



Using compute services in EOSC: Experiences and advices from EOSC-hub

EOSC-hub Webinar Programme

Tuesday 10 Nov 2020, 14:00 → 15:00



eosc-hub.eu



[@EOSC_eu](https://twitter.com/EOSC_eu)



- This webinar is recorded and will be published online
- Slides and recordings will be published in the EOSC-hub Webinars Page: <https://www.eosc-hub.eu/webinar-programme>
- Please keep your questions until the end
- Questions will be conducted by “raising your hand” through the Zoom functionality



Introduction to EOSC-hub

Francesca Spagnoli

EOSC-hub Innovation Management, Communication and Stakeholder Engagement
Project Manager, Trust-IT Services



eosc-hub.eu



@EOSC_eu



“Delivering the integration and management system of EOSC, acting as a European-level entry point for all stakeholders”

- *20 digital research infrastructures with EGI, EUDAT and INDIGO-DataCloud jointly offering services, software and data for advanced, data-driven research & innovation*

EOSC-hub: Integrating and managing services for the European Open Science Cloud

Grant Agreement ID 777536

Total budget: €33,287,542

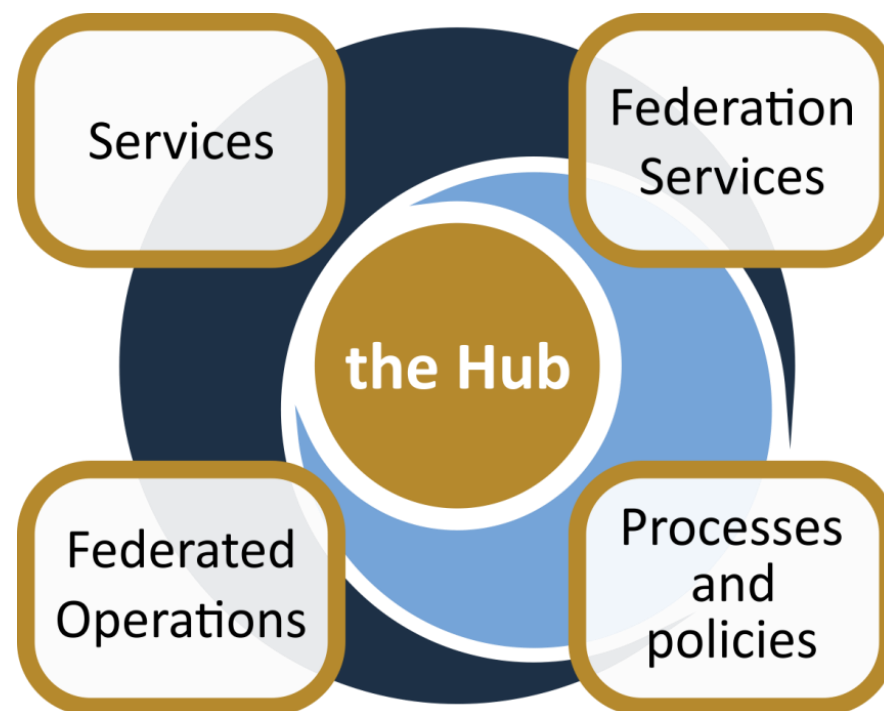
**100 Partners, 53 countries
+150 staff involved**

**Coordinator
STICHTING EGI**

Jan 2018 – Dec 2020 → March 2021

- Baseline services (storage, compute,...)
- Applications & tools
- Data
- Training & consultancy

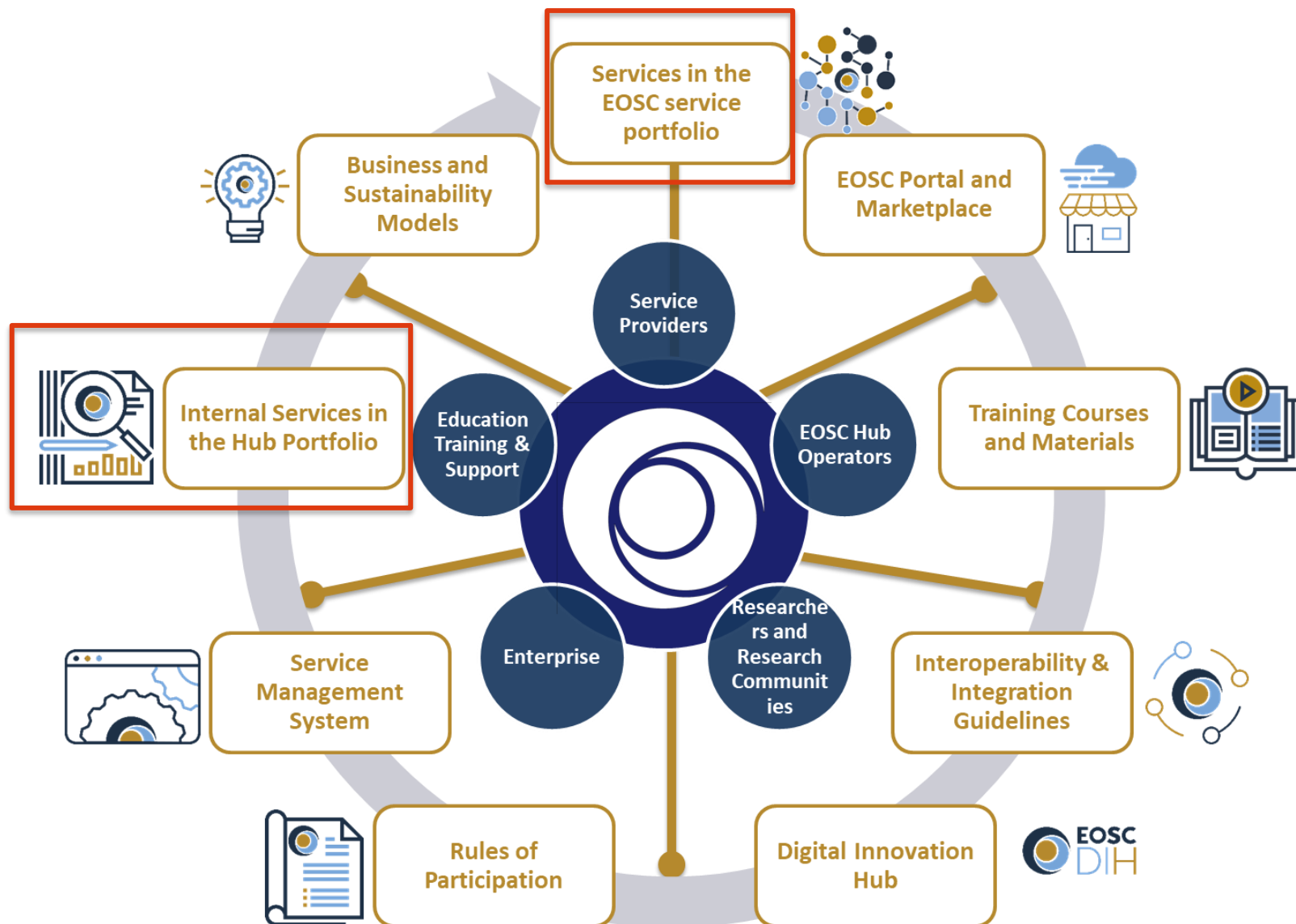
- Lightweight certification of providers
- SLA negotiation
- Customer Relationship Management



- EOSC Portal & Marketplace
- AAI
- Accounting
- Monitoring
- Helpdesk

- Security regulations
- Compliance to standards
- Terms of use
- FAIR implementation guidelines

EOSC-hub Key Exploitable Results to be tackled today





Using compute services in EOSC: Experiences and advices from EOSC-hub

Enol Fernández – enol.Fernandez@egi.eu



eosc-hub.eu

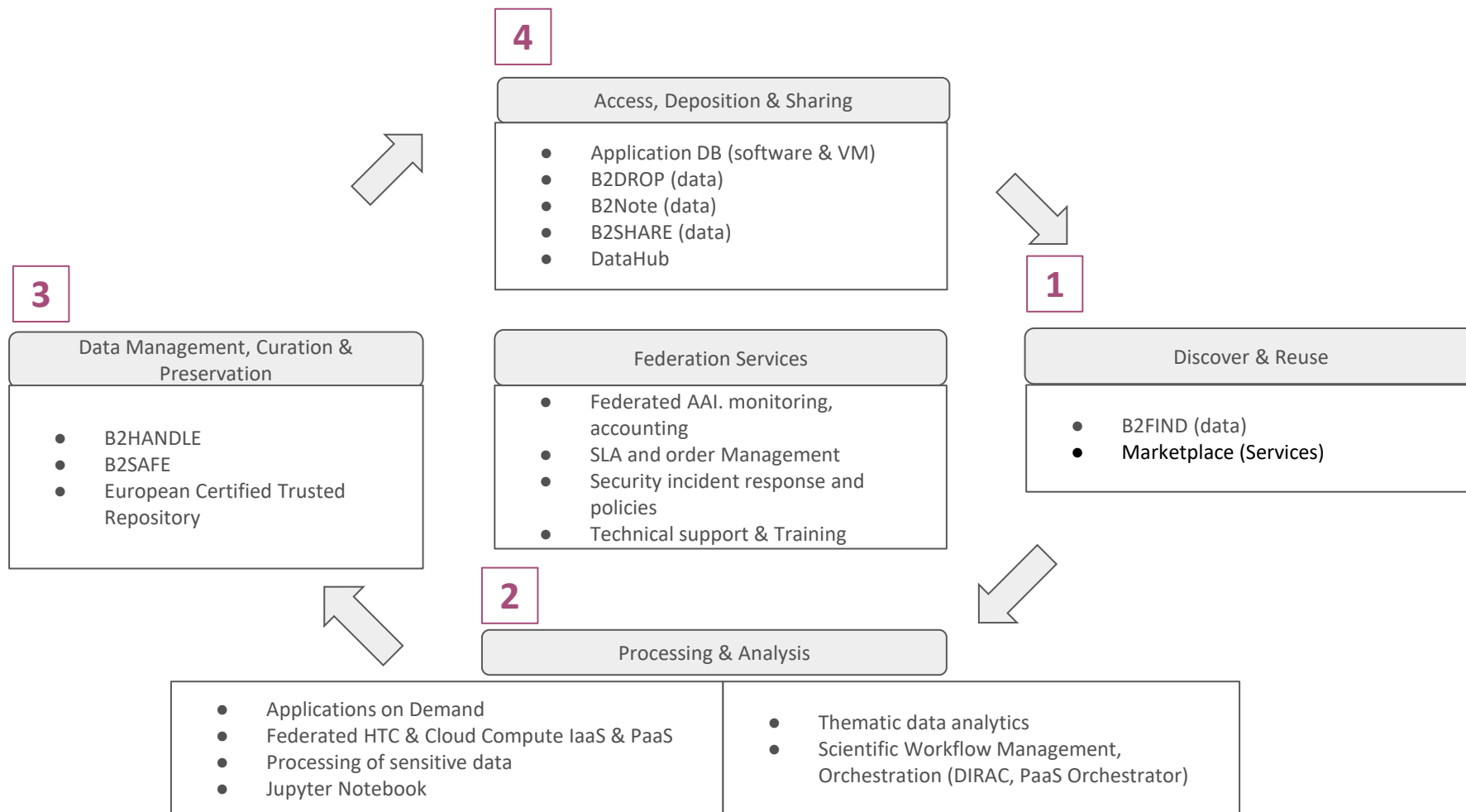
Dissemination level: Public



[@EOSC_eu](https://twitter.com/EOSC_eu)



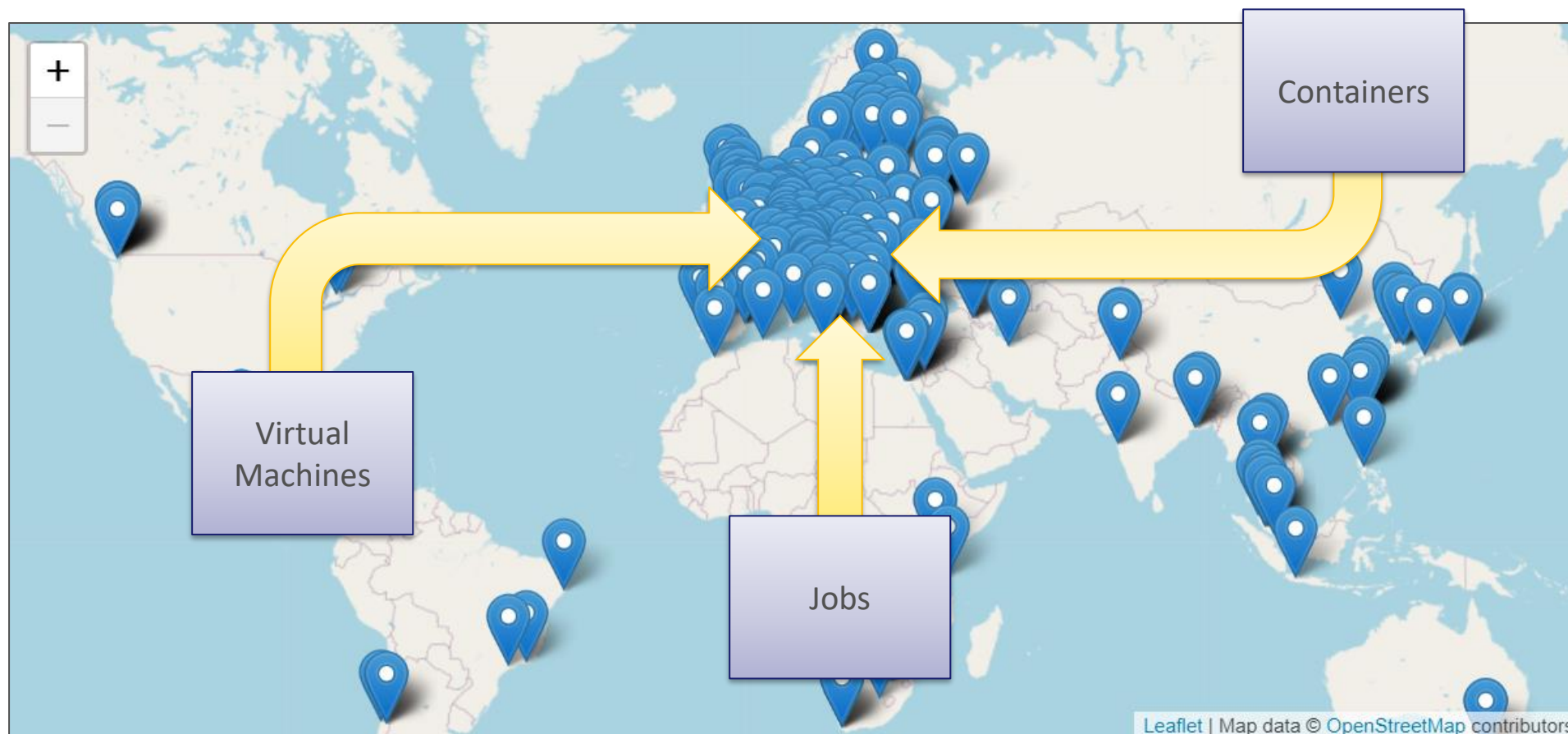
- Federated Compute
 - EGI Cloud Compute
 - EGI Cloud Container Compute
 - EGI High Throughput Compute
- Advanced features and services
 - udocker
 - Workflow Management: DIRAC, PaaS Orchestrator
- Applications services
 - Applications on Demand / EC3
 - Notebooks



