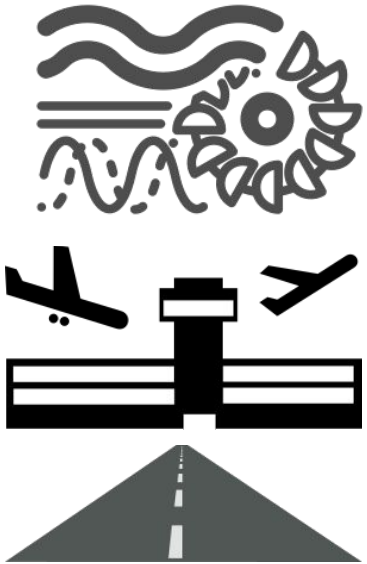
e-Infrastructures and how can we use them

www.incd.pt

The INCD project
22153-01/SAICT/2016
is co-funded by:

Infrastructures

airports, dams, roads, railroads ...



vs

Facilities and systems that provide important services serving countries or regions

e-Infrastructures

digital facilities and services, computing, storage, networks ...

e-Infrastructures - layers

Researchers, Engineers, etc from multiple organizations

e-Infrastructure encompass IT services

state-of-the-art data acquisition, data storage, data management, data integration, data mining, data visualization and other

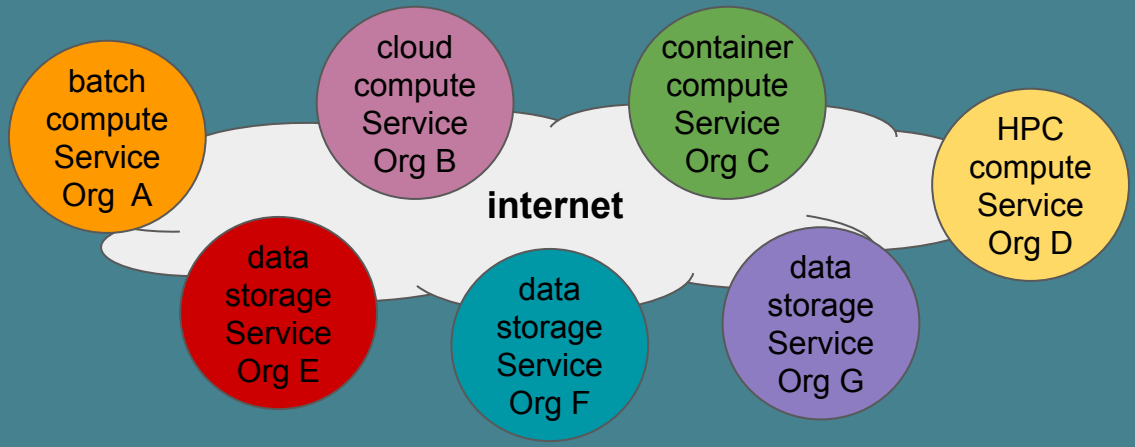
that exploit

Advanced Computers and Networks across locations

e-Infrastructures - distributed resources

Researchers, Engineers, etc from multiple organizations

e-Infrastructures integrate compute and data resources



e-Infrastructures - some examples

European scope

GÉANT

EUDAT

PRACE

Country scope

INCD Infraestrutura Nacional de Computação Distribuida

ES-NGI

SURF

SNIC Swedish National Infrastructure for Computing

srce

FRANCE GRILLES

Jisc

CSC

PSNC

MetaCentrum

SwiNG Supporting Swiss Scientific Computing

TÜBİTAK ULAKBİM

GA Gauß-Allianz

Domain specific

WLCG Worldwide LHC Computing Grid

elixir

emso ERIC

LifeWatch ERIC

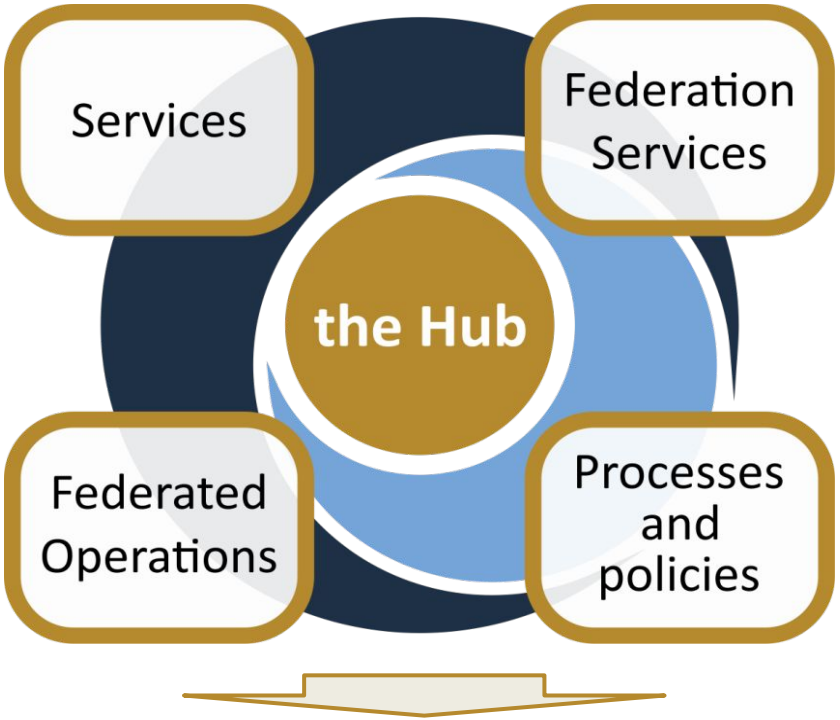
EPOS EUROPEANPLATEOBSERVINGSYSTEM

DARIAH-EU Digital Research Infrastructure for the Arts and Humanities

BBMRI-ERIC Biobanking and BioMolecular resources Research Infrastructure

Implementation of the European Open Science Cloud (EOSC) Hub: a federated integration and management system for EOSC

- Data repositories
- Thematic services
- Applications & tools
- Generic services (storage, compute, connectivity)…
- Training, consultants
- …



- Marketplace
- Authentication
- Authorization
- Accounting
- Monitoring
- …

- Certification of providers
- SLA negotiation
- Users management
- …

- Security regulations,
- Compliance to standards,
- Terms of use,
- Data best practices
- …

EUROPEAN OPEN SCIENCE CLOUD

Services by area

e-Infra

- EGI Federation
- EUDAT CDI
- INDIGO-DataCloud

