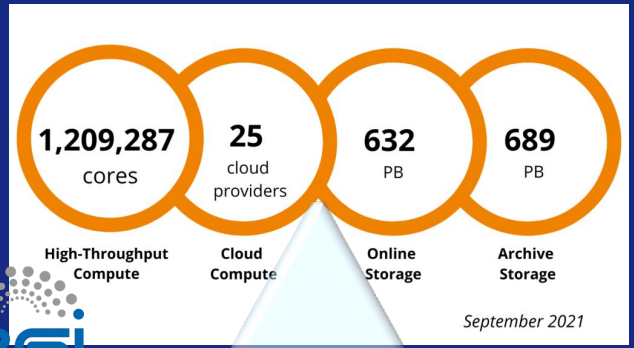
Other metrics

- Curated datasets: 2
- Patents: 2

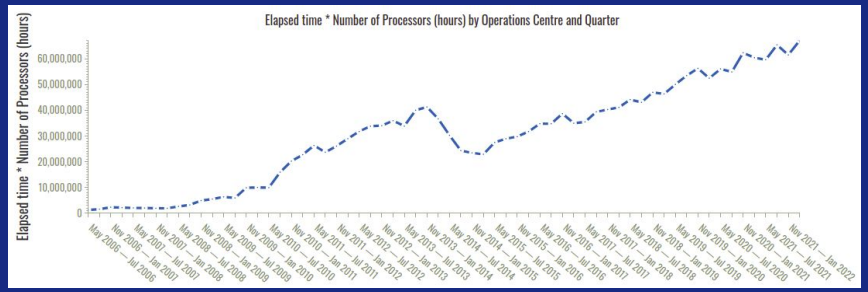
- FCT and international funded projects supported: 68
- CPCA supported projects: 50
- Artificial intelligence in public administration projects: 2
- organizations, research units and infrastructures: 62

Federation in EGI and IBERGRID

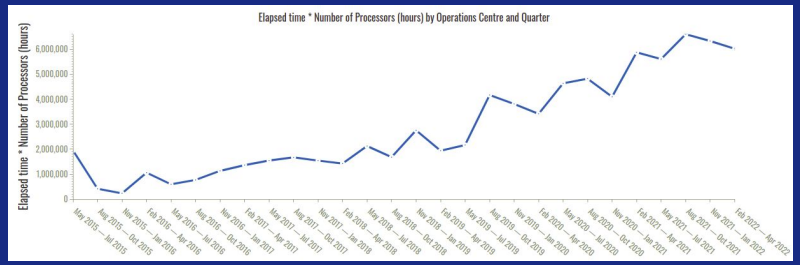
- **EGI** is a federation that joins national and regional infrastructures.
- **IBERGRID** is an Iberian federation.



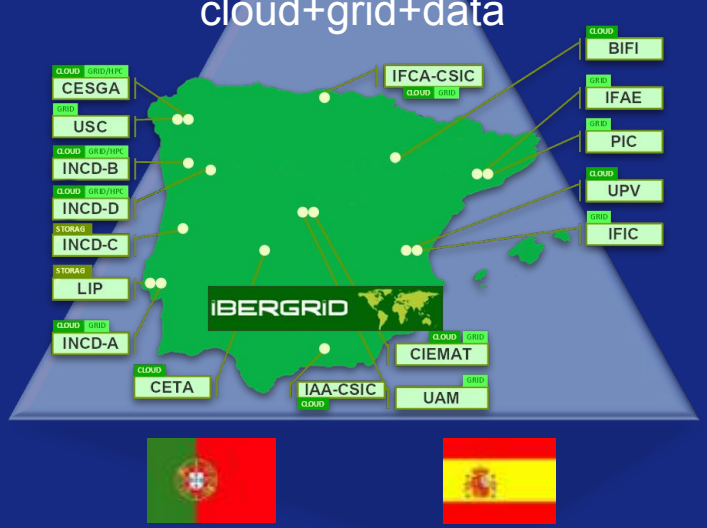
IBERGRID - 306 million jobs since 2006



IBERGRID - 1 million instantiated VMs since 2015



Operations Coordination



Advanced Computing Network - RNCA

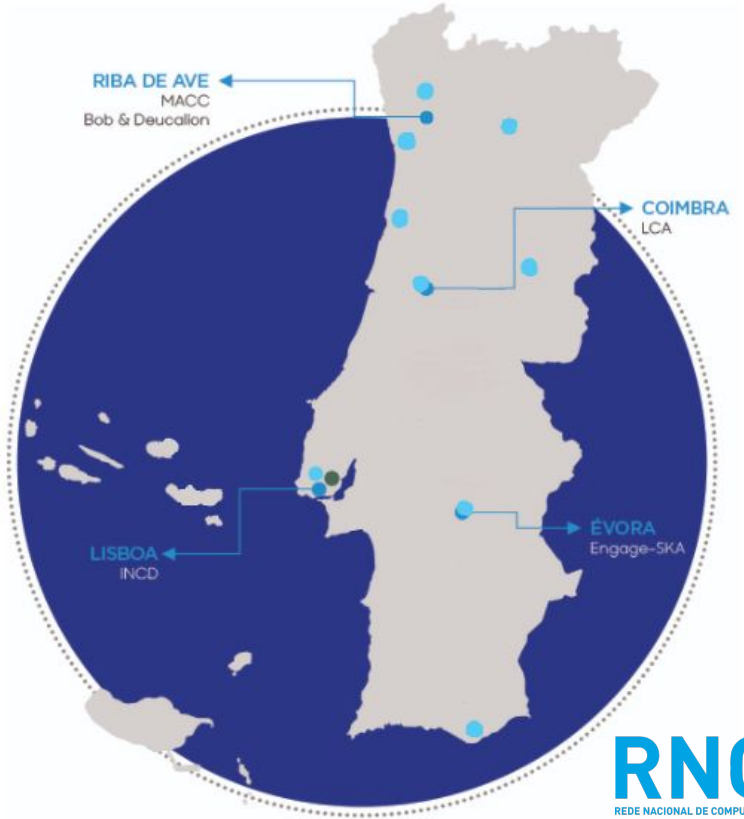
Infraestrutura Nacional de Computação Distribuída

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

Projects

Advanced computing **HEP and Big Data**

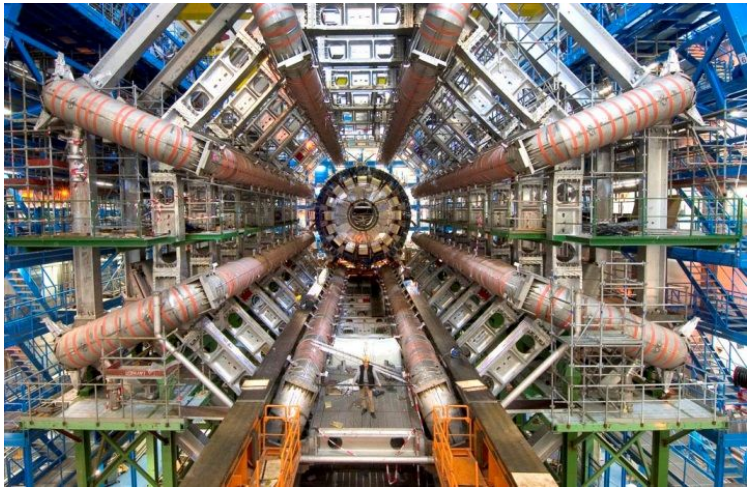
Software for High Energy Physics
- Development and optimization of code