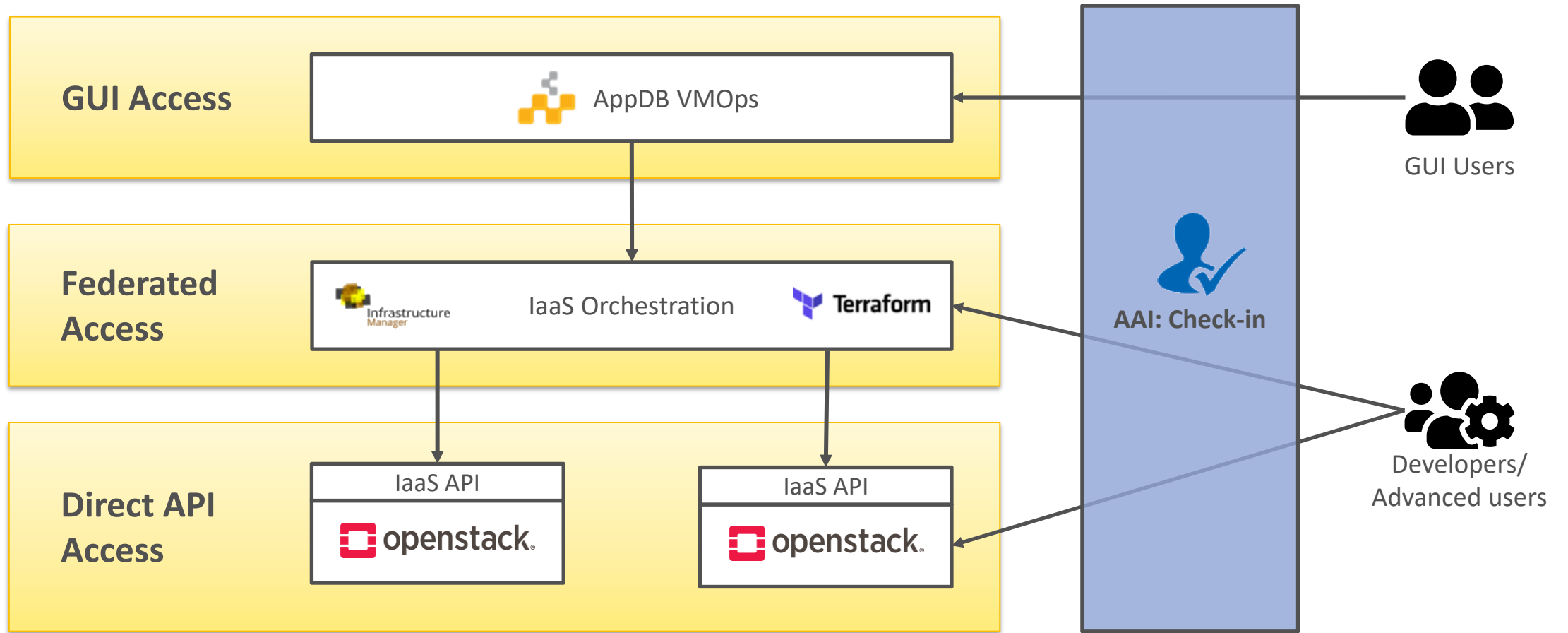


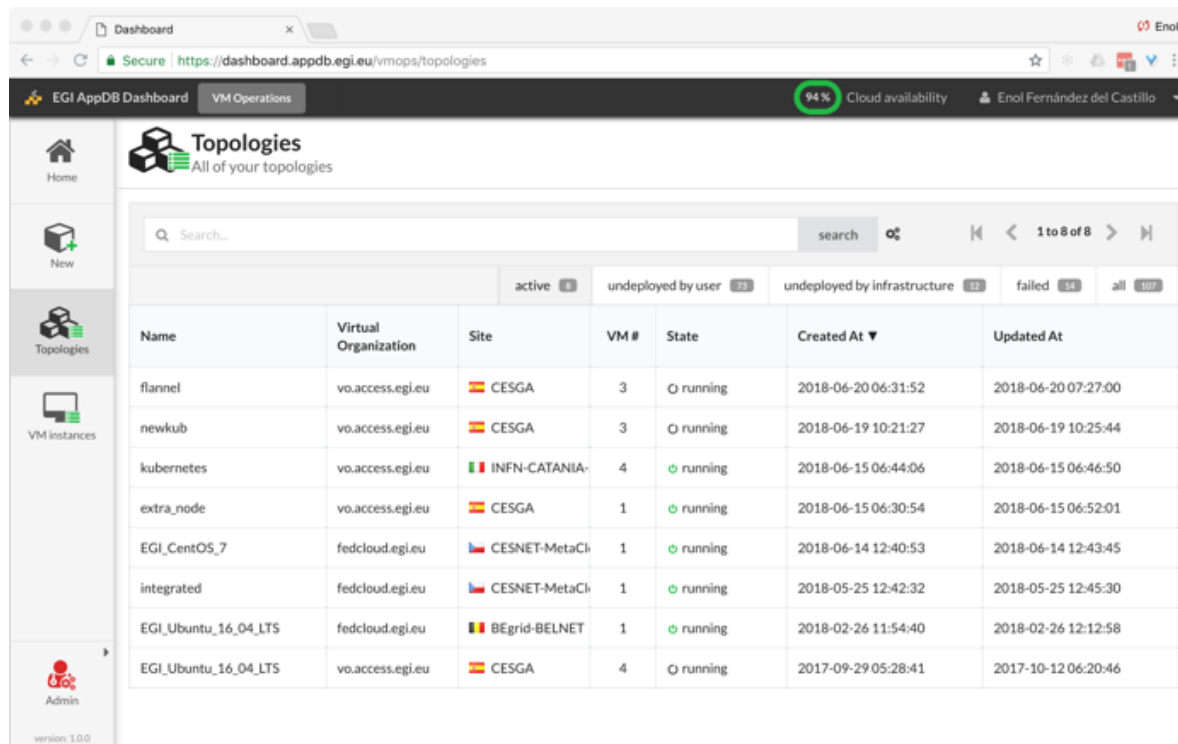
Federated Compute

Run your workloads on the distributed resources on the EGI Federation



- Distributed Infrastructure as a Service (IaaS) powered by the EGI Federated Cloud
 - Allows international collaborations to perform distributed data analysis with VM-based workloads
- Features:
 - Execution of VMs on a distributed infrastructure
 - Federated identity
 - Common VM image catalogue
 - GUI and CLI/API based access
 - Support for IaaS orchestration
 - Central accounting and monitoring





The screenshot shows the 'Topologies' section of the AppDB VMOps dashboard. It features a search bar, a filter menu, and a table of VM instances. The table columns are Name, Virtual Organization, Site, VM #, State, Created At, and Updated At. The table contains 10 rows of data, including instances like 'flannel', 'newkub', 'kubernetes', and 'EGI_Ubuntu_16_04_LTS'.

Name	Virtual Organization	Site	VM #	State	Created At	Updated At
flannel	vo.access.egi.eu	ESGGA	3	running	2018-06-20 06:31:52	2018-06-20 07:27:00
newkub	vo.access.egi.eu	ESGGA	3	running	2018-06-19 10:21:27	2018-06-19 10:25:44
kubernetes	vo.access.egi.eu	INFN-CATANIA	4	running	2018-06-15 06:44:06	2018-06-15 06:46:50
extra_node	vo.access.egi.eu	ESGGA	1	running	2018-06-15 06:30:54	2018-06-15 06:52:01
EGI_CentOS_7	fedcloud.egi.eu	CESNET-MetaCl	1	running	2018-06-14 12:40:53	2018-06-14 12:43:45
integrated	fedcloud.egi.eu	CESNET-MetaCl	1	running	2018-05-25 12:42:32	2018-05-25 12:45:30
EGI_Ubuntu_16_04_LTS	fedcloud.egi.eu	BEgrid-BELNET	1	running	2018-02-26 11:54:40	2018-02-26 12:12:58
EGI_Ubuntu_16_04_LTS	vo.access.egi.eu	ESGGA	4	running	2017-09-29 05:28:41	2017-10-12 06:20:46

- Single Web dashboard to manage VMs in the federation
 - Point-and-click wizard solution to create new VMs
- Integrated with all the features of the federation
- Powered by Infrastructure Manager orchestrator



- Deploy IaaS resources on all kinds of IaaS providers (including OpenStack)
- TOSCA standard support
- Integration with EGI Cloud features (image catalogue, authentication)



- Open Source tool with support for multiple IaaS providers (including OpenStack)
- Near API-level abstraction, but useful for interacting with different providers in a uniform way

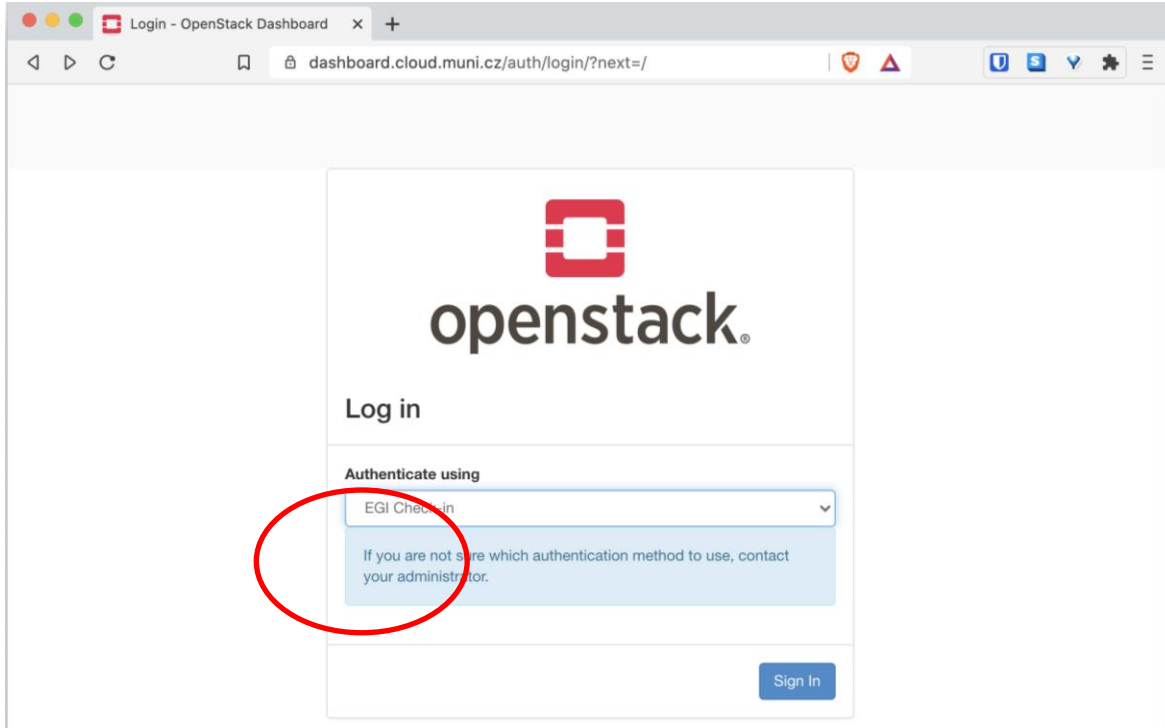
- IM is a service that deploys virtual infrastructures on top of Cloud resources.
- It uses RADL or TOSCA files to describe the infrastructure.
 - Infrastructure as code (IaC)
- The IM automates the deployment, configuration, software installation, monitoring and update of virtual infrastructures.
- It supports a wide variety of back-ends, thus making user applications Cloud agnostic.



- It features DevOps capabilities.
 - Based on Ansible.
 - Provides recipes for common deployments.
 - Also supporting cloud-init scripts.
- IM works as a service that offers several interfaces:
 - XML-RPC and REST APIs.
 - Command-line application.
 - Web-based GUI.
- It is distributed under a GNU GPL v3.0 open source license and its source code is available on GitHub.



<https://github.com/grycap/im>



```

fish /Users/enol
py[legicli] ~> env | grep OS | cut -b1-80
OS_ACCESS_TOKEN=eyJraWQioiJvaWRjIiwiaWxnIjoiaUlyNTYifQ.eyJzdWIiOiI1MjlhODdlNWw1M
OS_AUTH_TYPE=v3oidcaccessstoken
OS_AUTH_URL=https://stack-server.ct.infn.it:35357/v3
OS_IDENTITY_PROVIDER=egi.eu
OS_PROJECT_ID=2669272fcc0c4c6f842a1431eaa5456e
OS_PROTOCOL=openid
py[legicli] ~> openstack image list
+-----+-----+-----+
| ID | Name | Status |
+-----+-----+-----+
| d33874b0-6399-455a-af13-dc1ad3aea03e | DEBIAN 8 | active |
| ae2e46f6-8246-44ca-a370-4f60181b92bb | FEDCLOUD Image for EGI CentOS 7 [CentOS/7/VirtualBox] | active |
| 1c10174f-1334-446f-a42d-6e392b5319de | FEDCLOUD Image for EGI Centos 6 [CentOS/6/KVM] | active |
| c94522a1-91ec-4e10-a3e7-a45870f83acd | FEDCLOUD Image for EGI Docker [Ubuntu/18.04/VirtualBox] | active |
| d68d16db-9e70-44f3-b280-5ef503ac0dd0 | Ubuntu 16.04 | active |
| 8fc9e4ab-b9a1-406b-a2b8-18c5883096e3 | dci_bridge-212.189.145.29 | active |
| 3f2fc801-3b21-41b5-9c2f-a2e162fc25d9 | debian-9-openstack-amd64 | active |
| fed00e4b-cfad-45d0-9736-e70d4f7ad3a7 | galaxy | active |
| a1fa8338-c8b7-4d5b-908a-7e6e1e2d5a0f | galaxy | active |
| 97895def-0e5b-4889-bd6f-c49f3befd648 | larocca | active |
+-----+-----+-----+
py[legicli] ~> █

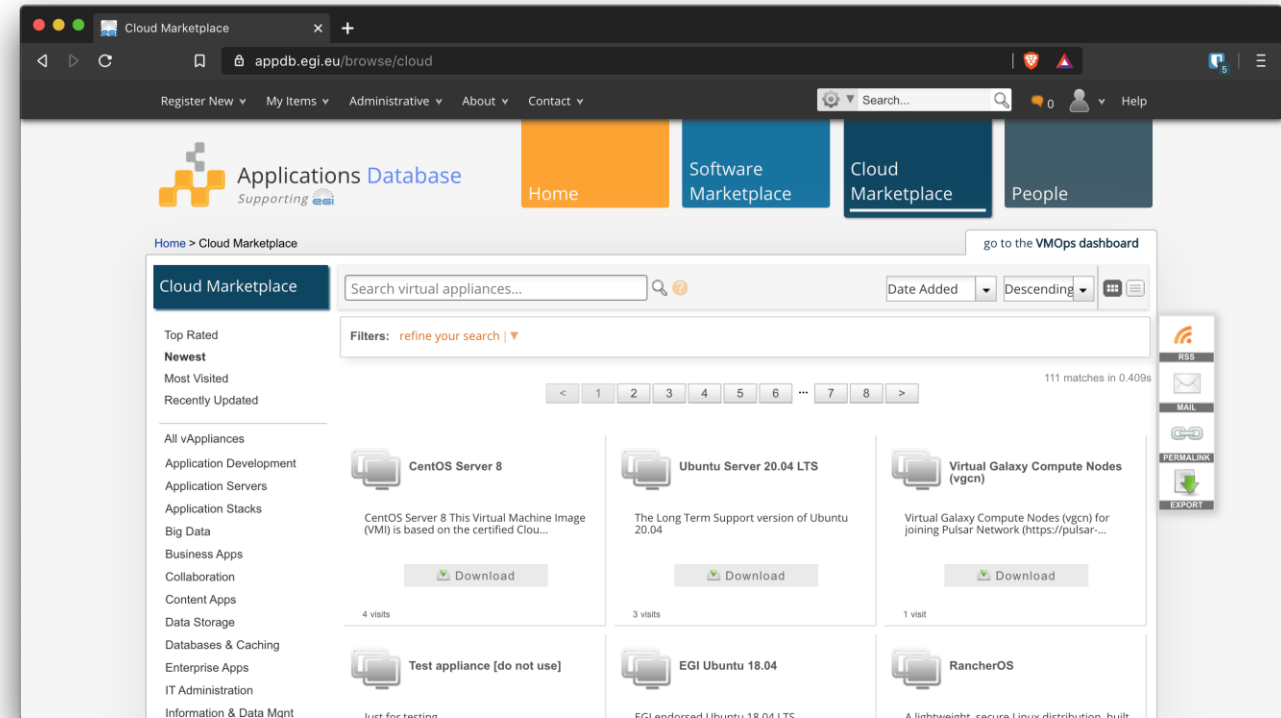
```