Humanities

- Language and literature (CLARIN)
- Arts (DARIAH)

Engineering

- Environmental engineering (sea vessels, LNEC)
- Civil Engineering (Disaster Mitigation)

Medical and Health Sciences

- Biological Sciences (ELIXIR)
- Structural biology (WeNMR)

Natural sciences

Physical Sciences

- Astronomy (LOFAR)
- Fusion (ITER)
- High Energy Physics (CMS and VIRGO)
- Space Science (EISCAT-3D)

Earth Science

- EO Pillar
- GEO
- Climate Research (ENES)
- Seismology (ORFEUS, EPOS)

Biological Sciences

- Marine and freshwater biology (IFREMER)
- Biodiversity conservation (LifeWatch)
- Ecology (ICOS)

EOSC-hub Marketplace

Platform where services can be

- Promoted
- Discovered
- Ordered
- Accessed

3 steps process for users

- Authentication & registration
- Product discovery and specification
- Ordering and request processing

About EGI

- Federated Infrastructure
  - Resource Providers: ~ **350 centres** federated into NGIs
  - Resources: **850,000 CPU cores**, **300PB disk**, **350PB archive storage**
  - Resource centre autonomy in resource allocation to VOs
- Distributed research communities
  - Users: ~ **31,000** grouped into **238 Virtual Organisations** (VOs)

With EGI, researchers can easily share ICT resources, data, applications, VREs, and training materials

EGI target users

- Multinational projects joining multiple organizations
  - EGI offers services to federate and share the data and compute resources of the participating research organizations
  - EGI offers services to coordinate the operation and usage of the federated resources
  - EGI offers training on IT service management and other topics

- Research groups
  - EGI offers seed computing and data storage capacity
  - EGI offers support and training

EGI services - batch computing

High-Throughput Compute
Batch processing to execute thousands of computational tasks to analyse large datasets

- Batch processing
  - Choose site where to run
  - Job is sent to batch cluster
  - Job waits in batch queue
  - Execution on Linux system

Workload Manager
Distribute processing tasks automatically across the computing resources in the EGI federation