ATLAS High-Level Trigger
- Algorithms parallelization for GPUs

Training in Scientific Computing
- Technical training
- Students MsC thesis

Understanding big data in physics
- POCI-01-0145-FEDER-029147

Now integrated in the distributed computing and digital infrastructures group



EuroCC
HPC Competence Center

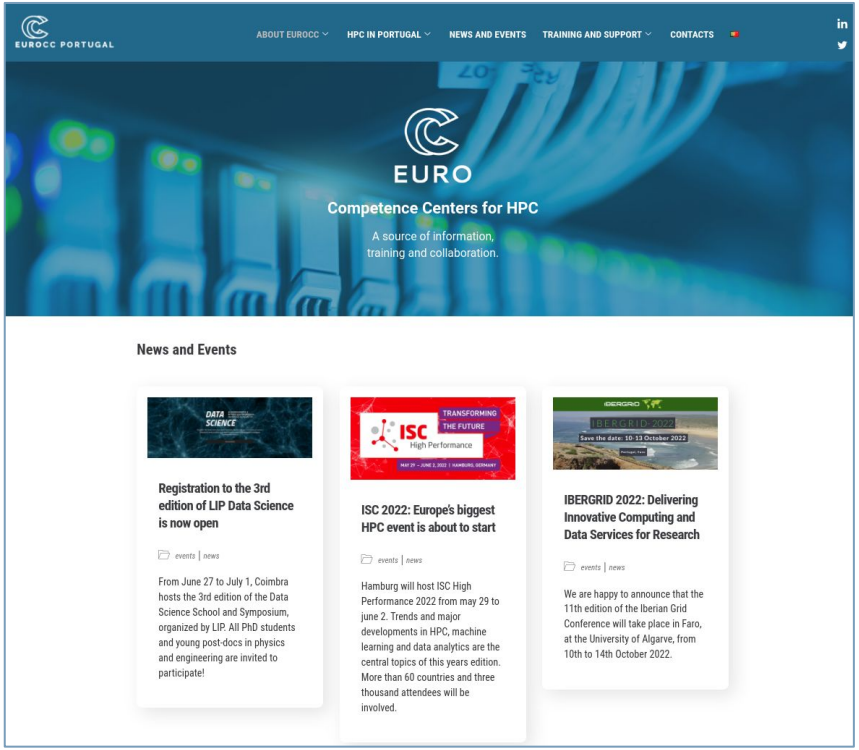
EuroHPC support project (2020-2022-2025)

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

LIP participation

- Technology transfer
- Training and skills development
- **Awareness and collaboration** (Task lead by LIP)
- Collaboration with industry
- Access to expertise and knowledge



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

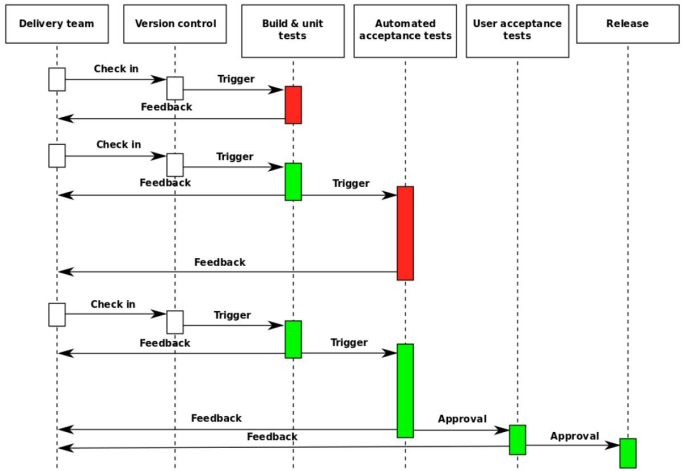
BigHPC
A framework for HPC

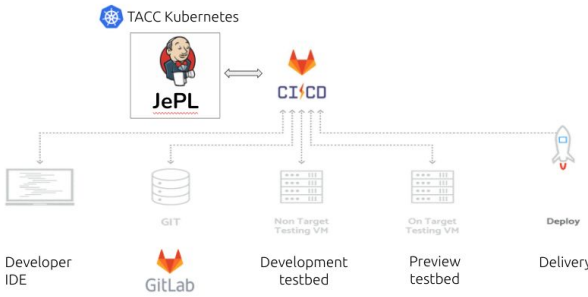
Portugal-UT-Austin project.

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.
LIP coordinates the quality assurance.