- **Common registry for Virtual Appliances (VA)**

- VM image + metadata
- Available for running at the EGI providers or on any hypervisor

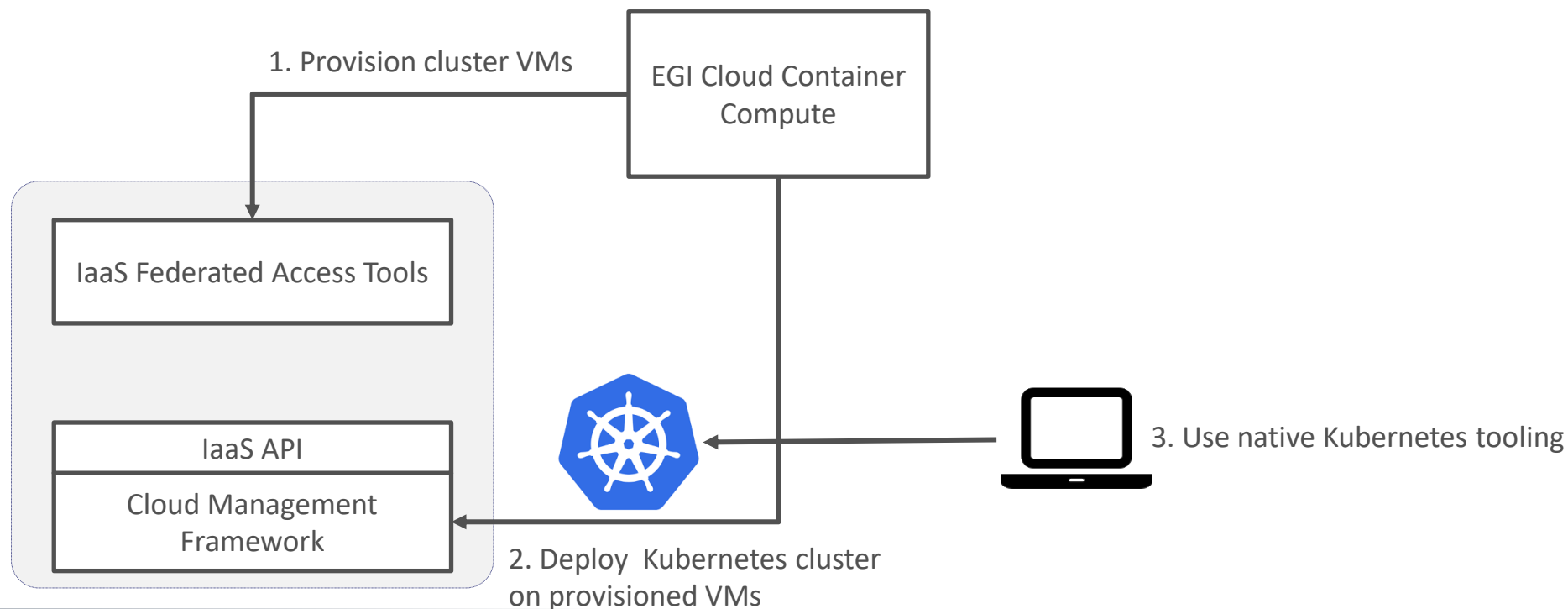
- **Community-level management of VAs**

- Get the needed software on all providers automatically
- Control what can be executed

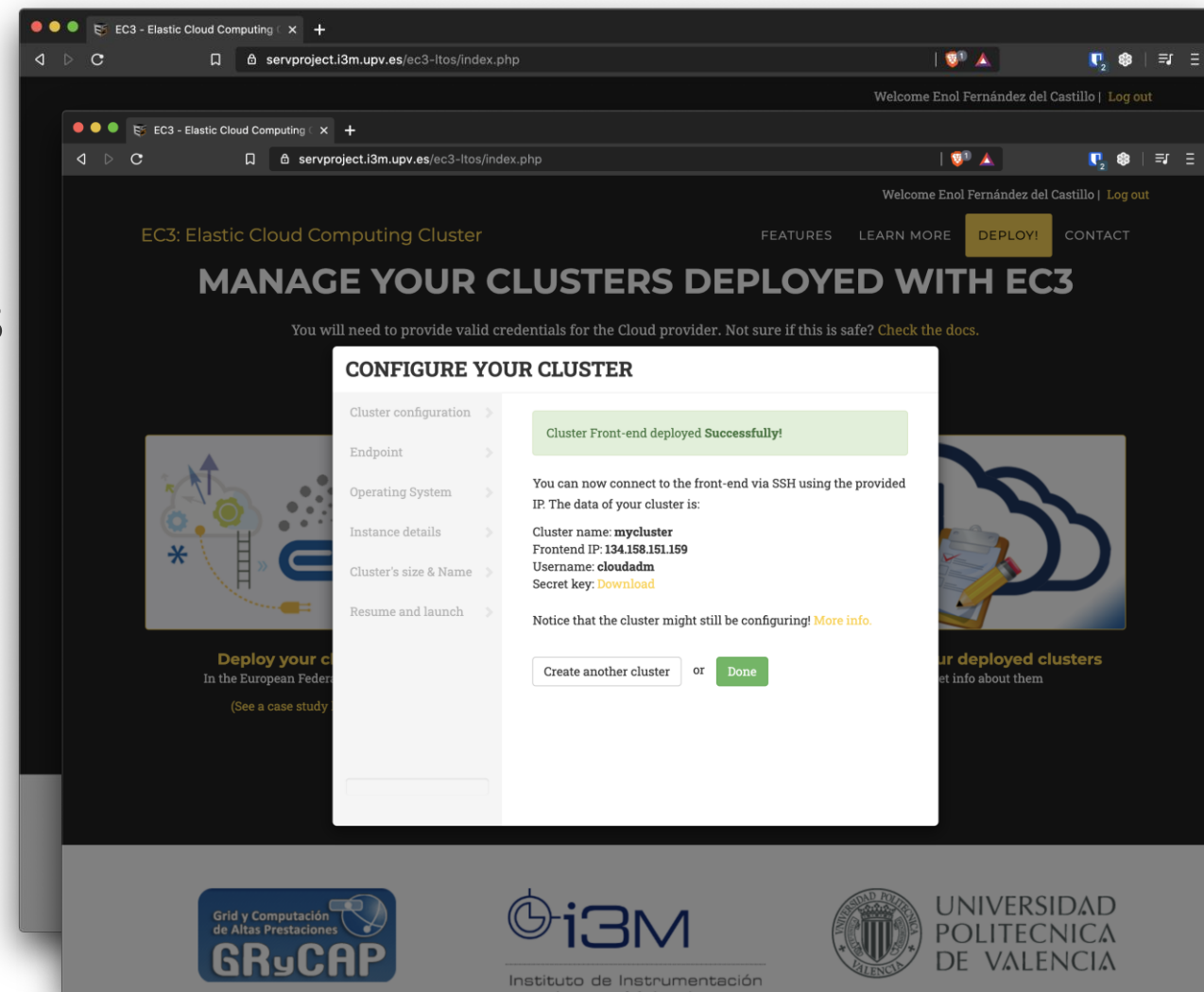


Automated provision of Kubernetes clusters on EGI Cloud Compute providers

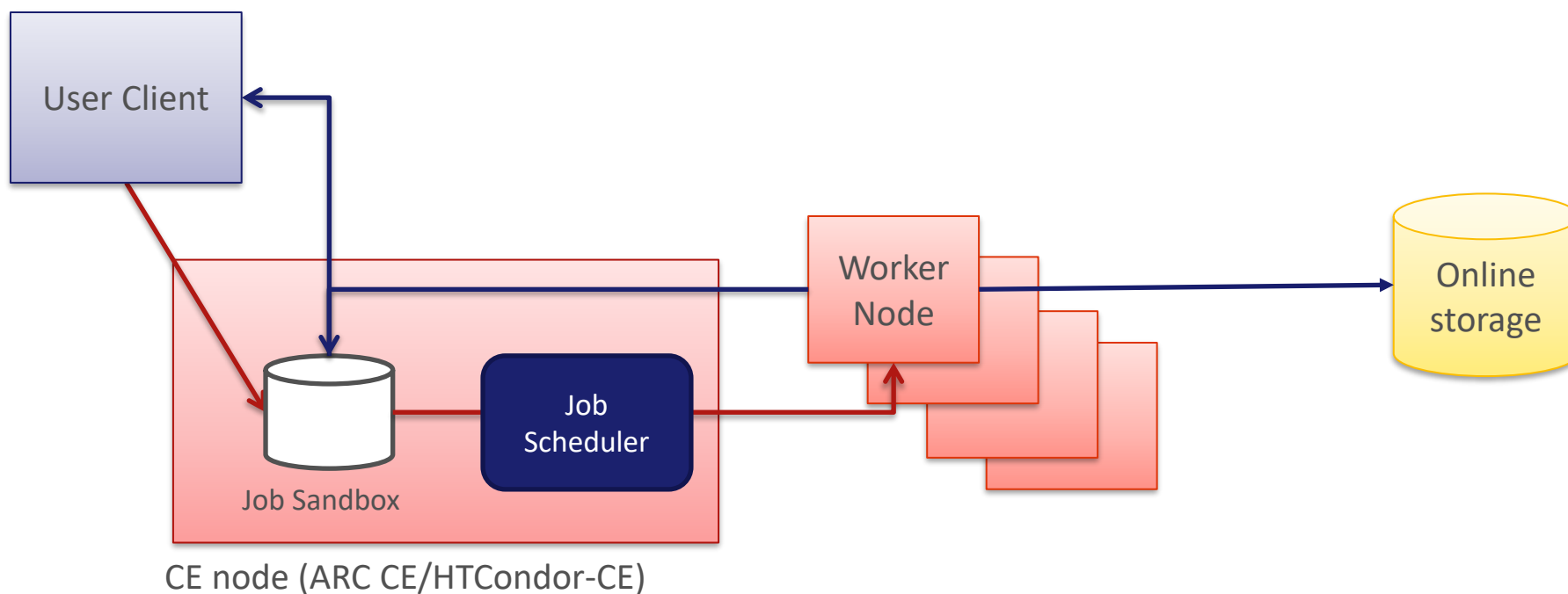
- Built on EC3 for scalable management of resources



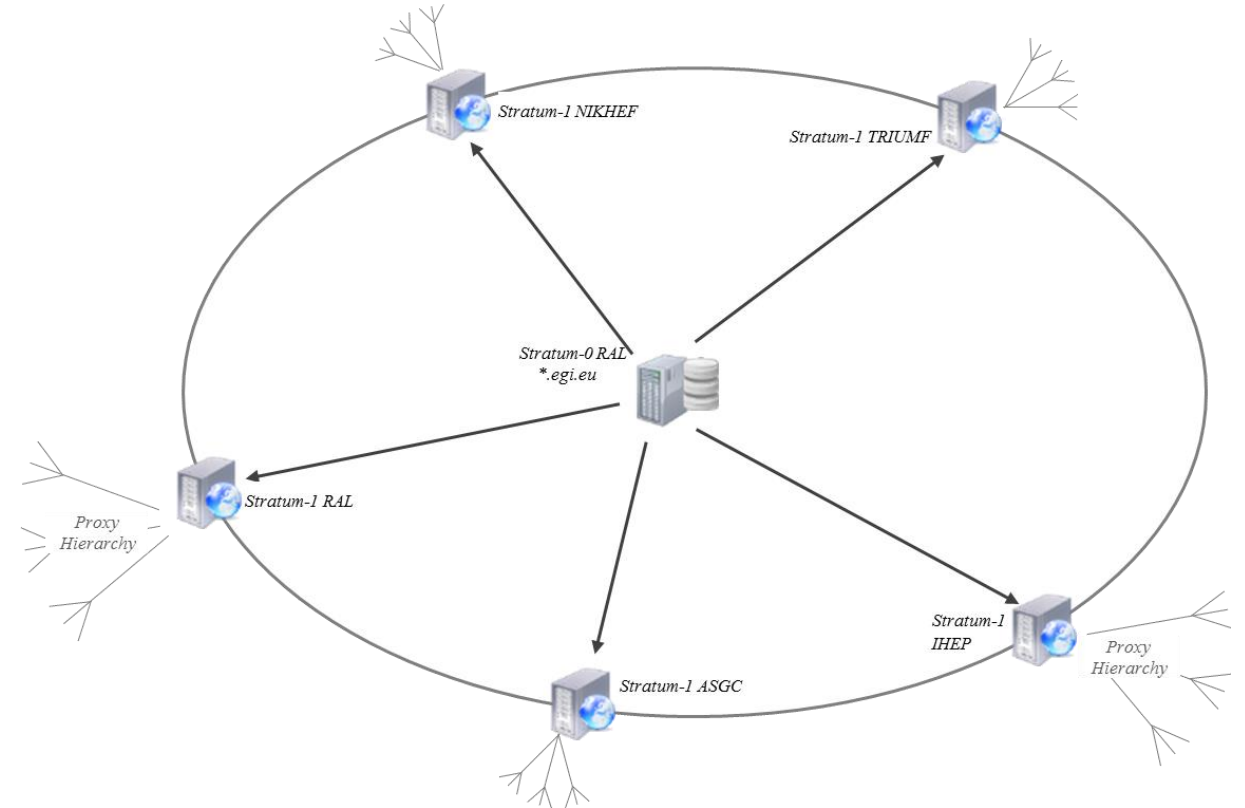
- EC3: Elastic Cloud Computing Cluster
- Deploys and configures Kubernetes
 - Elasticity – can grow/shrink depending on your load
 - GUI and CLI access
 - Integrated with Check-in and AppDB
- Builds on ansible and kubernetes



The EGI High-Throughput compute (HTC) provides users with the capability to access large amounts of computing resources, and to submit hundreds or thousands of computational tasks.



