Increasing Quality and Capacity of Thematic Services in EOSC-SYNERGY

19/10/2021

Ignacio Blanquer (UPV), Alberto Azevedo (LNEC), Thiago Emmanuel Pereira (UFCG), Manuel Pavesio-Blanco (INDRA), Salvador Capella-Gutierrez (BSC), Laura del Cano (CSIC-CNB), Rubio-Montero Antonio Juan (CIEMAT), Jan Astalos (IISAS), Tobias Kerzenmacher (KIT).

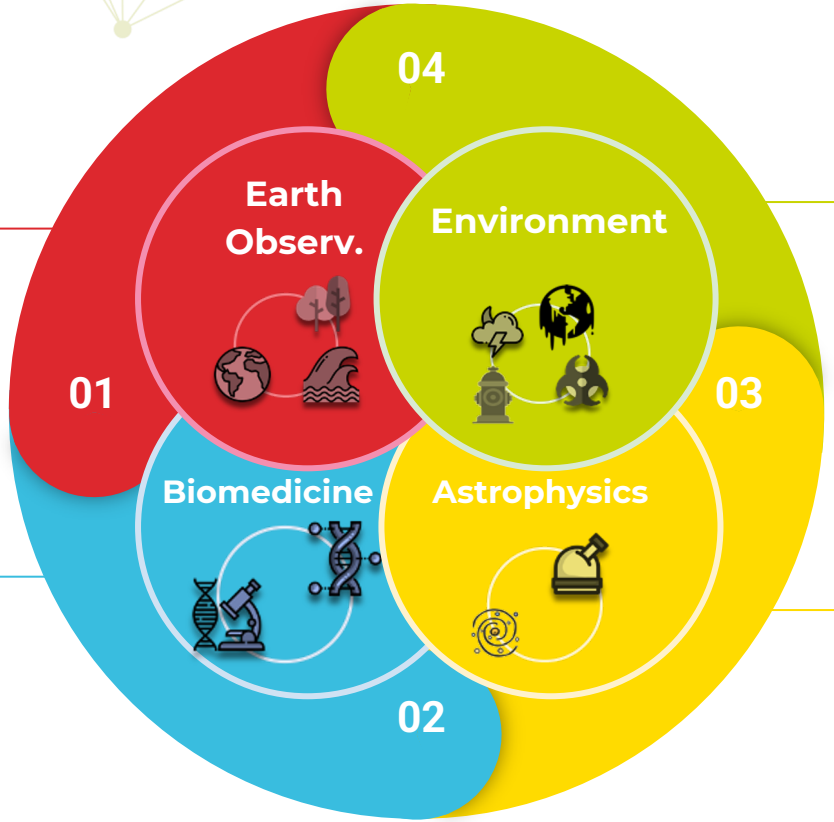
The Thematic Services Virtuous Cycle

Increase the capacity, performance, reliability and/or functionality

By means of best practices for adopting common EOSC core tools and services.

Increase service quality

FAIR data practices and software quality assessment.



Increase relevance of National Thematic Services

By expanding the use of the mature national services in an international scope.

Increase the number of users

By means of the integration in EOSC and the training.

Alignment with EOSC

- The adaptation, improvement and quality assessment of Thematic Services on a Federated Data Infrastructure strongly aligns with the objectives of EOSC and will develop best practices and experiences
  - The Thematic Services are part of the EOSC-Exchange data layer and rely on EOSC-Core services (*).
  - Fully focused to the the Operational Objective OO13 "*Continuously monitor and promote the increased uptake of core services and EOSC resources, access to EOSC Exchange tools and services and ensure a feedback loop with the users*" (**).
  - A key factor for the success of EOSC (*) is performance: how EOSC as an ecosystem operates and how the resources are used and acknowledged by the users.

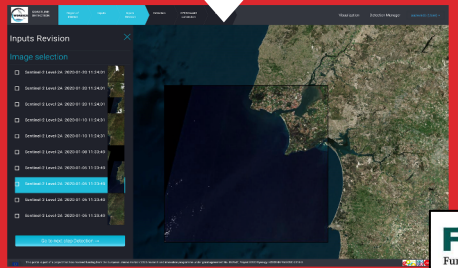
(*) Solutions for a Sustainable EOSC: An FAIR Lady report from the EOSC Sustainability Working Group, November 2020. https://bit.ly/3vomzSl
(**) The Strategic Research and Innovation Agenda (SRIA) (https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf)

0. Thematic Services in Earth Observation

WORSICA

Water Monitoring Sentinel Cloud Platform

A service for the detection of water using satellites, Unmanned Aerial Vehicles & in-situ data. WORSICA can be used for coastline detection, inland water bodies detection and water leaks detection on irrigation networks.