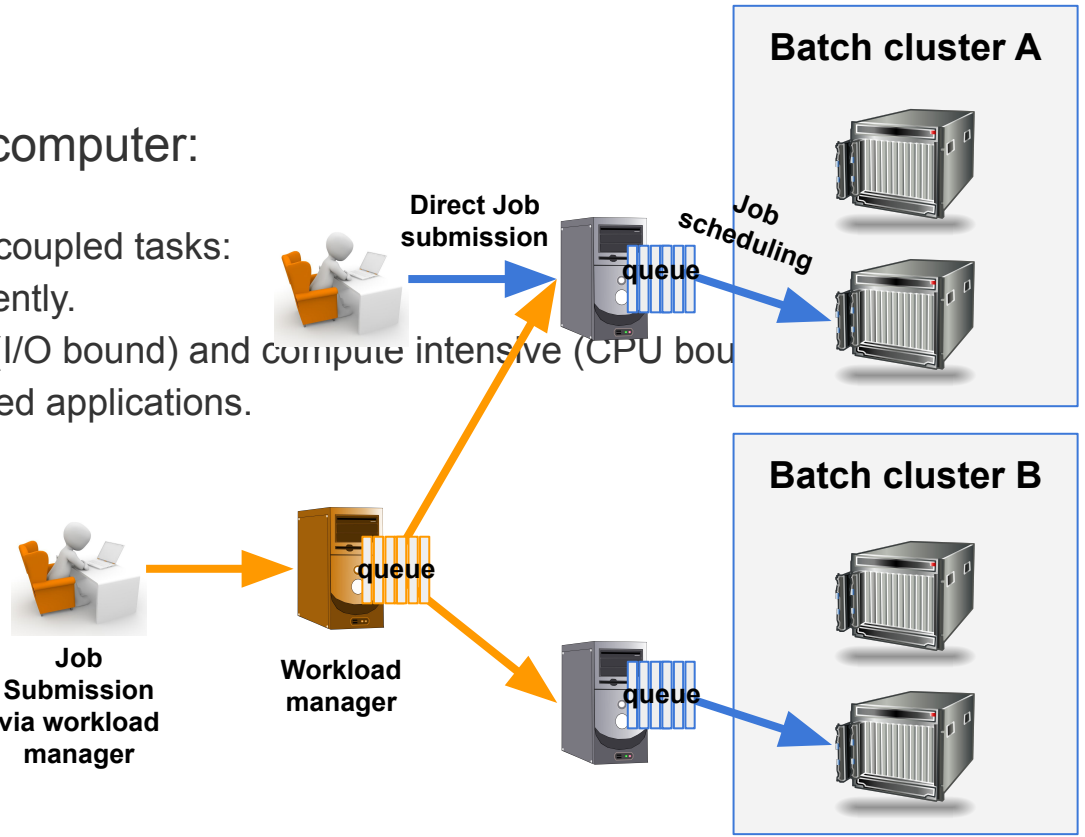
- Batch using multiple sites
  - Jobs are sent to the workload manager
  - Jobs are distributed across available sites
  - Jobs are processed as above

Why use High Throughput Compute service

Each program runs within a single computer:

- Execution of a large number of loosely-coupled tasks:
  - Each task is processed independently.
  - Adequate both for data intensive (I/O bound) and compute intensive (CPU bou
  - Adequate for serial or multithreaded applications.

- Examples:
  - Particle/Astroparticle physics
  - Data analysis
  - Genomics, protein folding
  - Machine learning applications
  - And many more...



EGI services - cloud oriented

Cloud Compute
Create and execute virtual machines on demand as in commercial IaaS clouds

- Cloud
  - Using a web portal or API
  - Start VM at remote sites
  - Liberty to configure the VMs
  - Run services or applications

Cloud Container Compute
Run docker containers in a lightweight Linux virtualized environment

- Containers
  - Similar to VMs but occupy less memory and disk space
  - Mostly used for services but can be used for applications

Using Cloud or Batch

Cloud computing and Linux containers:

- **Long running services** (months-years), scientific gateways and platforms.
- Virtualization: Virtual Machines or Containers.
- Highly customizable/Flexible environment - **Users choose**:
  - Operating System, Software and libraries, amount of resources (# Nodes, #CPUs, Memory, Storage, network topology).
  - On demand instantiation.
- Resources are administered/deployed by the users.
- Resources are isolated from other users.