- CernVM FileSystem provides scalable software distribution across the federation
- POSIX read-only filesystem in user space
- Uploader interface for software managers of each CVMFS repository
- Available at all providers of EGI HTC
- Files cached at providers for fast access to frequently used software



	Cloud Compute	Cloud Container Compute	High Throughput Compute
What is it?	Distributed IaaS	Kubernetes on top of EGI Cloud Compute	The <i>grid</i> , a scalable batch system
What do you run?	VMs	(Docker) Containers	Jobs
Software distribution	VM images via AppDB	Container images via registries & helm	CVMFS
Typical workloads	Lift and shift existing applications Specific OS (kernel) requirements Long running servers	Cloud-native containerised applications.	Execution of parallel computing tasks to analyse large datasets.
Pros / Cons	[+] Complete control on resources, run (almost) anything you'd like [-] Complex operation	[+] Hides complexity of Kubernetes setup [-] Kubernetes steep learning curve	[+] No management of resources, just submit jobs [-] Jobs may not match any computational need


Configurability
Abstraction



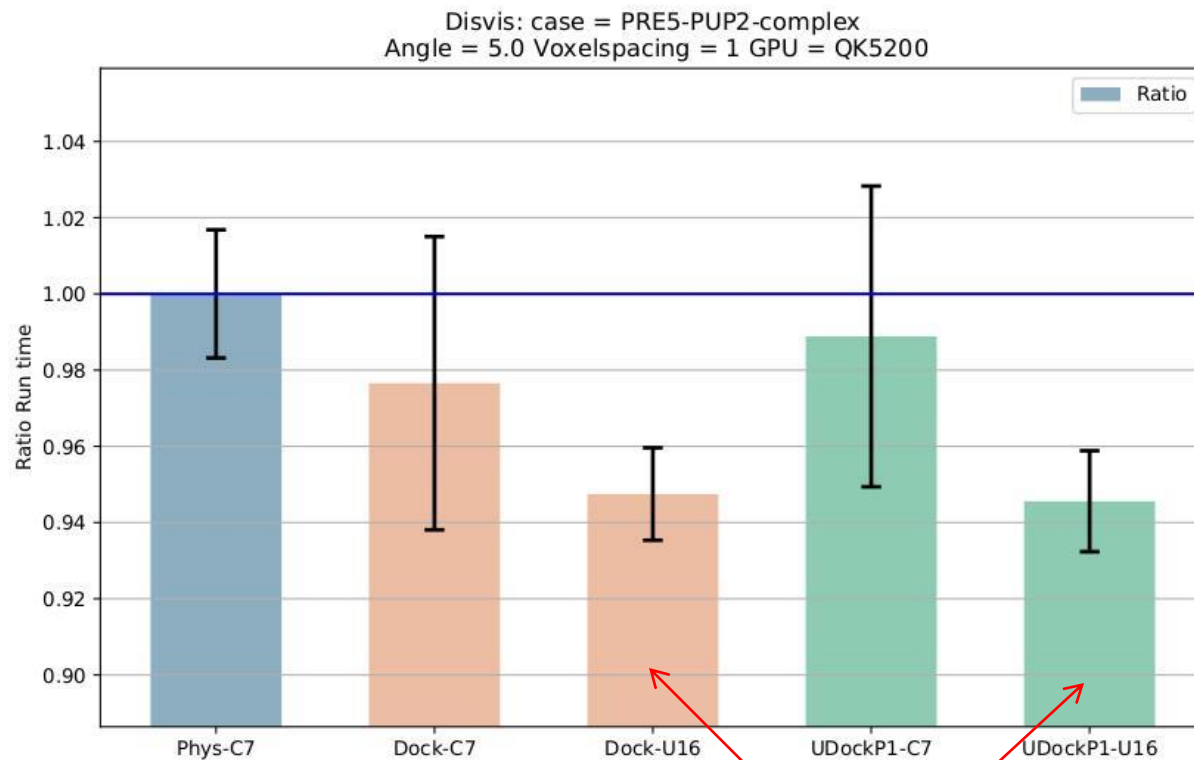
Advanced features and services

- Cloud Compute
 - Specific VM flavors with GPUs available at CESNET-MCC, IFCA-LCG2, IISAS, and NCG-INGRID-PT providers, more coming in the next months
 - VM images with GPU drivers ready to use
- Cloud Container Compute
 - Docker plugin and kubernetes configuration also ready
- High Throughput Compute
 - Selected sites support jobs with GPUs, just tag the jobs

- ◉ Runs applications encapsulated in docker containers:
 - without using docker
 - without requiring privileges
 - without system administrators intervention
 - without additional system software

- ◉ Other characteristics:
 - execution from the command line as a normal user
 - fork and execute model
 - normal process controls and accounting apply
 - suitable for interactive or batch systems

- ◉ Empowers end-users to run applications in containers



Better performance with Ubuntu 16 container

DisVis is being used in production with udocker

Performance with docker and udocker are the same and very similar to the host.

Using OpenCL and NVIDIA GPGPUs

udocker in P1 mode

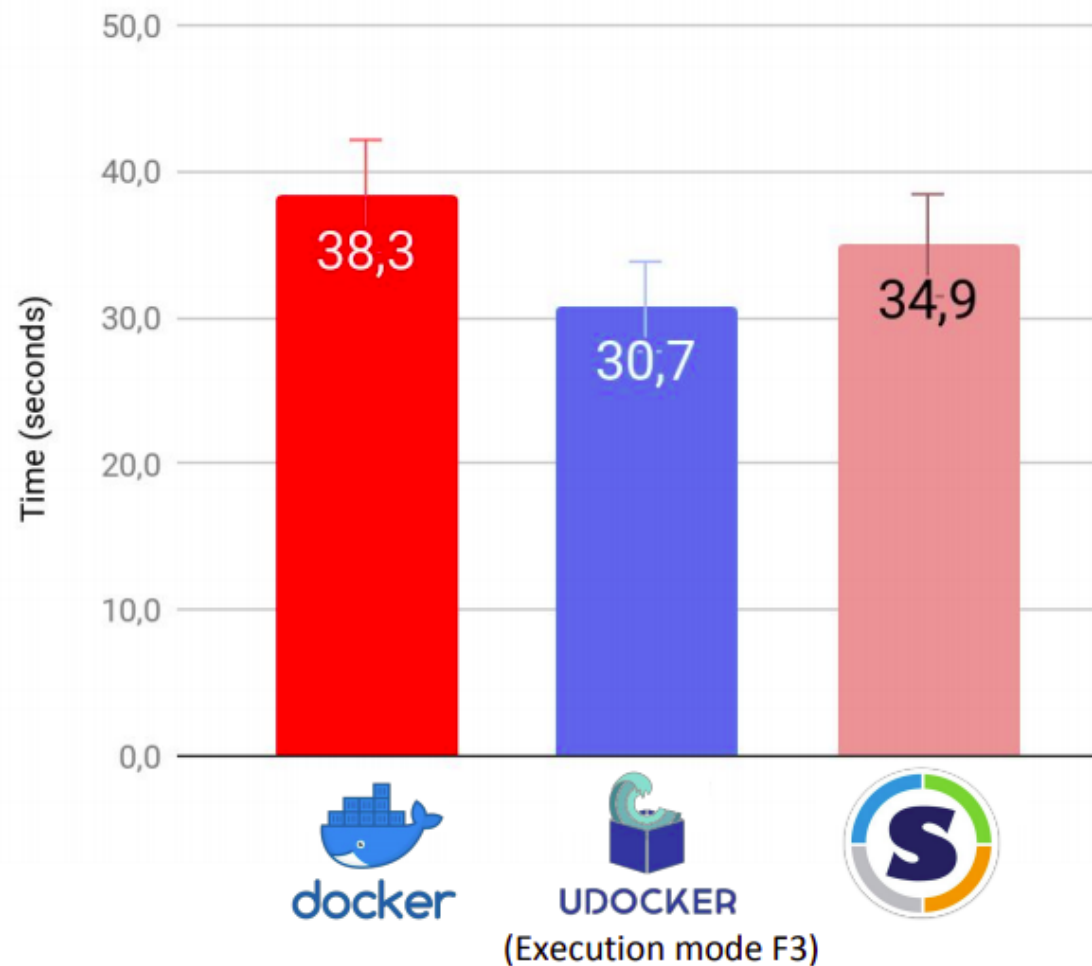
Container:

- Latest GPU version of Tensorflow (from Docker Hub).
- Train a model to recognize handwritten digits (the MNIST data set).

<https://github.com/tensorflow/models.git>



EXECUTION TIME



jorge-tip Update README.md
4 contributors

347 lines (277 sloc) | 13.3 KB

build passing



UDOCKER

udocker is a basic user tool to execute simple docker containers in user space without requiring root privileges. Enables download and execution of docker containers by non-privileged users in Linux systems where docker is not available. It can be used to pull and execute docker containers in Linux batch systems and interactive clusters that are managed by other entities such as grid infrastructures or externally managed batch or interactive systems.

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

- <https://github.com/indigo-dc/udocker/tree/master>
- <https://github.com/indigo-dc/udocker/tree/devel>

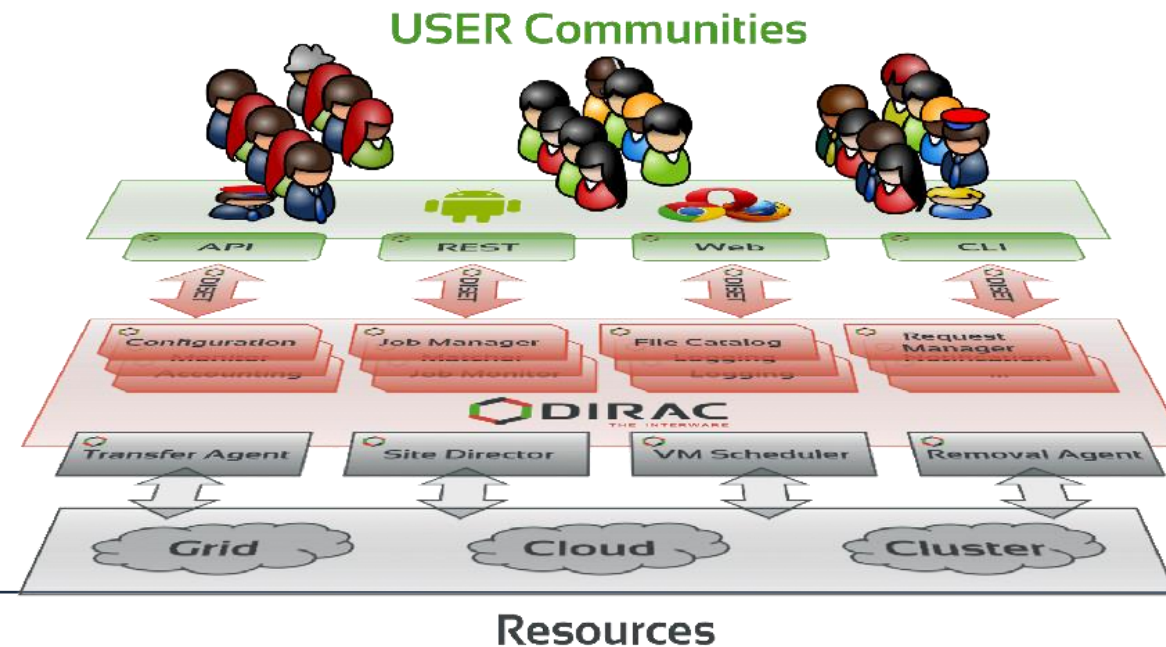
Python 2 & 3

- <https://github.com/indigo-dc/udocker/tree/devel3>

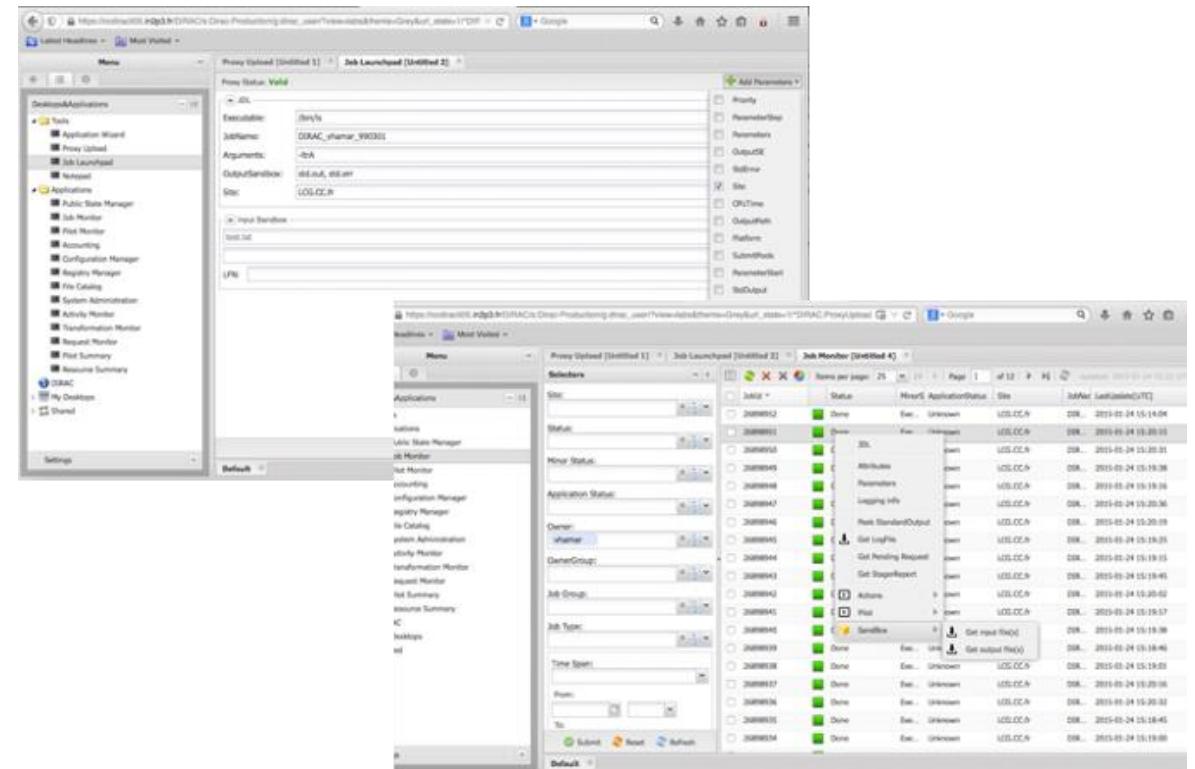
Full documentation also available on github