SAPS

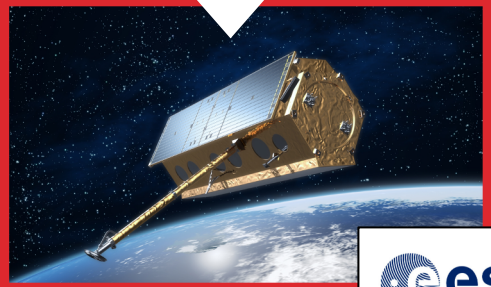
Surface Energy Balance Automated Processing Service

Used to estimate Evapotranspiration and other environmental data that can be applied, for example, on water management and the analysis of the evolution of forest masses and crops.

GCore

Acquisition, cataloguing and processing EOS data

G-Core is a production-ready technology used as a service at ESA's and national programs that provides a Data Manager for spatial and non-spatial purposes and a framework for third-party processors.

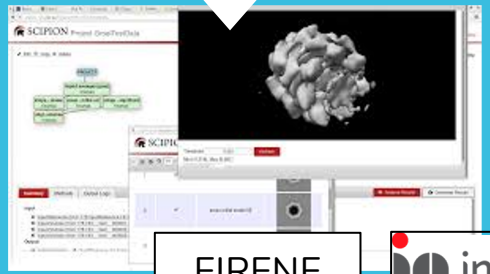


0. Thematic Services in Biomedicine & Astrophysics

SCIPION

CryoEM data processing for Structural Biology

ScipionCloud service will allow users from Instruct to deploy a dynamic cluster in the cloud to keep processing the data acquired at the facility.



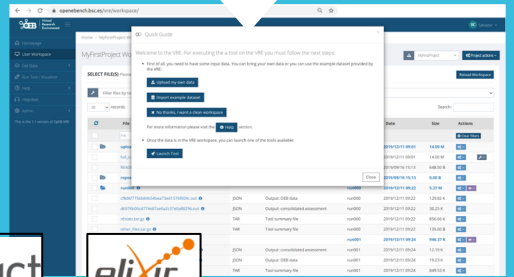
EIRENE

instruct ERIC

OpenEBench

ELIXIR benchmarking and technical monitoring platform

Used to evaluate bioinformatics tools, OpenEBench is an observatory for SW quality based on the automated monitoring of FAIR for research software metrics and indicators.



elixir

LAGO

Latin American Giant cosmic ray Observatory

LAGO is a cosmic ray observatory made of a network of water- Cherenkov detectors (WCD) spanning over different altitudes and latitudes making research on High Energy Physics, Space weather, etc.

LAGO

0. Thematic Services in Environment

UMSA

Untargeted Mass-Spectrometry Analysis

UMSA aims at processing data to correlating the whole spectra with other data to work with more complex hypotheses on the impact of environment in human health.

EIRENE

MSWSS

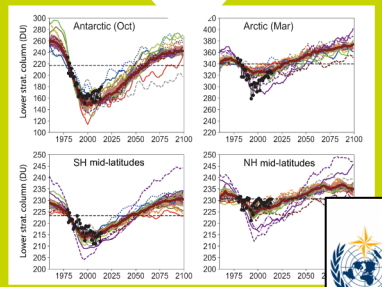
Water Supply Systems modeling and analysis

MSWSS integrates the analysis and simulation of toxics in drinking water supply networks to allow operators and researchers to analyse hazardous events.

O3AS

Ozone Analysis Service

The O3AS service provides a tool to extract O_3 time series from large climate prediction model data to produce figures of stratospheric O_3 trends to estimate when O_3 recovers.

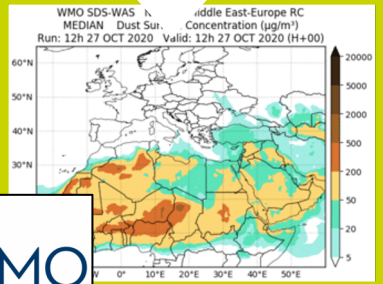


WMO

SDS-WAS

A Service related to the mineral dust forecast

SDS-WAS aims to support institutional entities to warn about possible dust events and to foster the study of dust-related phenomena.



1. Increase the capacity, performance, reliability and/or functionality: Improvements due to EOSC-Synergy

- Integration of **standardized AAI IdPs** to facilitate user management.
- Improvement of **processing backends** by replacing single computing instances with batch job queues, container management platforms or clients to high-throughput computing backends.
- **Publishing** the output results in persistent repositories.
- **PID annotation** of output data and integration in official harvesters.
- Improving **repeatability** and **platform-agnosticism** by describing the application topologies as code using standard TOSCA language.
- **Self-management** of resources to reduce maintenance costs.