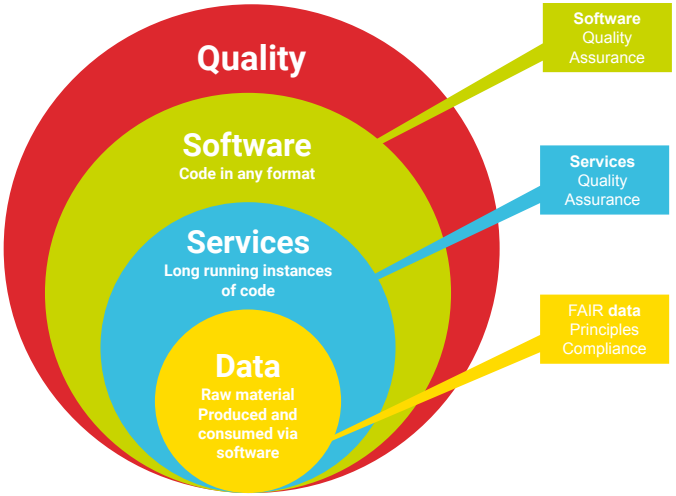


EOSC-Synergy
Quality for EOSC

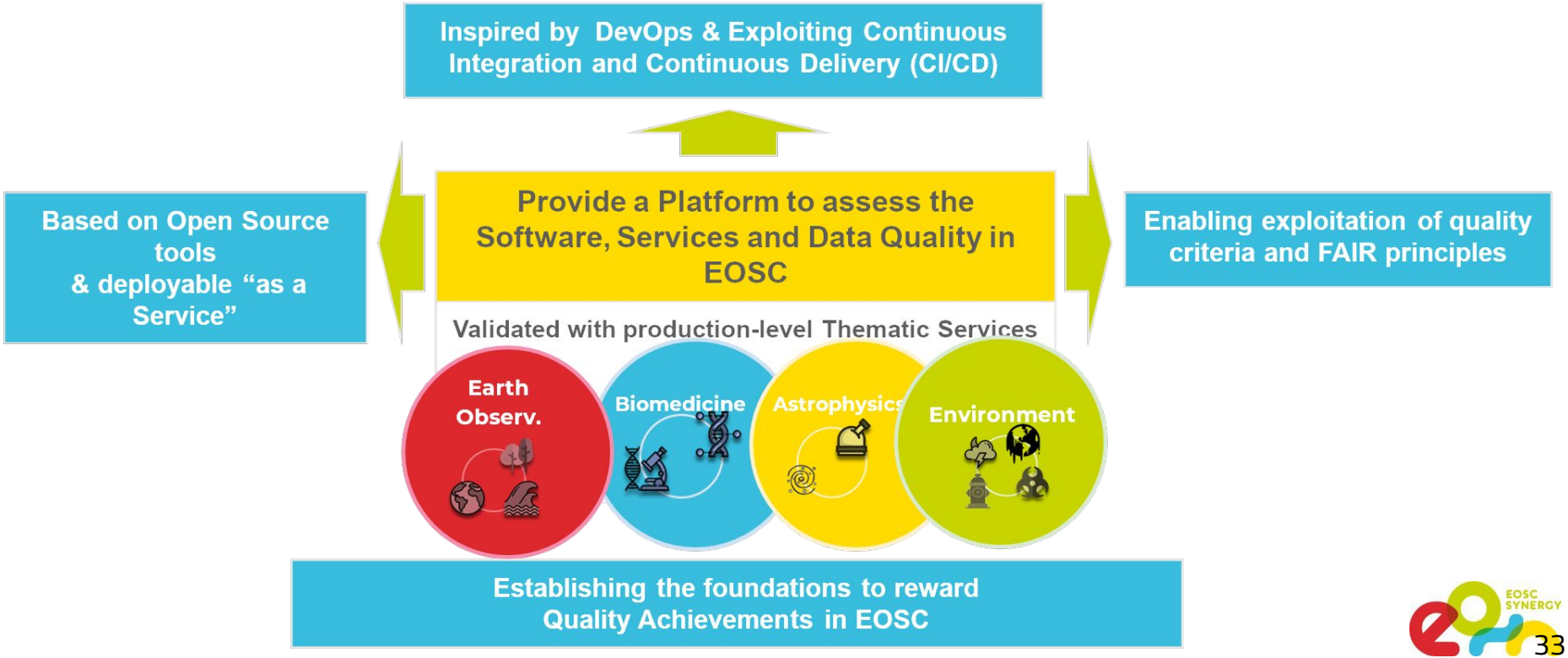
European Open Science Cloud (EOSC) project.

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.



EOSC-Synergy - quality work in a nutshell

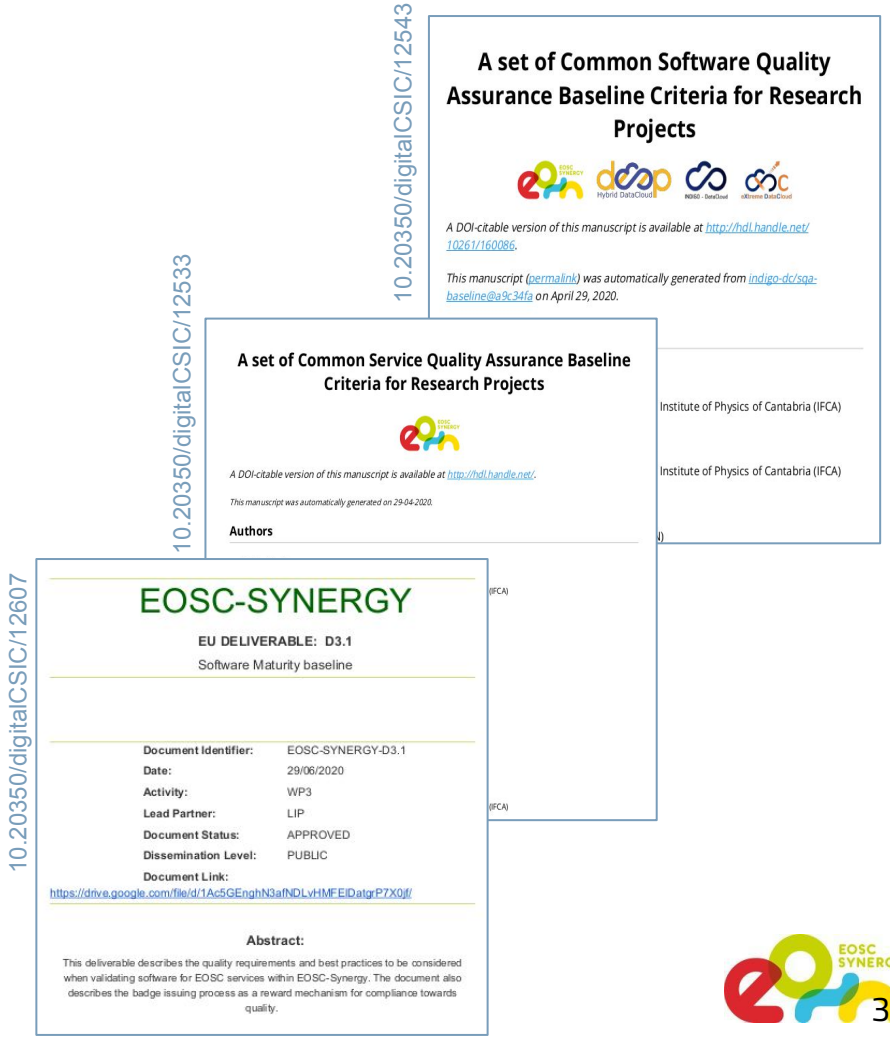


EOSC-Synergy
Quality Assurance baselines

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.

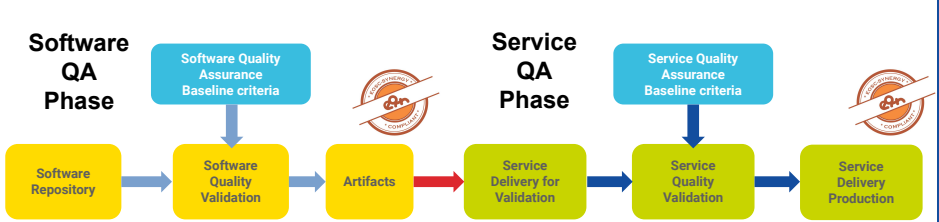
The baseline documents are openly developed on github and open to contributions.