- DIRAC provides:
 - a framework for building distributed computing systems aggregating multiple types of resources
 - an integrated solution with a reach set of ready to use services for managing computing resources, application workloads and data
- A framework shared by multiple experiments, both inside HEP, astronomy, and life sciences



- ◉ DIRAC provides access to computing and storage resources
 - EGI grid and cloud resources, storage elements
 - User community provided resources (computing clusters, file servers)
- ◉ User access DIRAC with :
 - Command line
`dsub /bin/echo "Hello world!"`
 - Web Portal
 - Python and REST APIs
 - Jupyter notebooks (prototype)
- ◉ Support for massive workload and data operations
- ◉ Services customizable for the needs of particular communities

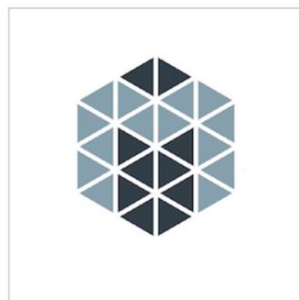




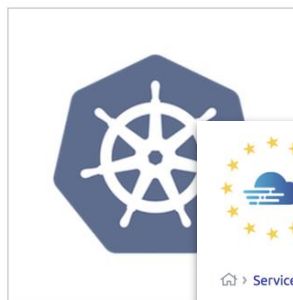
Services



Slurm cluster

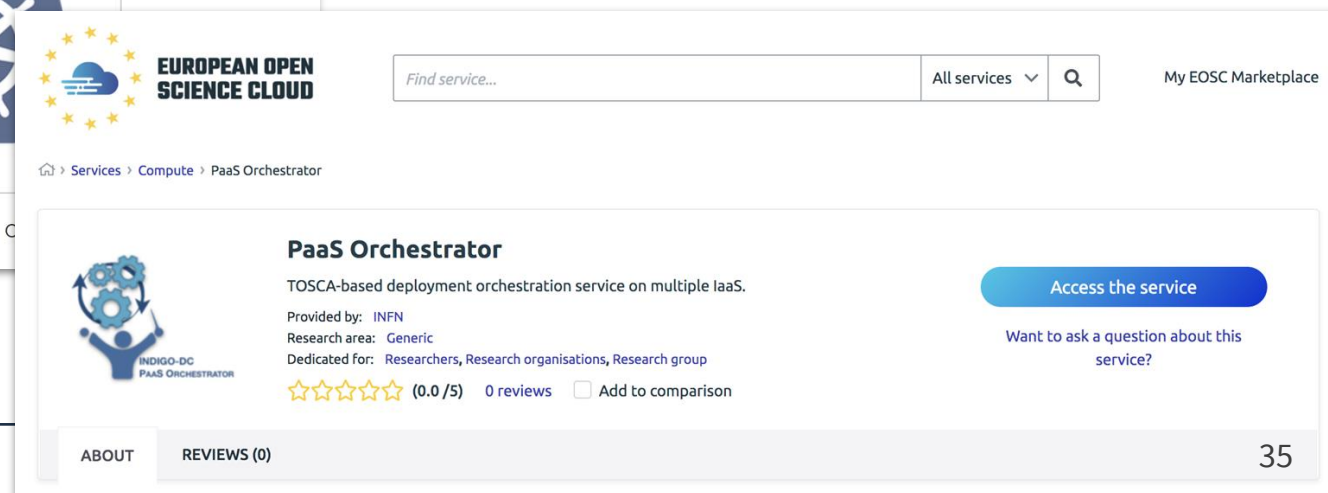


Apache Mesos cluster



Kubernetes cluster

The PaaS Orchestrator service allows to coordinate the provisioning of virtualized compute and storage resources on distributed Cloud infrastructures and the deployment of dockerized services and jobs on Container Orchestration Platforms (like Apache Mesos).



Find service... All services My EOSC Marketplace

Services > Compute > PaaS Orchestrator

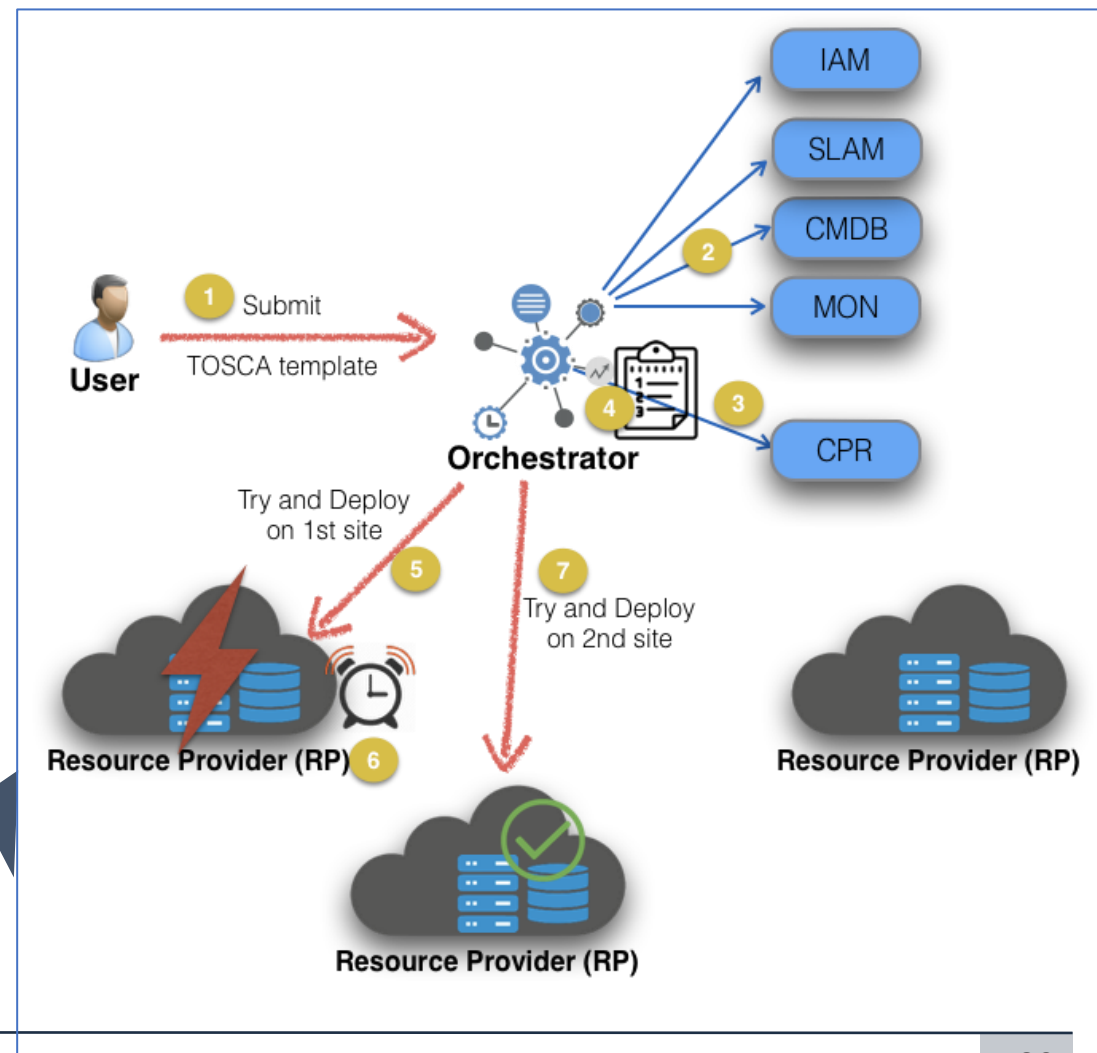
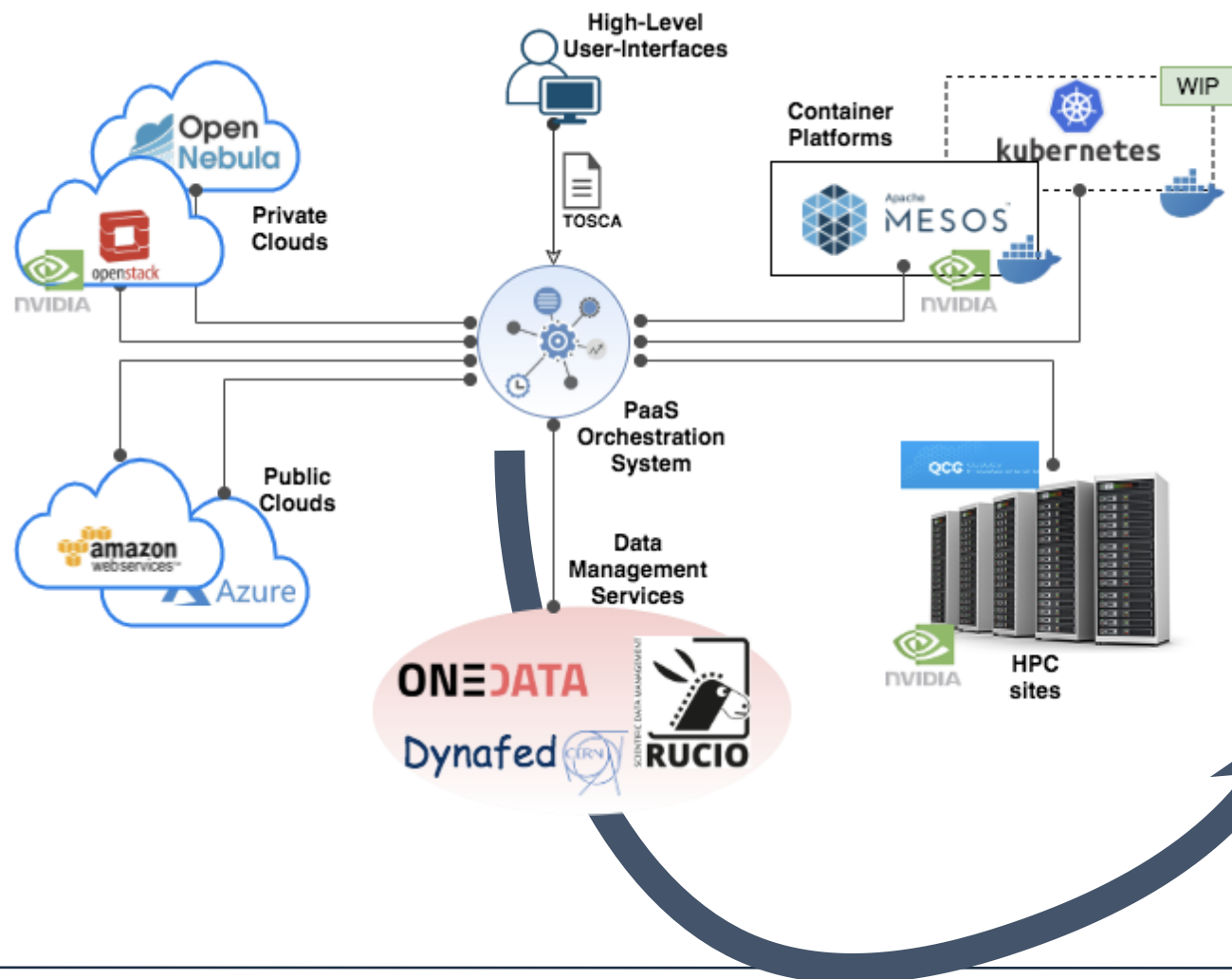
PaaS Orchestrator
TOSCA-based deployment orchestration service on multiple IaaS.
Provided by: INFN
Research area: Generic
Dedicated for: Researchers, Research organisations, Research group
☆☆☆☆☆ (0.0 / 5) 0 reviews Add to comparison

Access the service

Want to ask a question about this service?

ABOUT REVIEWS (0) 35

<https://indigo-paas.cloud.ba.infn.it>



- Fully automated provisioning of virtualized compute and storage resources on different Cloud Management Frameworks (like OpenStack, OpenNebula, AWS, etc.), submission of batch jobs on HPC clusters (through QCG gateways), deployment of dockerized services and jobs on Mesos clusters
 - A plugin for Kubernetes is under development
- Automatic selection of the best resource providers based on criteria like user's SLAs, services availability and monitoring metrics and data location
 - Integration with Onedata, Dynafed and Rucio APIs
- Automatic re-submission of the deployment request in case of failures or timeout (for a configurable number of retries)
- Support for data orchestration workflow with Rucio (Cern)
 - Submission of data replication rules
 - Trigger processing jobs on data ingestion events



Applications

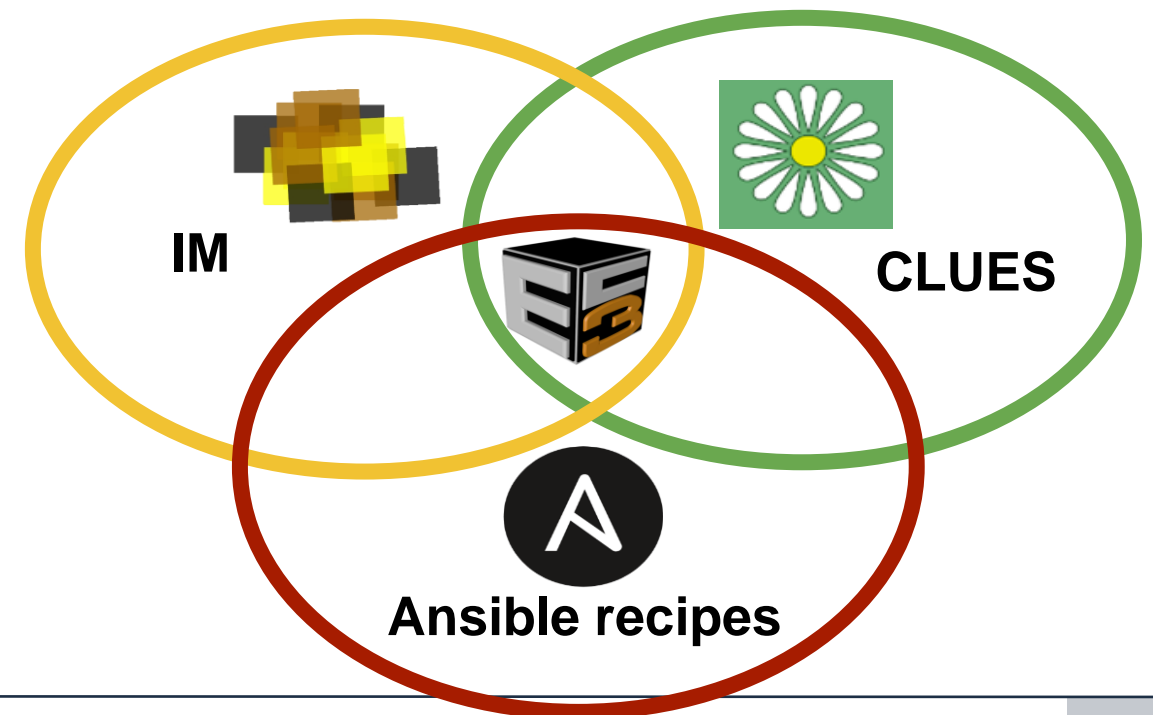
Online scientific applications and application-hosting frameworks with computing and storage for compute-intensive data analysis.