Batch processing and Grid computing:

- Applications with **limited execution time** (minutes-weeks).
- “Almost” static environment:
  - Operating System, software, compilers and libraries are deployed by the system administrator.
  - Users request the sysadmins to deploy specific software, libraries, compilers to their needs.
- Applications are scheduled into a queue for execution until enough resources became free.
- Resources are shared with other users.

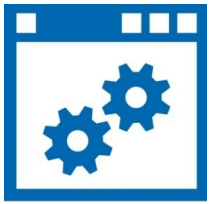
EGI services - storage oriented

Online storage
Store, share and access files and their metadata.

Archive storage
Backup data for long-term storage and future reuse.

- Storage short-term
  - Can be combine with the compute services
  - Storage space for data processing input and output
  - Backend storage for data catalogues
- Storage long-term
  - Archival for retention
  - Release online storage capacity
  - Store large amounts of data

EGI services - data analysis oriented



Applications on demand
Access to common applications and frameworks.

Notebooks
Create interactive documents with live code, visualisations and text.

- Available applications
  - Application and data analysis frameworks: Jupyter Notebook, Docker, Apache Tomcat, Hadoop, Marathon, and Chronos
  - Life sciences: Galaxy, ClustalW2, Chipster, NAMD and AutoDock Vina
  - Data analysis: GnuPlot, Octave and the Statistical R for Computing
  - Humanities: the parallel Semantic Search Engine
- Data analysis notebooks
  - Browser-based tool for interactive data analysis
  - EGI storage and compute services.
  - Based on the JupiterHub technology.

INCD - Portuguese Distributed Computing Infrastructure

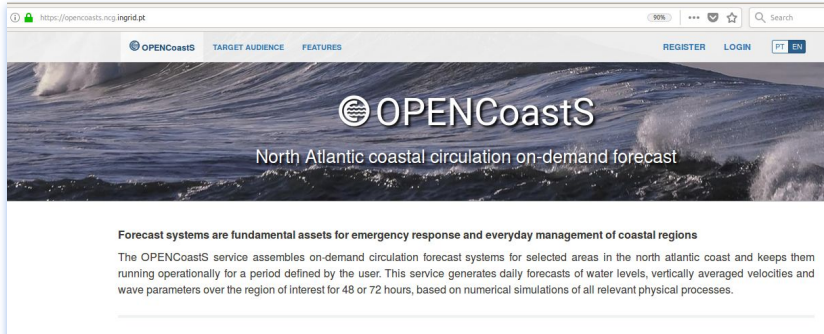
- **INCD** is a **digital infrastructure**:
  - Funded by FCT in the context of the Portuguese Roadmap of Research Infrastr
- Goals:
  - Provide **computing** and **data services** for the academic
  - Support participation of Portuguese researchers in:
    - **National R&D projects.**
    - **Large international projects.**
    - **Scientific infrastructures** (such as CERN, ESFRIs).
    - **Other** research projects & activities.
- Computing Services:
  - Cloud.
  - HTC and Grid.
  - HPC.

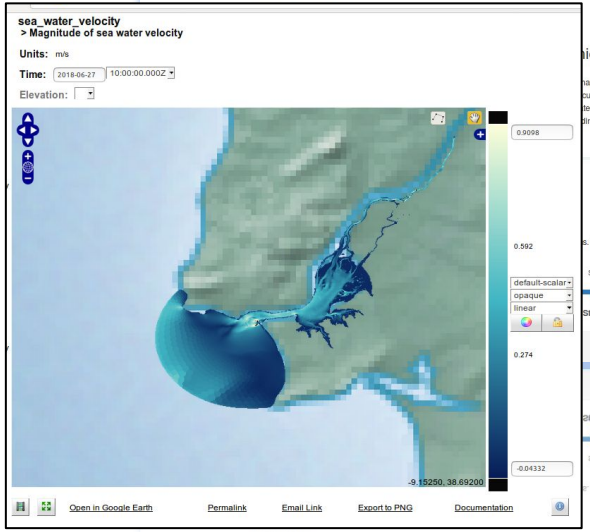
Deploying the OPENCoastS service in EOSC

- **On-demand circulation forecasts for selected coastal areas**:
  - 2D velocities and wave parameters.
  - Over the user's spatial region of interest for periods of 72 hours.
  - **Numerical simulations** of all relevant physical processes: **parallel using MPI**
  - In the European Atlantic coast.
- **Thematic Service** for the European Open Science Cloud (**EOSC**)