Best Practices Elicitation - Gaps and Bottlenecks

DM	**Data Management**	• Ensure bandwidth for downloading data from operational providers • Need for Persistent Data Storage for resulting data with POSIX interface. • Data access control mechanisms. • Need of fata catalogues for data discovery.
WM	**Workload Management**	• Need for processing resources. • Need for containerised workloads.
DP	**Processing Resources**	• Need for Resource Management Services. • Non-standard configuration (RAM and GPU). • Dynamic infrastructures.
AAI	**Authentication and Authz Infrastructure**	• Use of federated Identity Providers. • Identity Delegation.

1. Increase the capacity, performance, reliability and/or functionality: Adoption of EOSC Services

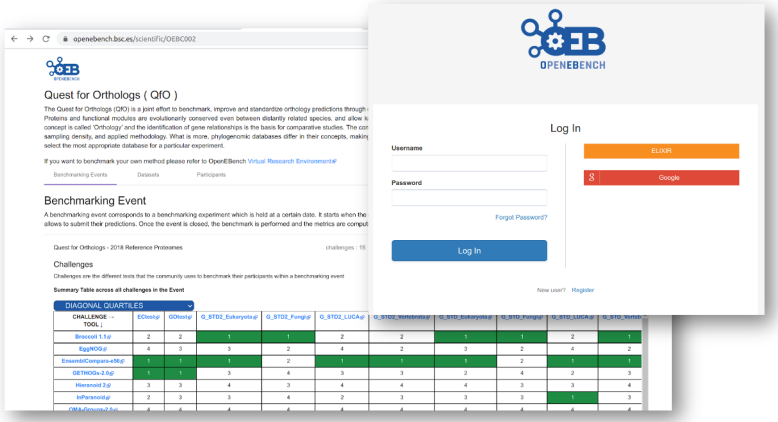
Service	WORSICA	G-Core	SAPS	Scipion	LAGO	SDS-WAS	UMSA	MSWSS	O3AS	OpenEBench
AAI	***EGI Check in***, WATTS	Kerberos LDAP & CAS User/pwd	***EGI Check in***	***EGI Check in***	***eduTEAMS+ EGI Check-in***	***B2ACCESS***	***EGI Check in*** & Life- science AAI	***EGI Check in***	***EGI Check in***	Life Sciences AAI
Workload Mng.	ArcCE, Batch (SLURM)	GCore+ K8s	K8s	Batch (SLURM)	Batch (SLURM)	Batch (SLURM)	Batch (SLURM) in ***IM/EC3*** (in Galaxy)	Batch (SLURM) in ***EC3*** (in Galaxy)	Cluster batch (SLURM) & K8s	GA4GH WES/TES stack + NextFlow
Resource Mng.	***IM (TOSCA)***	***IM / EC3***	***IM / EC3***	***IM / EC3***	***Local clusters & IM+EC3***	***EC3***	***IM / EC3***	***IM / EC3***	***IM***	one
Data Storage	Nextcloud, Dataverse	ElasticSearch for the catalogue	OpenStack Swift	Local + OneData	***EGI DataHub ONEDATA***	***B2HANDLE /B2SAFE***	Local + S3	Local + Dataverse	WebDAV	Local + ***B2SHARE***

Services in EOSC Marketplace

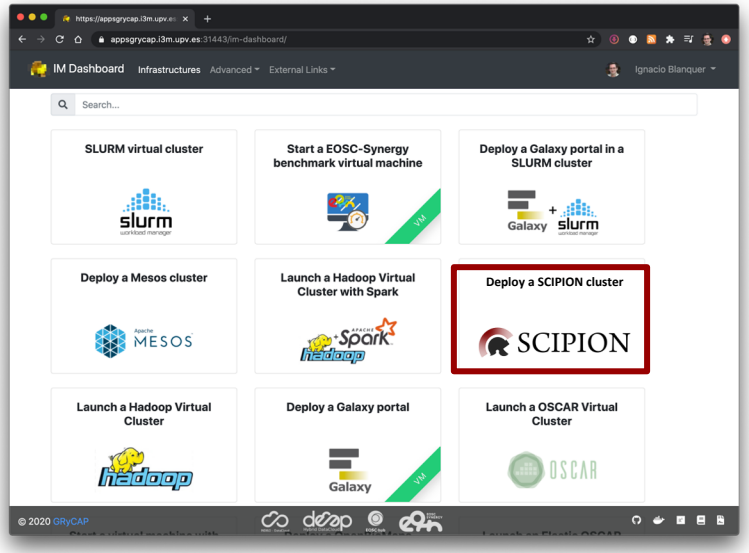
D4.2 - First prototype of the EOSC Thematic services - http://dx.doi.org/10.20350/digitalCSIC/12610

1. Thematic Services: Different Usage Modes

- A single entry portal that manages users, data access, processing and visualization
  - Supported by a shared or dedicated pool of resources.