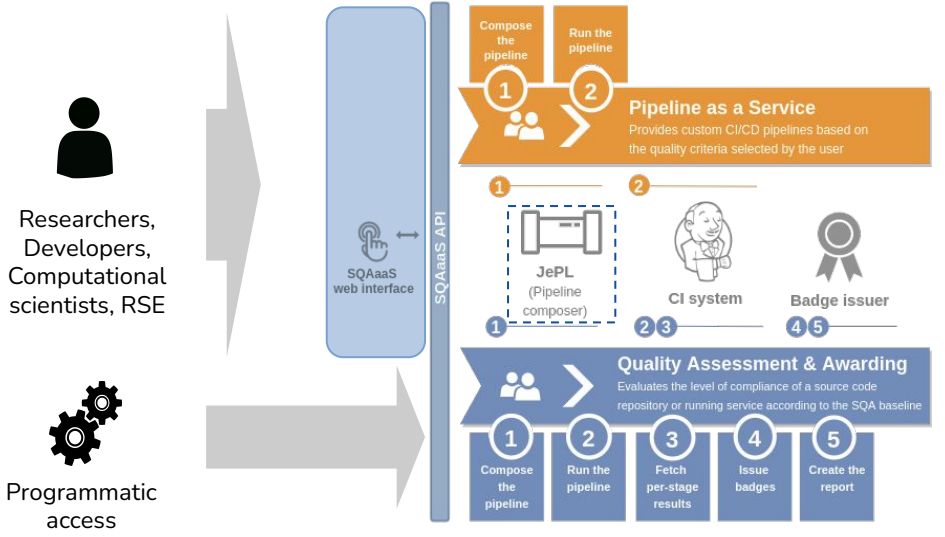


EOSC-Synergy
SQAaaS

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.

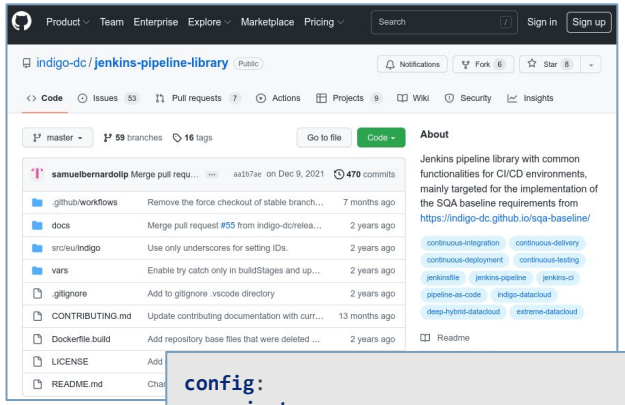




EOSC-Synergy
JePL

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

```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
```