Includes:

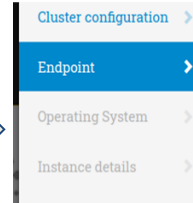
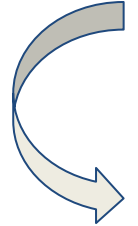
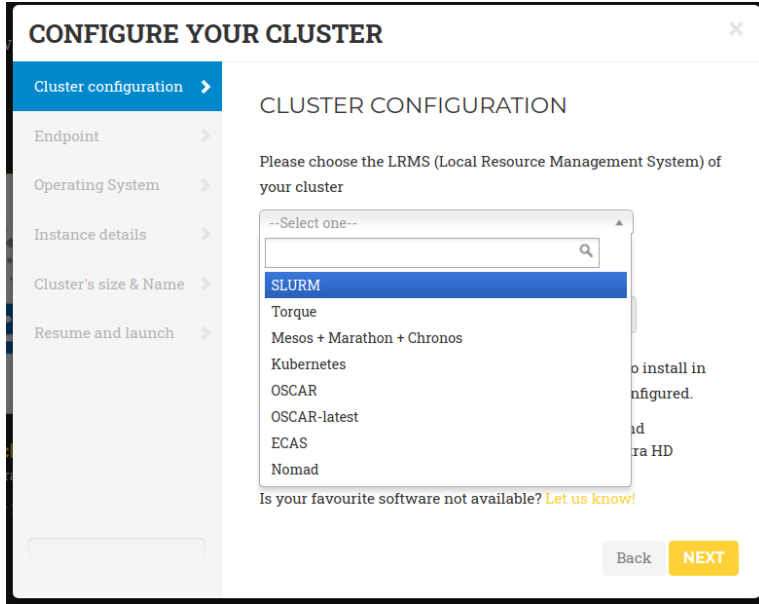
- Scientific applications: Chipster, NAMD, ECAS, many others...
- Access to cloud/storage resources to host and scale up the applications
- Application hosting frameworks (to tun and operate your own applications)
 - Elastic Cloud Compute Cluster (EC3): deployment of elastic virtual clusters in the cloud
 - Science Software on Demand (SSoD): a programmable interface of a RESTful API Server to provide an easy access PaaS layer by leveraging recent web technologies
- AppDB VMOps Dashboard

Elastic Cloud Computing Cluster (EC3) is a tool to deploy virtual elastic clusters on top of Infrastructure as a Service (IaaS) providers, either public (such as Amazon Web Services, Google Cloud or Microsoft Azure) or on-premises (such as OpenNebula and OpenStack, ...)

- **EC3** deploys and configures **virtual elastic** clusters. It relies on **IM** to deploy the machines and on **CLUES** to automatically manage the elasticity.
- Offers a set of predefined templates to configure the resources through **Ansible**:
 - Kubernetes, Mesos, SLURM, Torque, SGE, HTCondor, Nomad.

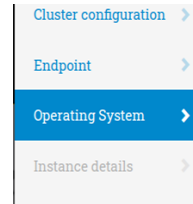


Wizard with 6 simple steps



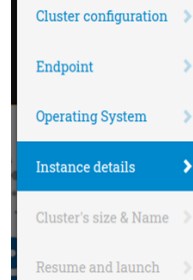
ENDPOINT

FedCloud endpoint:



OPERATING SYSTEM

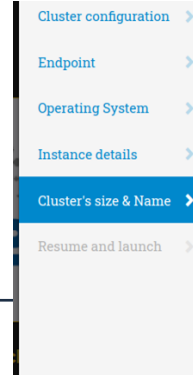
What OS distribution do you like your cluster to have?



INSTANCE DETAILS

Front-end instance type:

Working nodes instance type:



CLUSTER'S SIZE & NAME

Introduce the maximum number of nodes of your cluster (without including the front-end node).
Note that EC3 will initially provision only the front-end node and it will dynamically deploy additional working nodes as necessary.

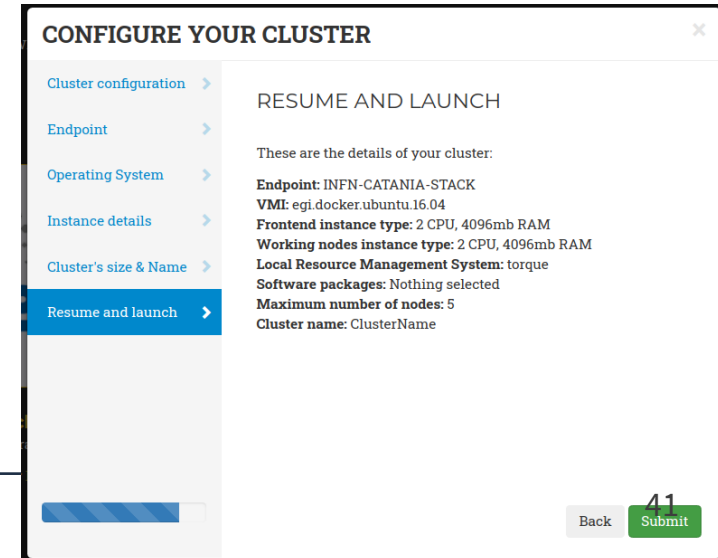
Cluster name (must be unique and without spaces):



AND
DEPLOY
YOUR
CLUSTER



10/11/2020

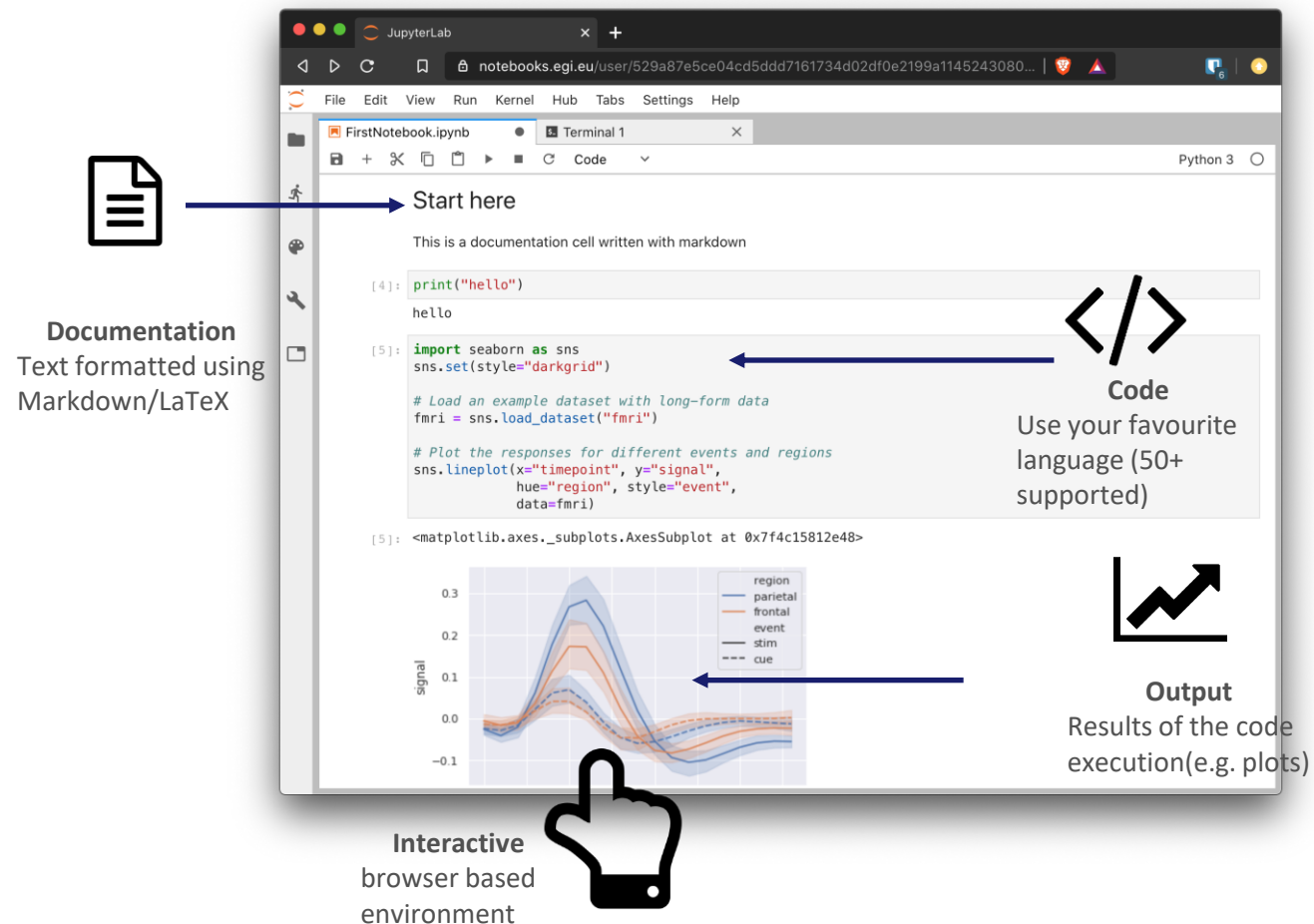


- JupyterHub hosted in the EGI Cloud

- Offers Jupyter notebooks ‘as Service’
- One-click solution: login and start using

- Main Features:

- Login with the EGI AAI Check-In service
- Persistent storage for notebooks
- Use EGI computing and storage resources from your notebooks



Documentation
Text formatted using Markdown/LaTeX

Code
Use your favourite language (50+ supported)

Output
Results of the code execution (e.g. plots)

Interactive browser based environment

```

[4]: print("hello")
hello

[5]: import seaborn as sns
sns.set(style="darkgrid")

# Load an example dataset with long-form data
fmri = sns.load_dataset("fmri")

# Plot the responses for different events and regions
sns.lineplot(x="timepoint", y="signal",
             hue="region", style="event",
             data=fmri)

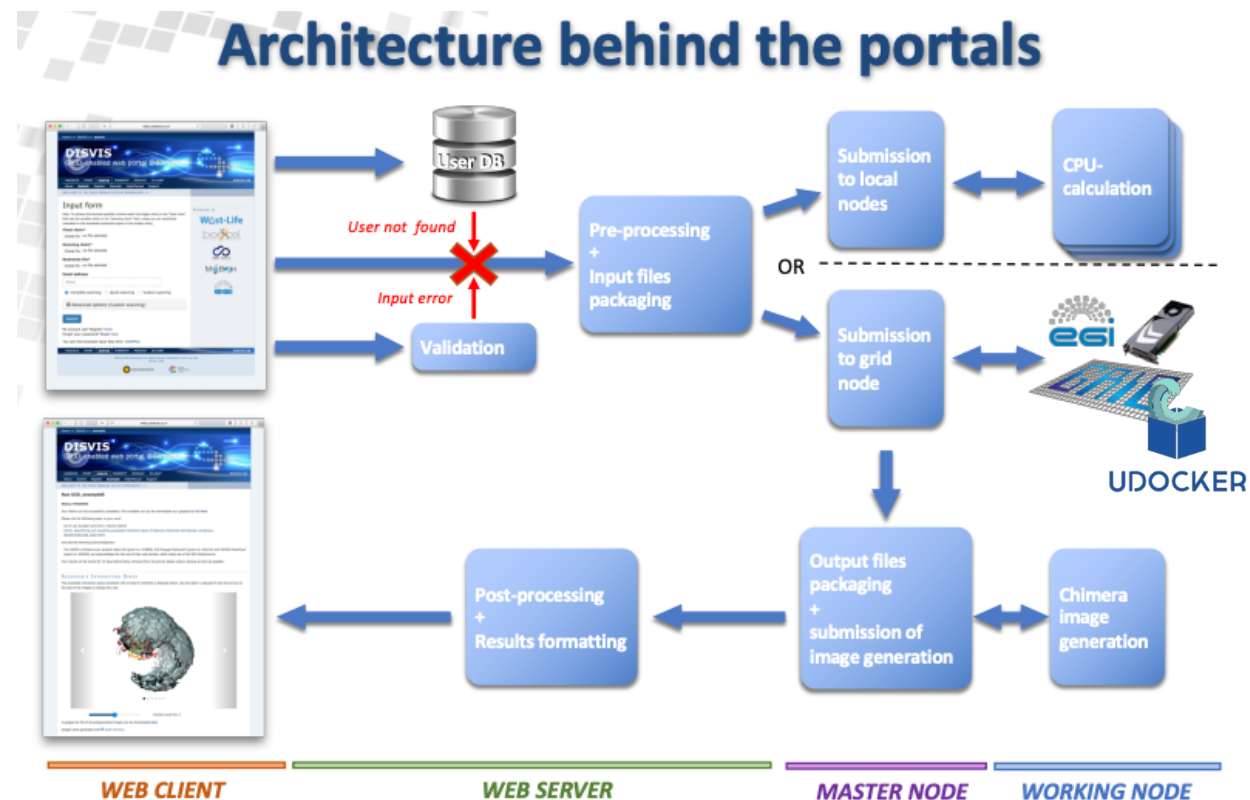
[5]: <matplotlib.axes._subplots.AxesSubplot at 0x7f4c15812e48>
    
```