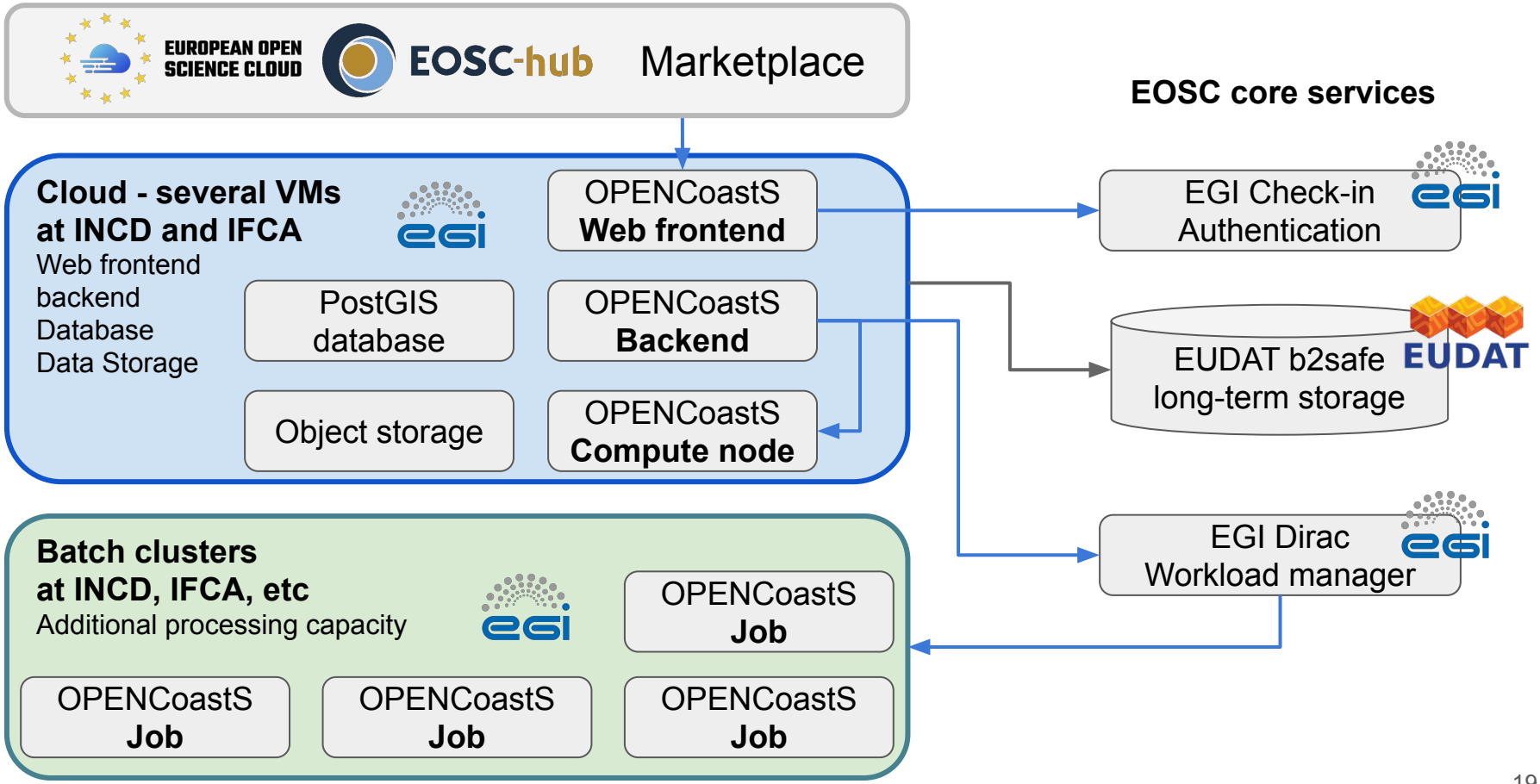
  https://opencoasts.ncg.ingrid.pt

- **e-Infrastructures: INCD (pt) + IFCA (es)**

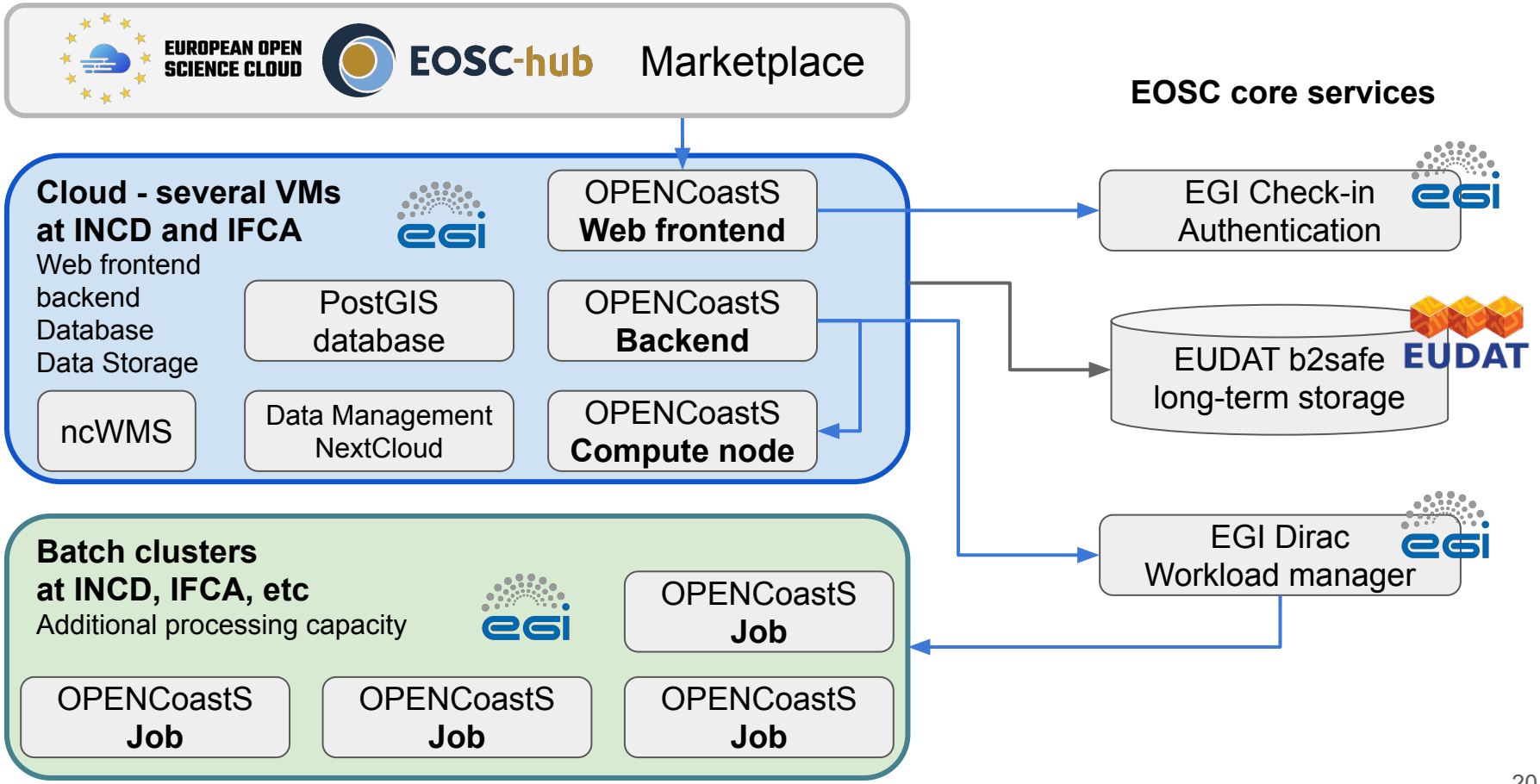




Deploying the OPENCoastS service in EOSC



Deploying the OPENCoastS service in EOSC



Benefits of using e-Infrastructures

- Access to scientific computing and data services tailored to research and engineering
- Explore capacity and capabilities available at national and international level
- Access a large portfolio of scientific computing and data services with added-value
- Combine different services to match the needs of your applications
- Single authentication and authorization mechanisms to access all services
- Better connectivity through the academic networks in Europe and elsewhere
- Better redundancy through the usage of resources at several locations
- Interconnect your compute and storage capacity with your project partners
- Store, find and share data collections and analyse data in the cloud
- Create services and portals to explore and share data

Contact and present your use case

https://www.eosc-hub.eu/

https://www.egi.eu/services/

https://eudat.eu/

For Portuguese users

Infraestrutura
Nacional de
Computação
Distribuída

https://www.incd.pt

helpdesk@incd.pt