docker-compose.yml

```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

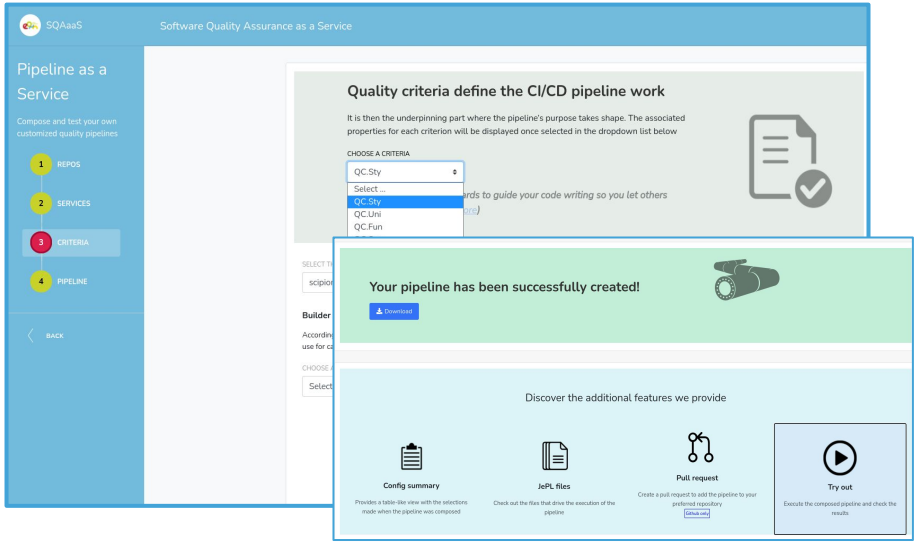
EOSC-Synergy
SQA as a Service

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.



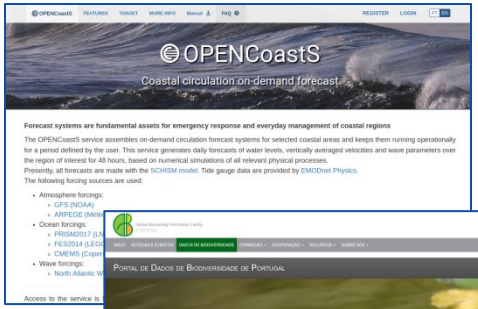
https://www.eosc-synergy.eu/technical-areas/quality/

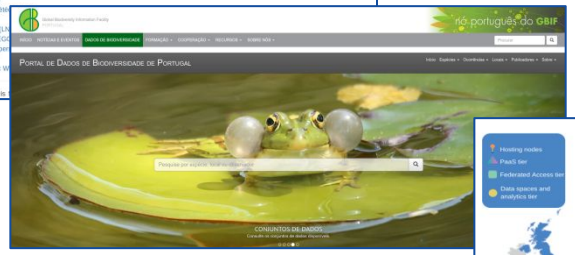
EGI-ACE
Computing for EOSC

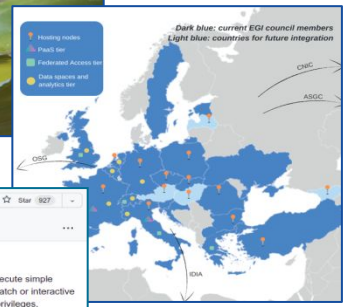
European Open Science Cloud project.

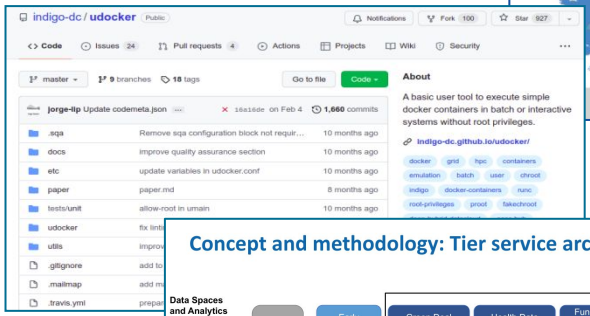
EGI-ACE (2021-2023) Advanced Computing for EOSC, with LIP work focused on

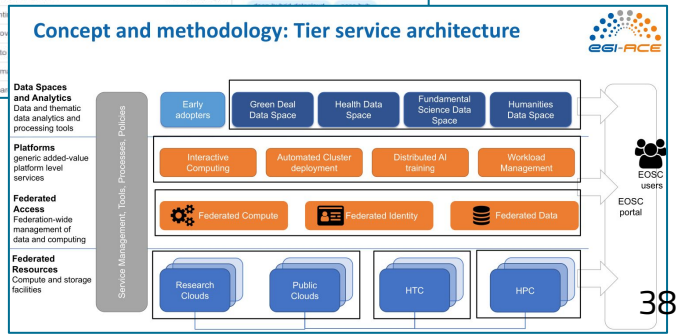
- HPC integration, software management, thematic services.
- Provisioning and support of cloud services.
- Software management for EGI.
- Implement EGI software repositories.











EGI-ACE
Coastal forecasts

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

- Water levels and velocities, wave characteristics
- Forecasts using the SCHISM model

Atmosphere forcings:

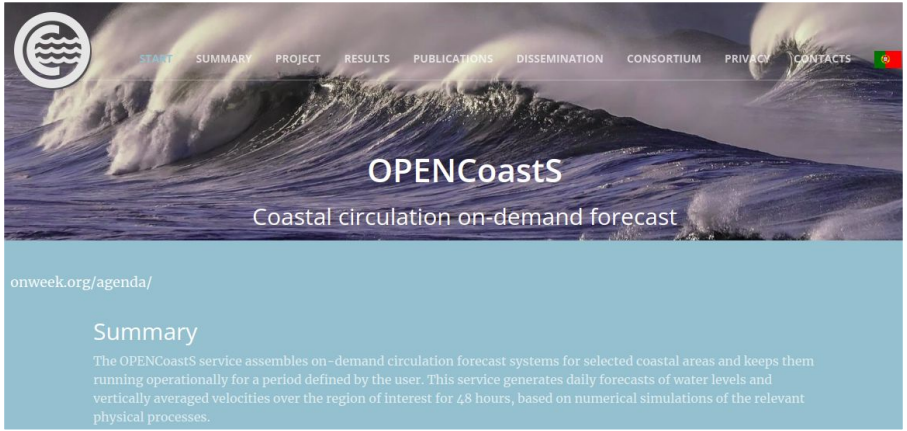
- GFS (NOAA)
- ARPEGE (MétéoFrance)

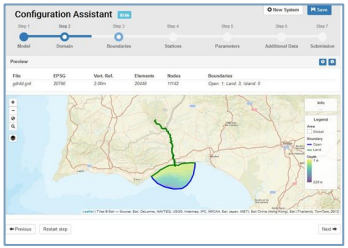
Ocean forcings:

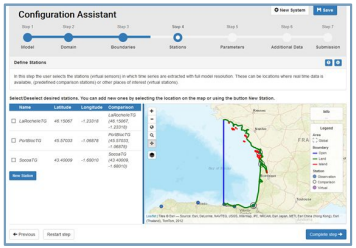
- PRISM2017 (LNEC)
- FES2014 (LEGOS)
- CMEMS (Copernicus)

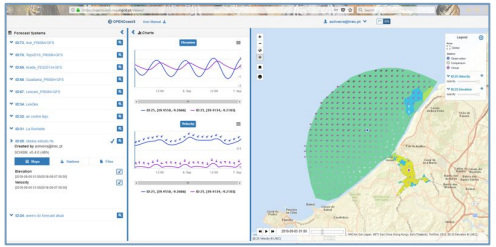
Wave forcings:

- North Atlantic WW3









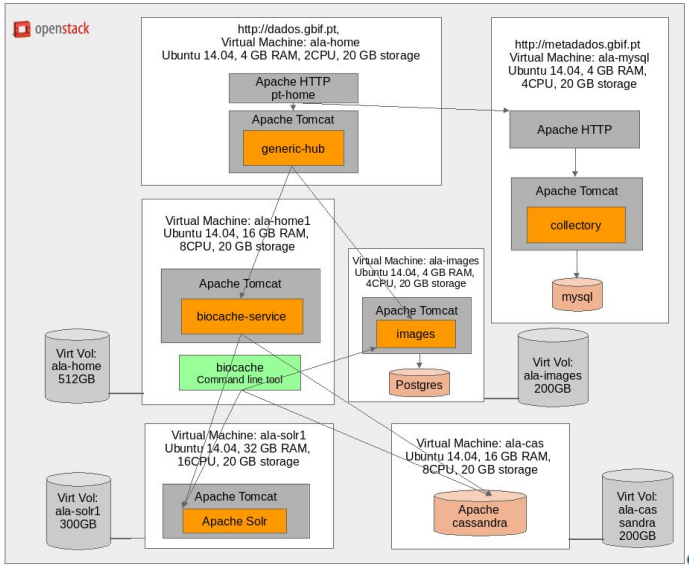