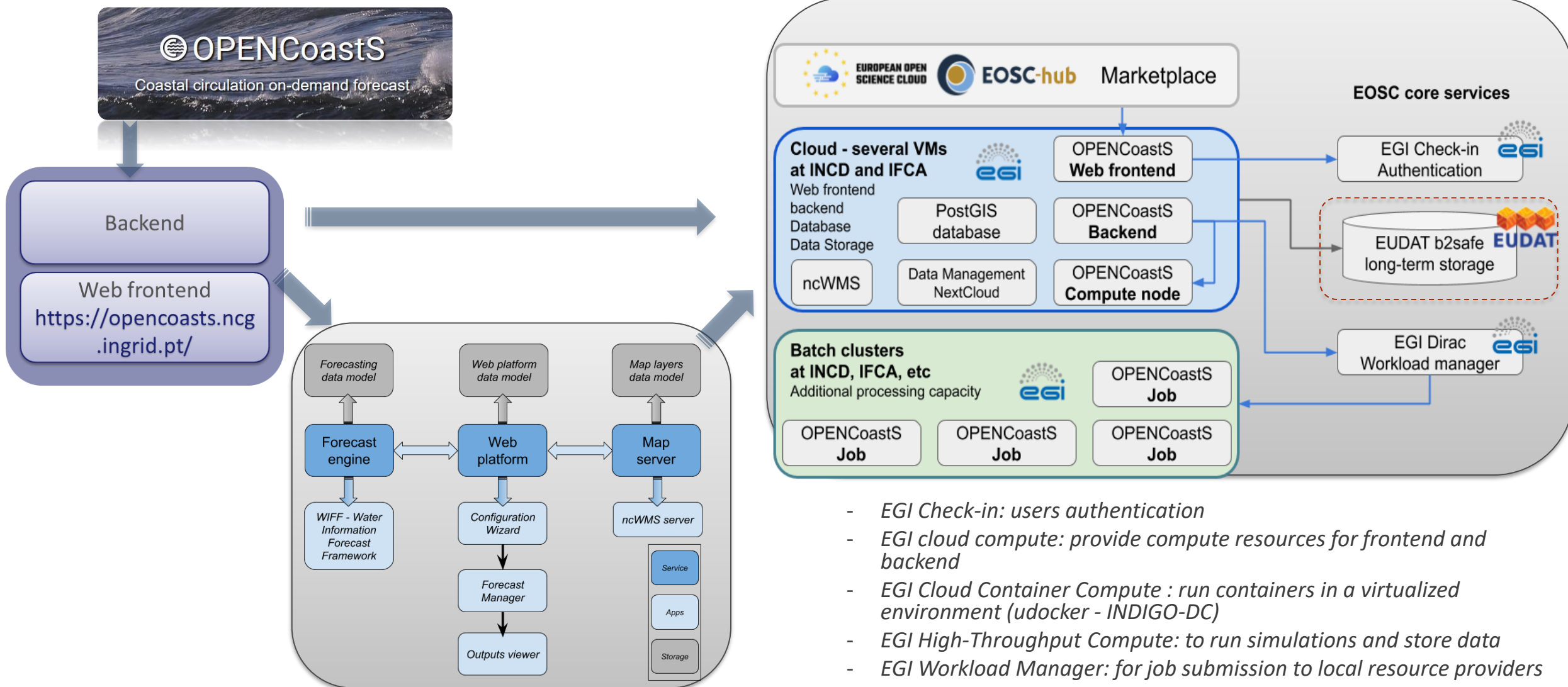



Use of Compute Services in EOSC-hub


The **WeNMR suite** is composed of seven individual platforms:

- **AMPS-NMR**: a web portal for Nuclear Magnetic Resonance (NMR) structures
- **CS-ROSETTA**: to model the 3D structure of proteins
- **DISVIS**: to visualise and quantify the accessible interaction space in macromolecular complexes
- **FANTEN**: for multiple alignment of nucleic acid and protein sequences
- **HADDOCK**: to model complexes of proteins and other biomolecules
- **POWERFIT**: for rigid body fitting of atomic structures into cryo-EM density maps
- **SPOTON**: to identify and classify interfacial residues as Hot-Spots (HS) in protein-protein complexes



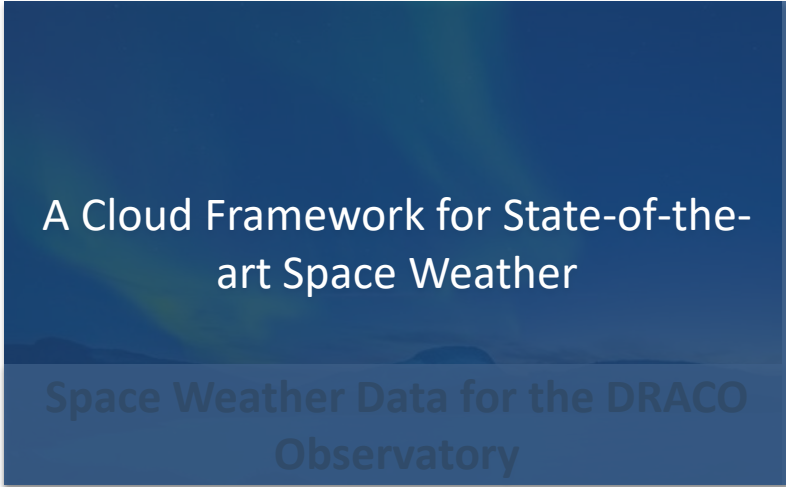


- EGI Check-in: users authentication
- EGI cloud compute: provide compute resources for frontend and backend
- EGI Cloud Container Compute : run containers in a virtualized environment (udocker - INDIGO-DC)
- EGI High-Throughput Compute: to run simulations and store data
- EGI Workload Manager: for job submission to local resource providers

A background image of a seaport with large cranes and shipping containers, overlaid with a dark blue semi-transparent filter.

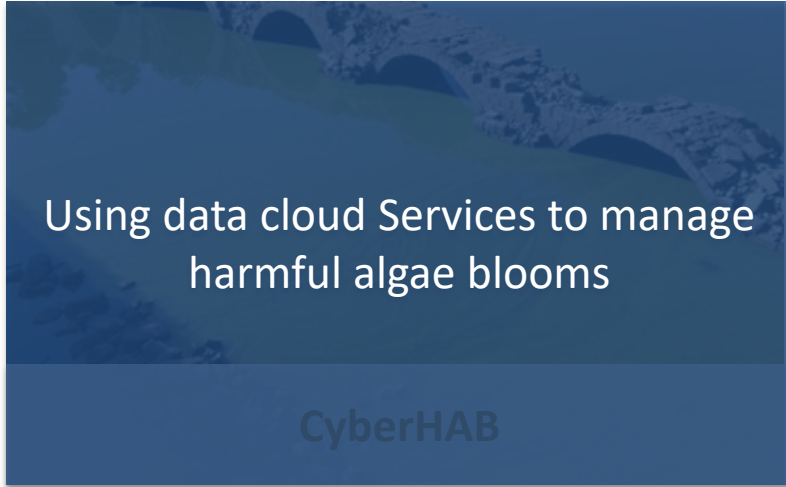
Improving safety and operational performance of seaports

ACTION Seaport

A background image of a colorful aurora borealis in a dark sky, overlaid with a dark blue semi-transparent filter.

A Cloud Framework for State-of-the-art Space Weather

Space Weather Data for the DRACO Observatory

A background image of a large, green, circular algae bloom in the ocean, overlaid with a dark blue semi-transparent filter.

Using data cloud Services to manage harmful algae blooms

CyberHAB

A background image of a world map with various data points and lines, overlaid with a dark blue semi-transparent filter.

Solutions to secure online services from Botnets Attacks

Bot Mitigation Engine

A background image of a person in a dynamic pose, possibly a gymnast or athlete, overlaid with a dark blue semi-transparent filter.

Analysing Sport Performance through a Cloud-hosted platform

Sport Smart Video Analysis

A background image of a modern living room interior with a brick wall and furniture, overlaid with a dark blue semi-transparent filter.

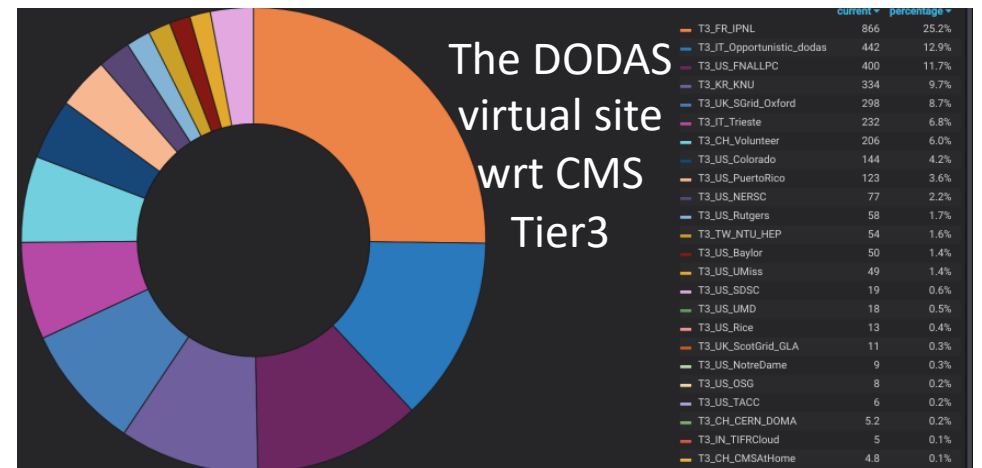
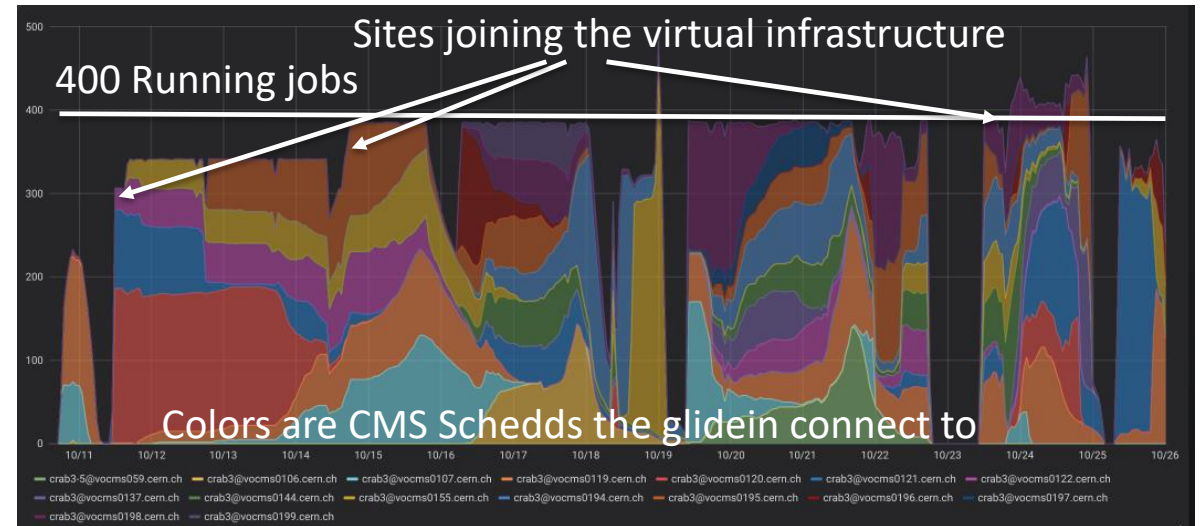
Platform-as-a-Service Data Analytics for the Furniture Industry

Furniture Enterprise Analytics Datafurn

DODAS (Dynamic On Demand Analysis Service) is a PaaS that instantiates on-demand container-based clusters to execute software applications

DODAS was used to build a single virtual infrastructure on top of 4 EGI Cloud providers

- Executing actual experiment workload
- Distributing the load across providers



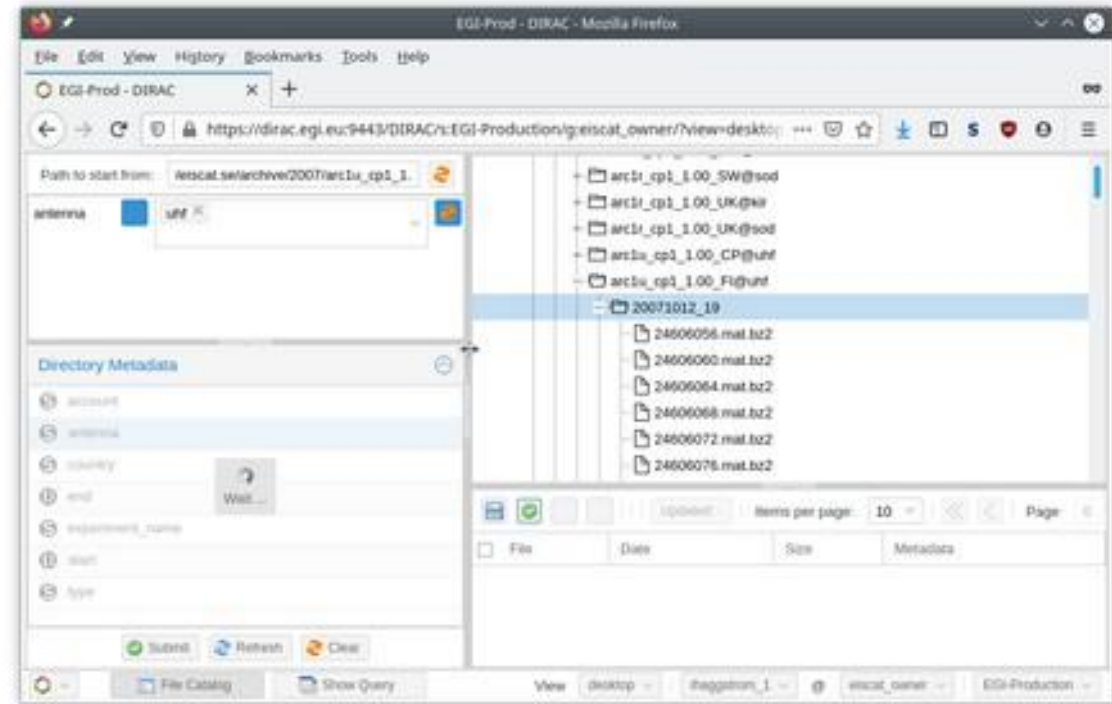
- Show how federated EOSC resources can facilitate a range of Sentinel data applications across agricultural user domains
 - Use of Big Data Analytics to multi-annual high resolution Copernicus Sentinel data and EU open access reference data sets
 - Agriculture use case: cross-border EU region of NL and Germany
- Objectives:
 - (1) project EOSC as reference platform hosting permanent Sentinel data archive, (2) methods to handle a wide set of technical requirements for Sentinel data use



The EISCAT_3D Competence Centre



- deploy and integrate necessary tools, services and infrastructures
 - Data management and processing
- DIRAC interware
 - integration component
 - a single access point towards e-Infrastructures
- EUDAT's B2 services
 - unify the data management
 - discovery system across different storages
 - storage access management
- EGI and INDIGO services
 - deploying the software stack on HPC/HTC systems including release management
- provide secondary services for production operation
 - user authentication and access control