https://opencoasts.ncg.ingrid.pt

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.



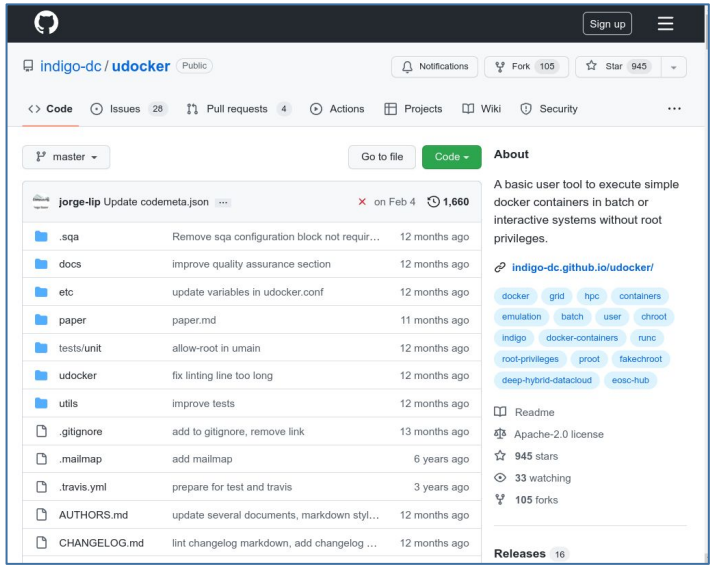
EGI-ACE
udocker

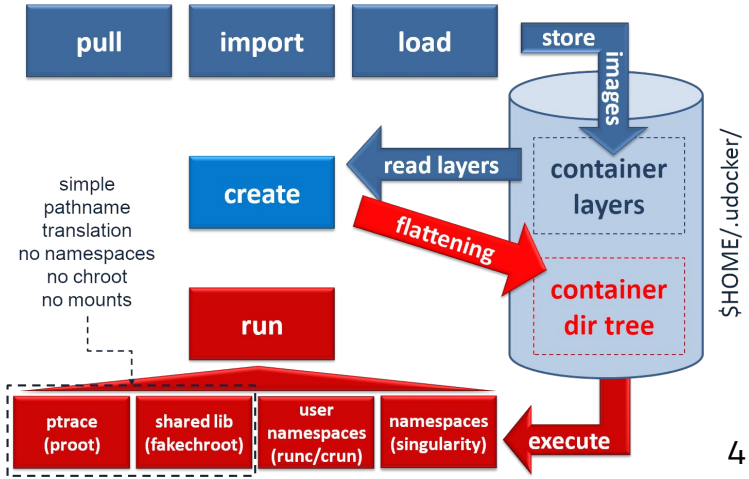
UDOCKER

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

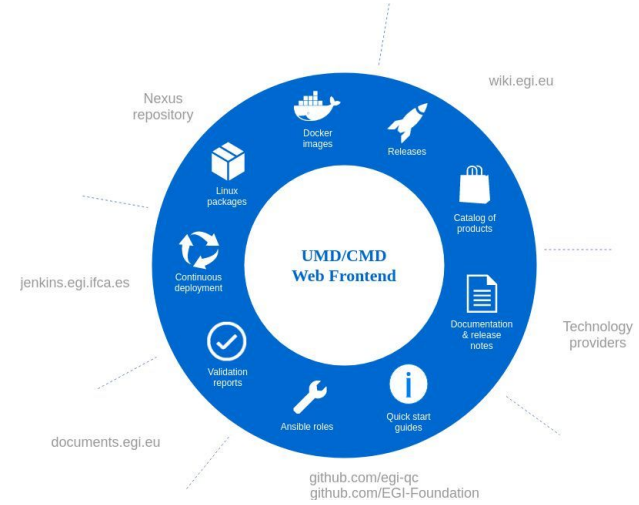


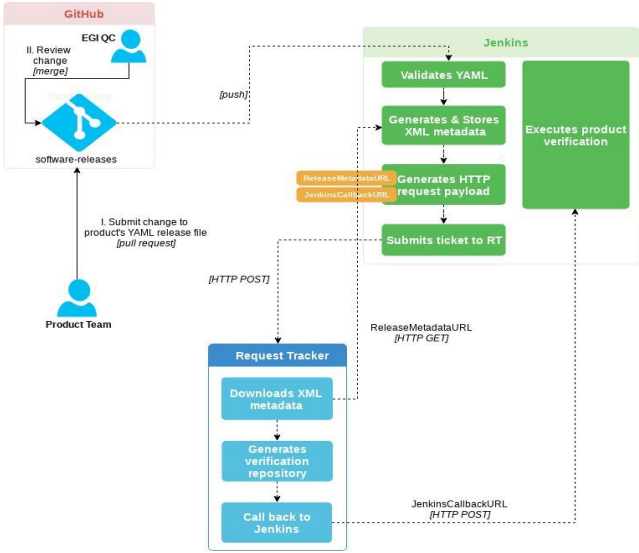


EGI-ACE
Software quality

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.





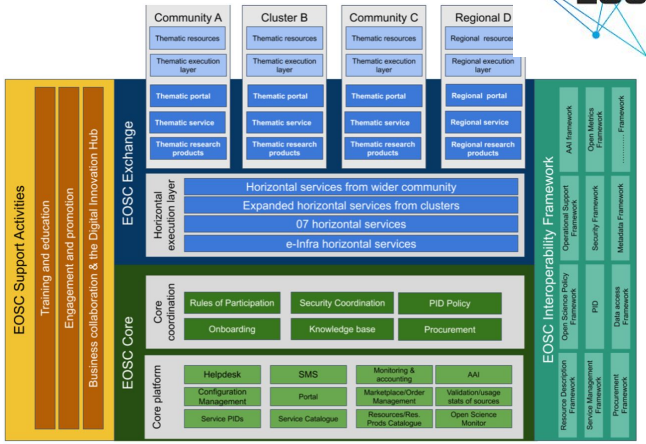
EOSC-Future
IT services for EOSC

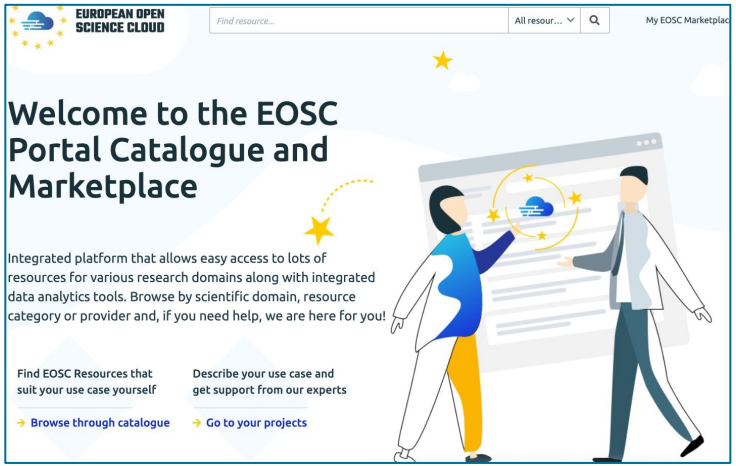
European Open Science Cloud project.

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

LIP participation in:

- IT service management for EOSC, managing the complex EOSC services ecosystem.
- Training





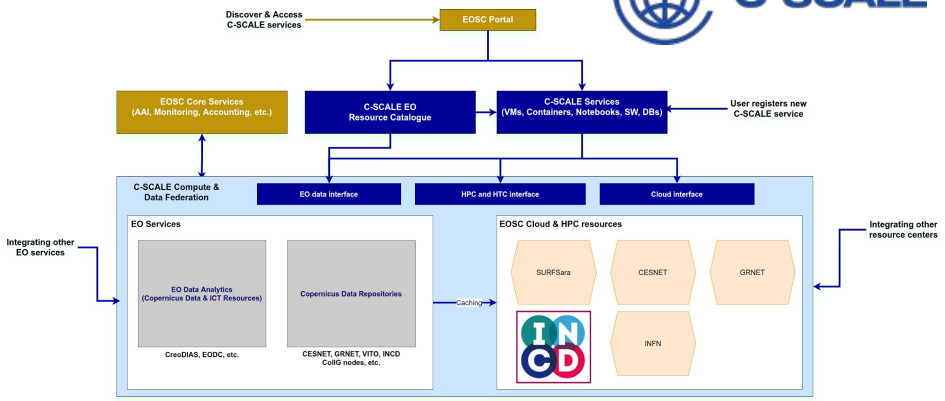
C-Scale
Copernicus

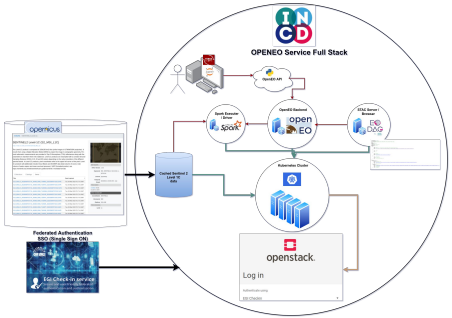
European Open Science Cloud project.

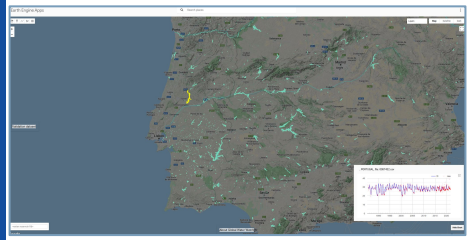
C-Scale (2021-2023) aims to federate European EO infrastructure services:

- 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).

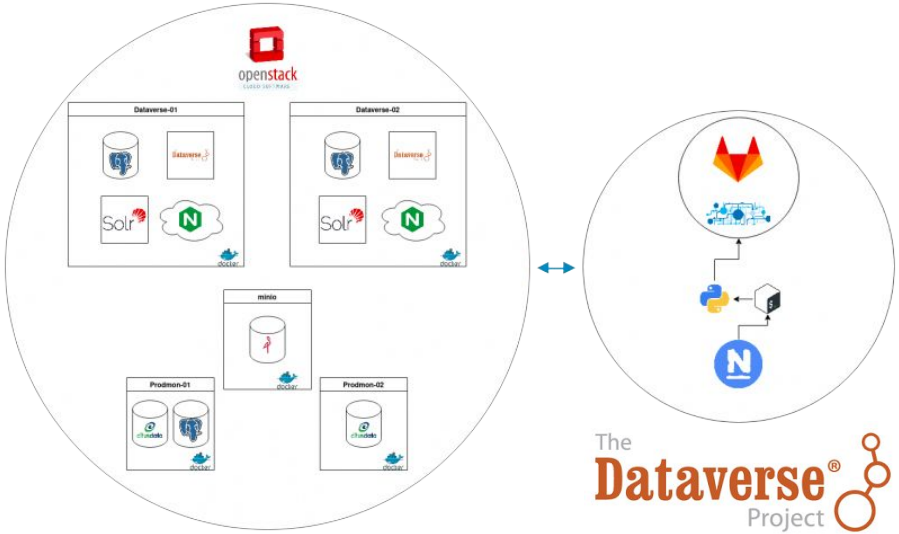
Relies on the top-of-atmosphere reflectance images from Landsat 4,5, 7, and 8 and will be extended to use Sentinel-2 MSI Level-1C data.

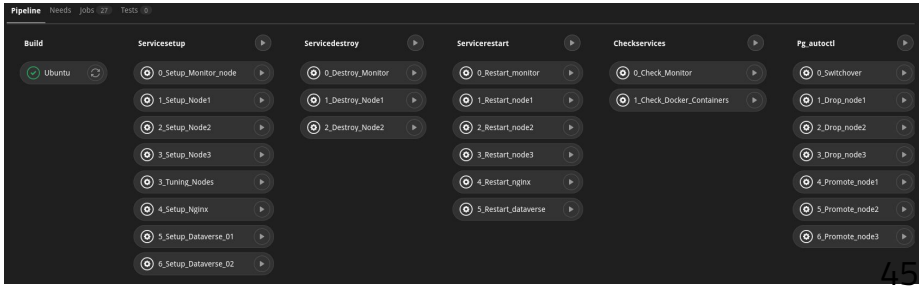
Research data **Repositories**