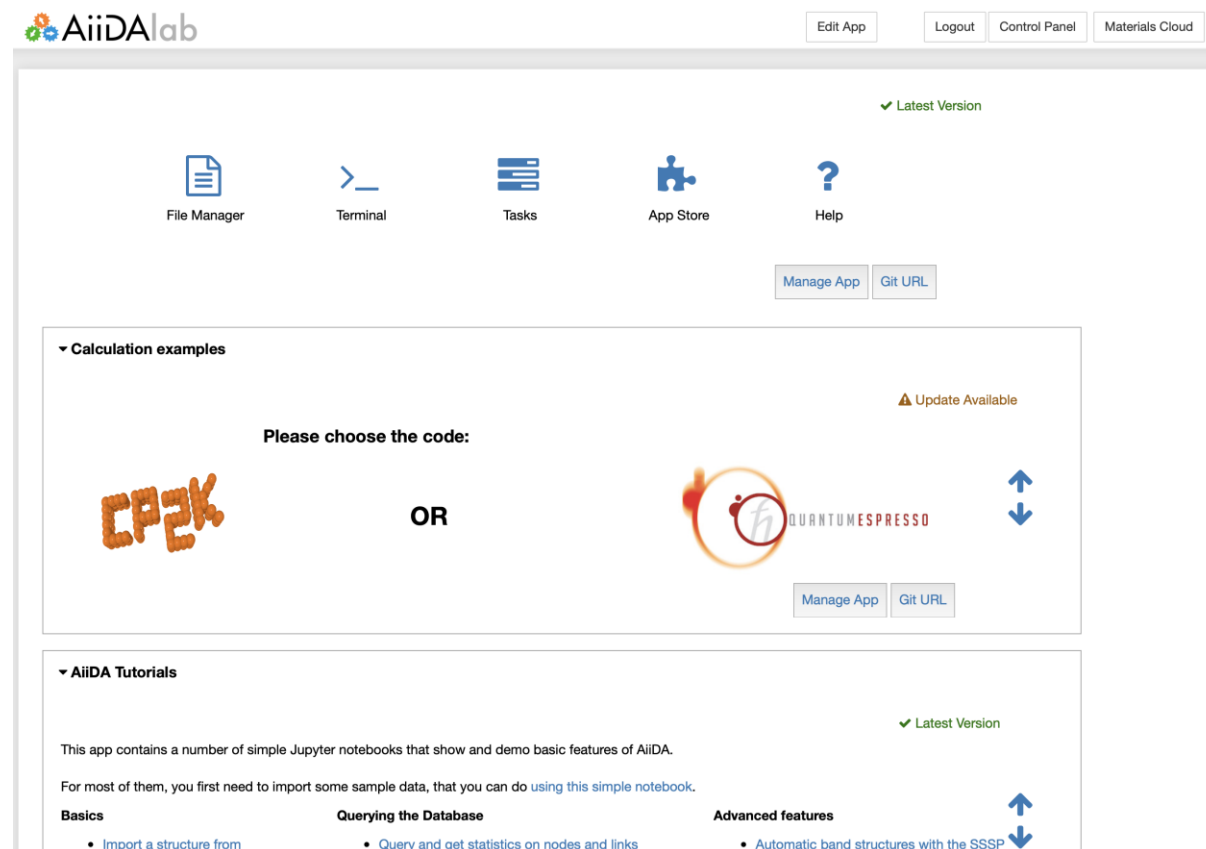


- Cloud data generation platform and data analysis platform:

- Based on AiiDA + Jupyter + App Mode

In EOSC-hub:

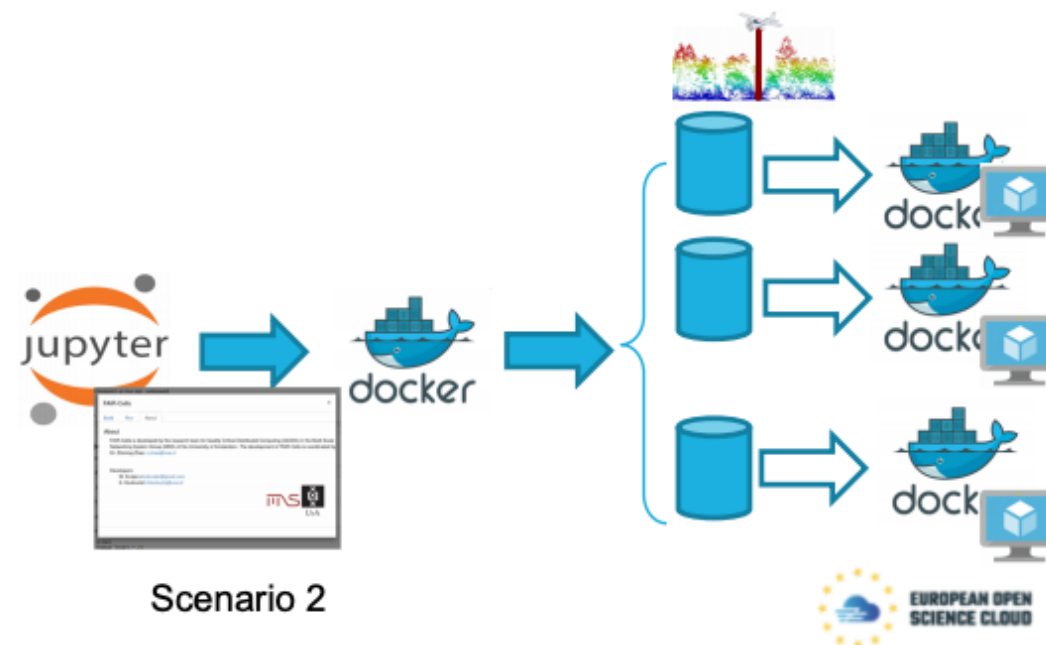
- Provide access to Kubernetes infrastructure to support the deployment and operation of an open AiiDA lab instance
- Support the authentication and authorisation of users into AiiDA lab with EOSC users



The screenshot shows the AiiDA lab web interface. At the top, there is a navigation bar with the AiiDA lab logo and buttons for 'Edit App', 'Logout', 'Control Panel', and 'Materials Cloud'. Below the navigation bar, there is a row of icons for 'File Manager', 'Terminal', 'Tasks', 'App Store', and 'Help'. Underneath these icons are buttons for 'Manage App' and 'Git URL'. The main content area is divided into two sections. The first section, titled 'Calculation examples', features a central prompt 'Please choose the code:' with two options: 'CP2K' (represented by a stylized orange logo) and 'QUANTUMESPRESSO' (represented by a red and white logo). There are 'Manage App' and 'Git URL' buttons for each option, and a green checkmark indicating the 'Latest Version' for both. The second section, titled 'AiiDA Tutorials', contains a paragraph of text and a list of tutorial topics: 'Basics', 'Querying the Database', and 'Advanced features'. Each topic has a list of sub-topics and a green checkmark indicating the 'Latest Version'.

EOSC DevOps framework and virtual infrastructure for ENVRI-FAIR common FAIR data services

- ENVRI-FAIR connects the Environmental Research Infrastructure (ENVRI) community to the EOSC
- Use cases:
 - Automated Cloud execution for data workflow
 - Continuously testing and integration for ENVRI services
 - Notebook based environment for FAIR data access and processing

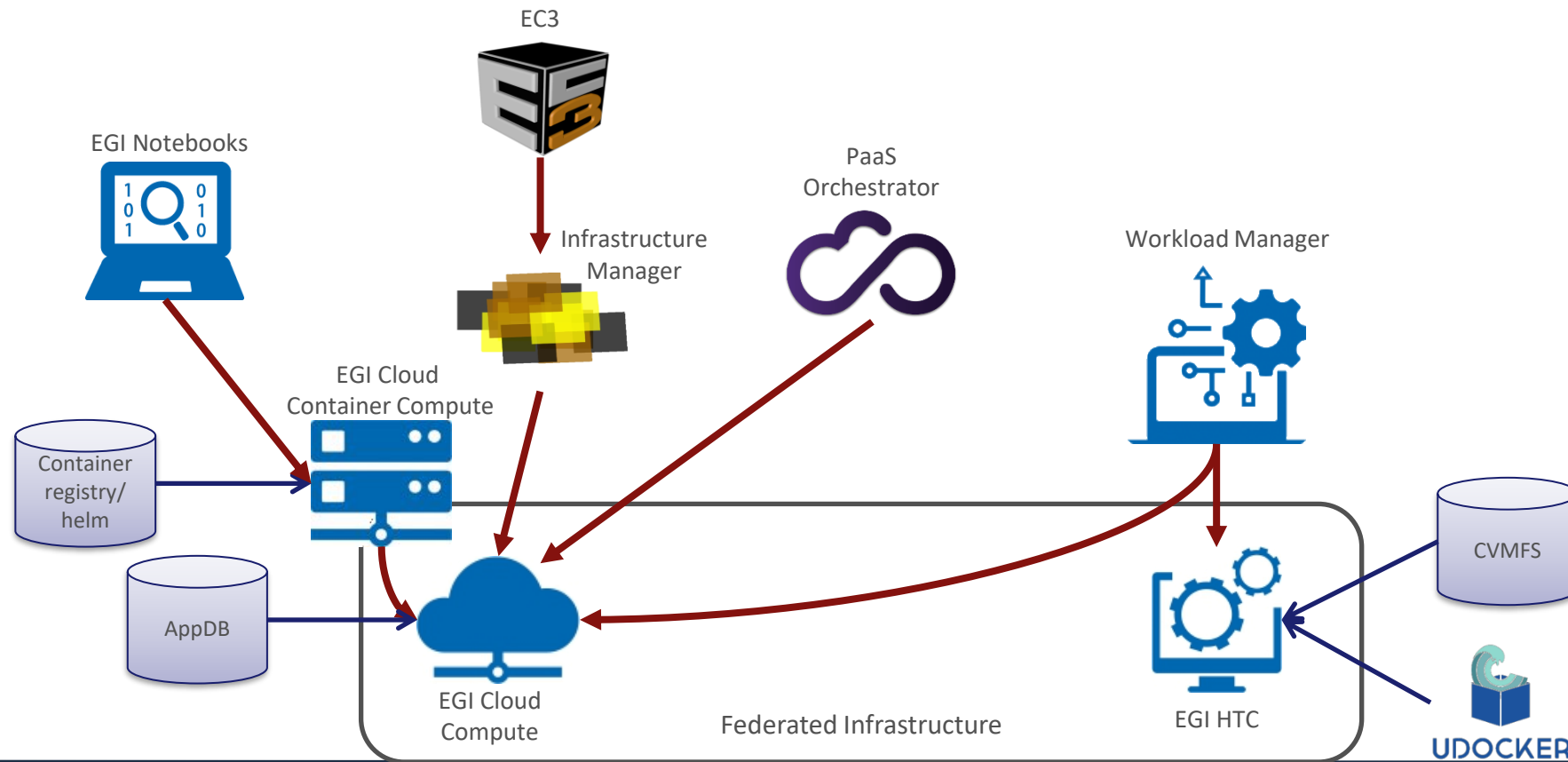




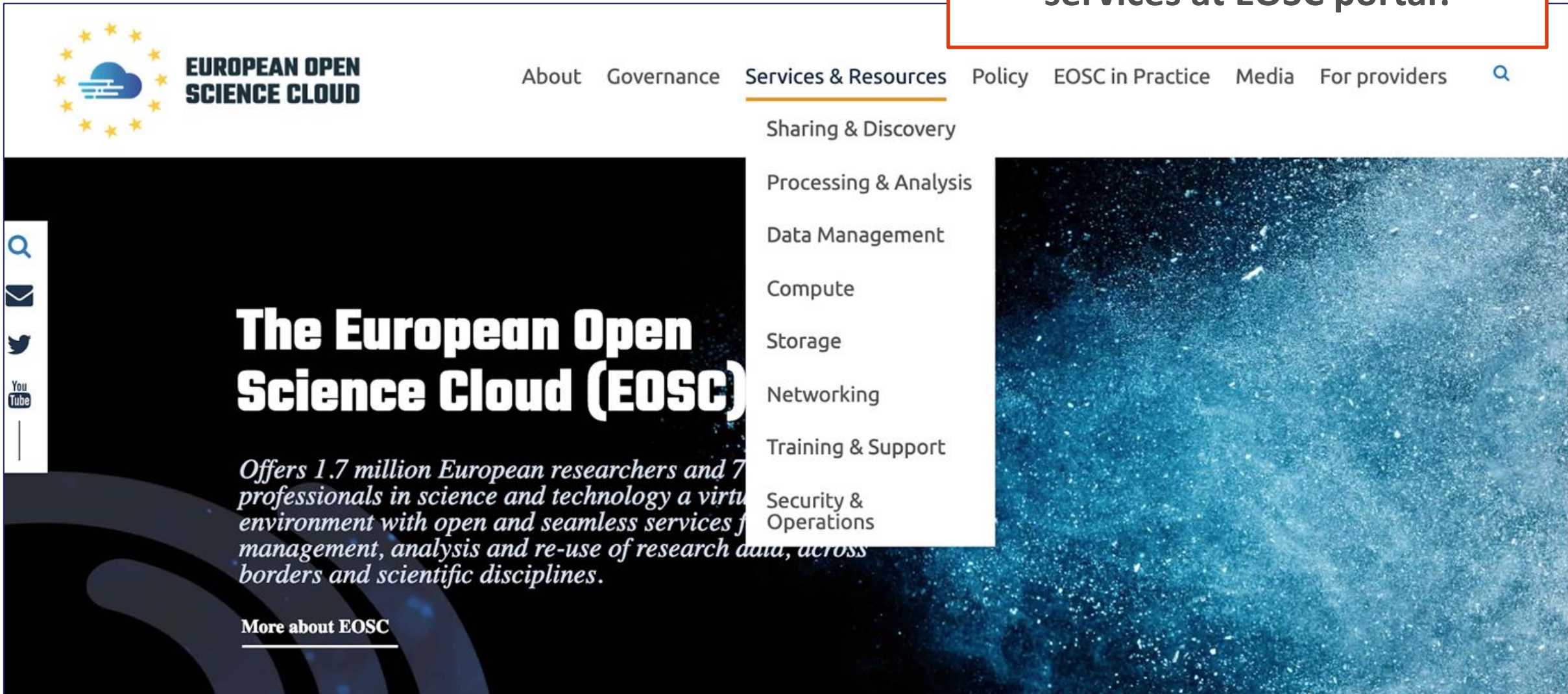
Summary

EOSC-hub Summary

- EOSC-hub brings together computing services into a federated solution: from powerful and fully-customisable Infrastructure as a Service to high-level and interactive computing platforms.



Get Started by requesting services at EOSC portal!



The screenshot shows the EOSC Portal website. At the top left is the European Open Science Cloud logo, featuring a blue cloud with white lines and a yellow sun, surrounded by twelve yellow stars. To its right is the text "EUROPEAN OPEN SCIENCE CLOUD". The navigation menu includes "About", "Governance", "Services & Resources" (which is underlined and has a dropdown menu), "Policy", "EOSC in Practice", "Media", and "For providers". A search icon is located on the far right of the navigation bar. The dropdown menu for "Services & Resources" lists: "Sharing & Discovery", "Processing & Analysis", "Data Management", "Compute", "Storage", "Networking", "Training & Support", and "Security & Operations". The main content area features a large blue and black background with the text "The European Open Science Cloud (EOSC)" in white. Below this, it says "Offers 1.7 million European researchers and 700,000 professionals in science and technology a virtual environment with open and seamless services for the management, analysis and re-use of research data, across borders and scientific disciplines." At the bottom left of the main content area, there is a link "More about EOSC" with a horizontal line underneath. On the left side of the page, there is a vertical sidebar with icons for search, email, Twitter, and YouTube.

**Thank you
for your attention!**

Questions?



EOOSC-hub

 eosc-hub.eu  [@EOOSC_eu](https://twitter.com/EOOSC_eu)



This material by Parties of the EOOSC-hub Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).