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.

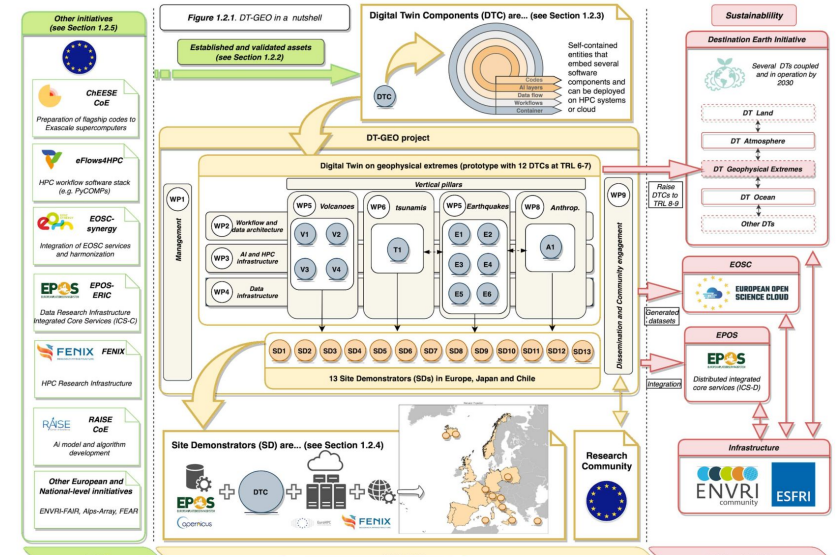


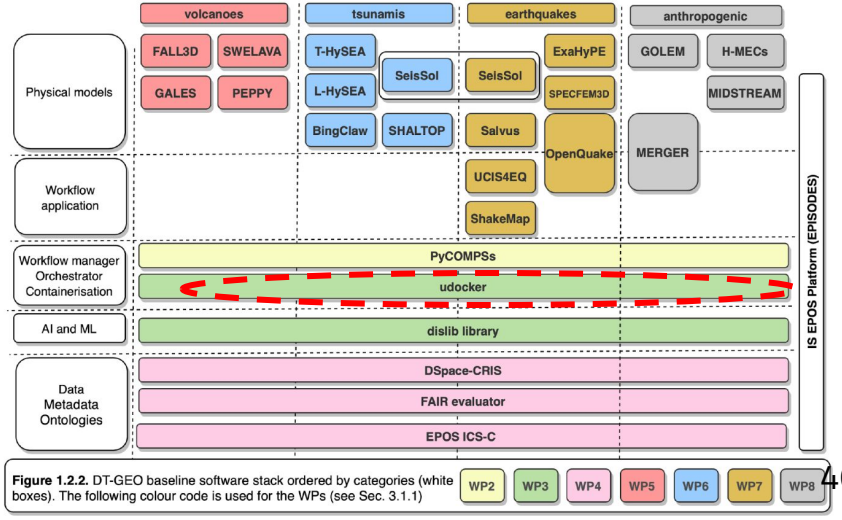


New projects (2022-2025) **Looking ahead**

New projects starting in September focused on:

- **Digital Twins**: **DT-GEO** (INFRA-2021-TECH-01)
  - Digital Twin of geophysical extremes dealing with geohazards from earthquakes, volcanoes, and tsunamis.
  - Aiming future integration with Destination Earth Initiative.
  - Urgent computing, early warning forecast and rapid post-event assessment.
- LIP contribution:
  - Software and Service Quality assessment
  - udocker integration with workflow managers and in HPC

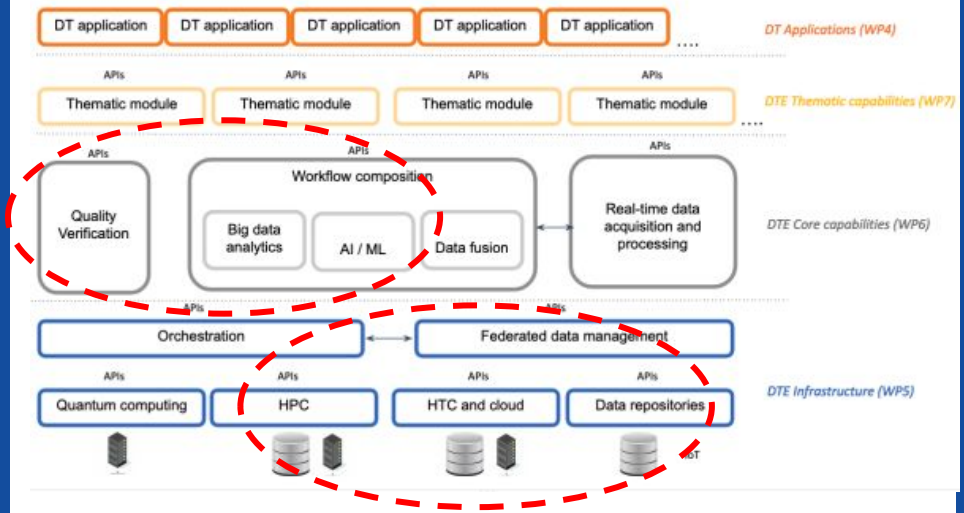


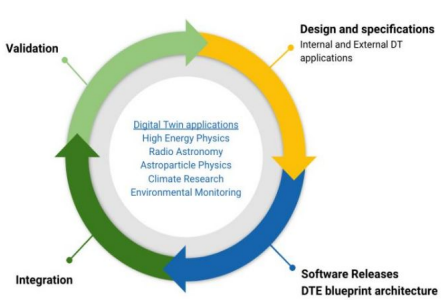


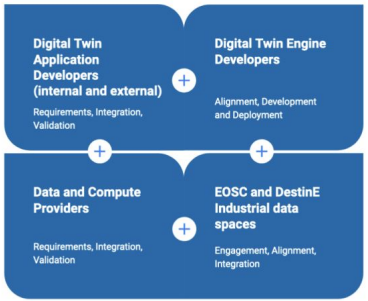
New projects (2022-2025) **Looking ahead**

New projects starting in September focused on:

- **Digital Twins**: **InterTwin (INFRA-2021-TECH-01)**
  - Develop a common approach to the implementation of DTs applicable across the whole spectrum of scientific disciplines
  - DT of LHC detector components
  - DT for noise simulation next-generation radio-telescopes
  - DT Virgo interferometer
  - DT Standard Model
  - DT of the Earth - Liaison with Destination Earth
- LIP contribution:
  - Software release and management
  - Quality and validation for applications and services







New projects (2022-2025) **Looking ahead**

New projects starting in September focused on

- **Artificial Intelligence**: **AI4EOSC** (INFRA-2021-EOSC-01)
  - Advanced services for Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) models and applications in the EOSC.
  - Event-driven data processing based on serverless computing.
  - Agrometeorological forecasts
  - Integrated plant protection scenario
  - Automated Thermography
- LIP contribution:
  - Software, services and applications quality, data FAIRness
  - Integration of udocker for serverless computing

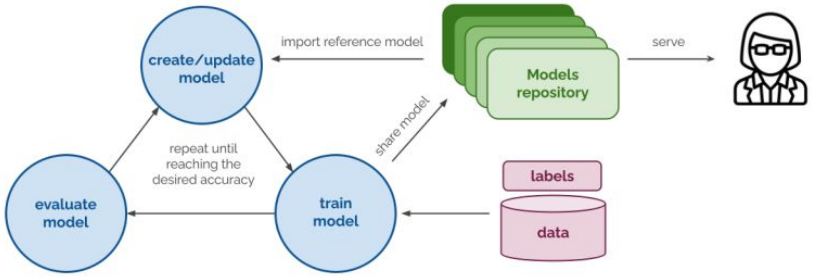
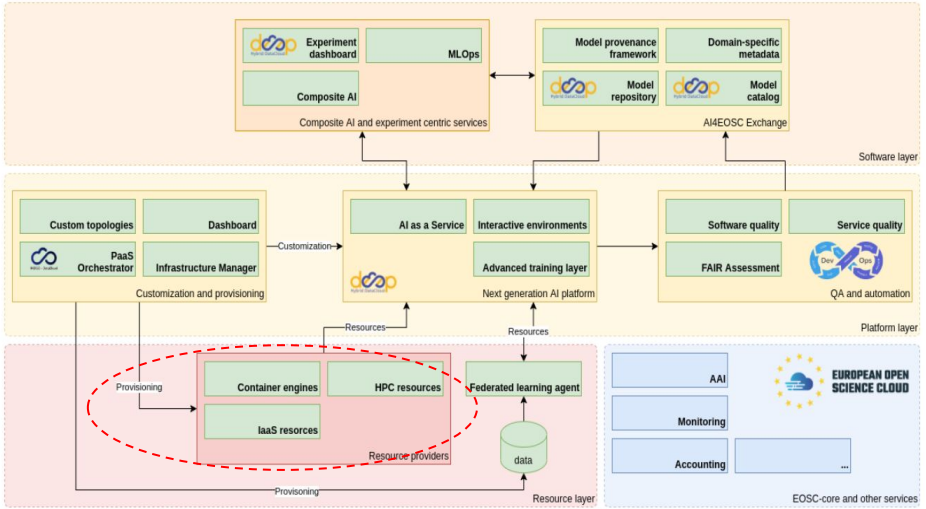


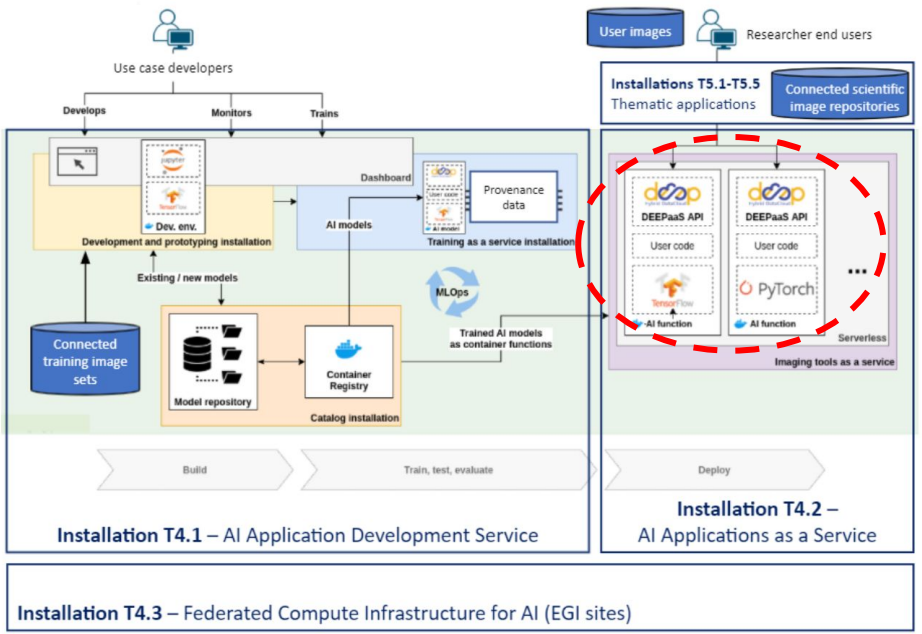
Figure 1: Machine learning application development lifecycle.



New projects (2022-2025) **Looking ahead**

New projects starting in September focused on

- **Artificial Intelligence**: **iMagine** (INFRA-2021-SERV-01)
  - 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 monitorization and analysis
- LIP contribution:
  - DEEP AI Application Development Service
- INCD contribution:
  - Federated Compute Infrastructure for AI and support



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 !

LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Thanks!

Discovery through science

Innovation through technology

Sharing with People