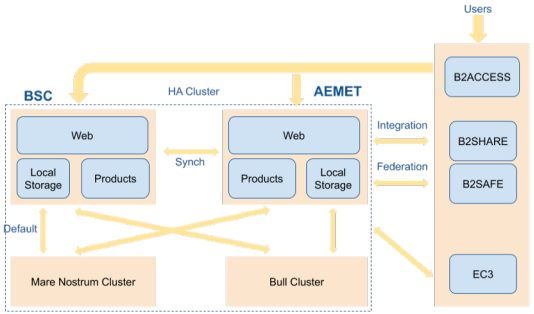
- Each user deploys his/her own instance
  - A combination of TOSCA recipes with Docker containers.

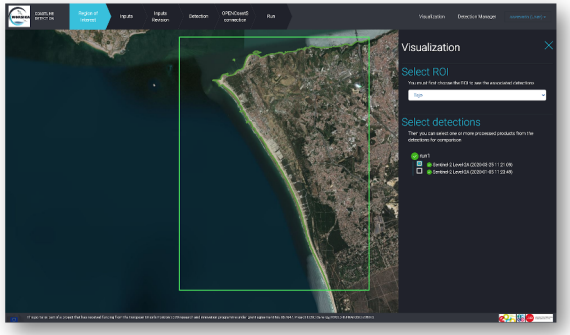


Computing Backends

- A Dynamic dedicated cloud backend
  - An elastic cluster that shrinks or grows according to the workload.
- An HPC batch queue
  - A dedicated slot in an HPC batch queue.
- High-throughput Computing Back-end
  - Using EGI HTC resources for massive Batch jobs.

```
$ clues status
node                                  state     enabled    time stable
-------------------------------------------------------------------
wn1.localdomain                        used     enabled      18h53'26"
wn2.localdomain                       powon     enabled      00h03'35"
wn3.localdomain                       powon     enabled      00h02'59"
wn4.localdomain                       powon     enabled      00h02'31"
wn5.localdomain                         off     enabled      00h01'20"
```

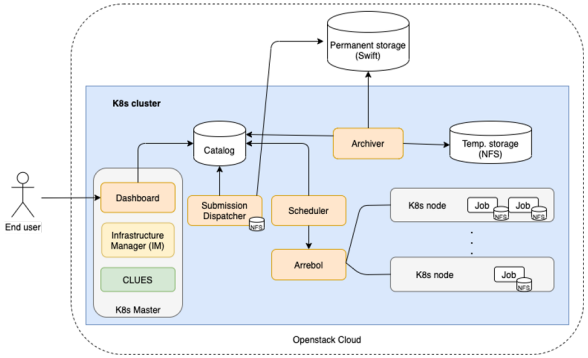




Different workloads

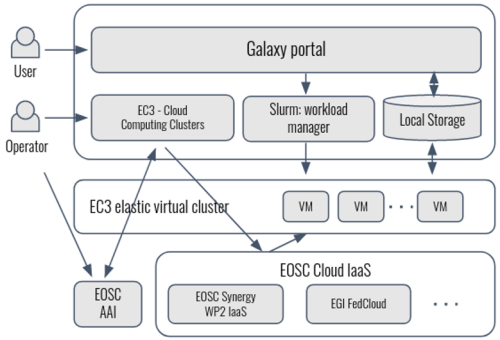
Containers on a K8s back-end

- K8s jobs submitted directly from the service.



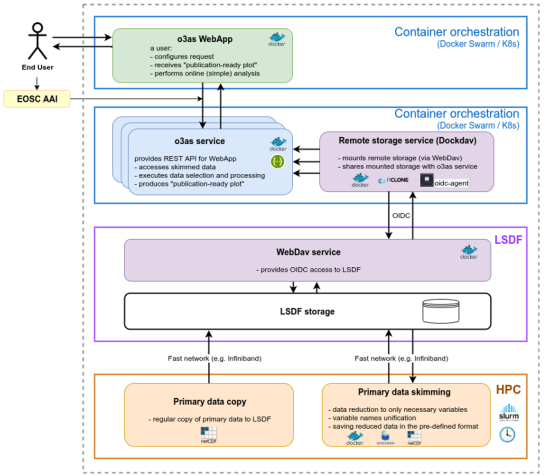
Batch jobs on SLURM and Galaxy

- Slurm jobs on a dedicated batch queue with shared filesystem.



Batch jobs on HPC

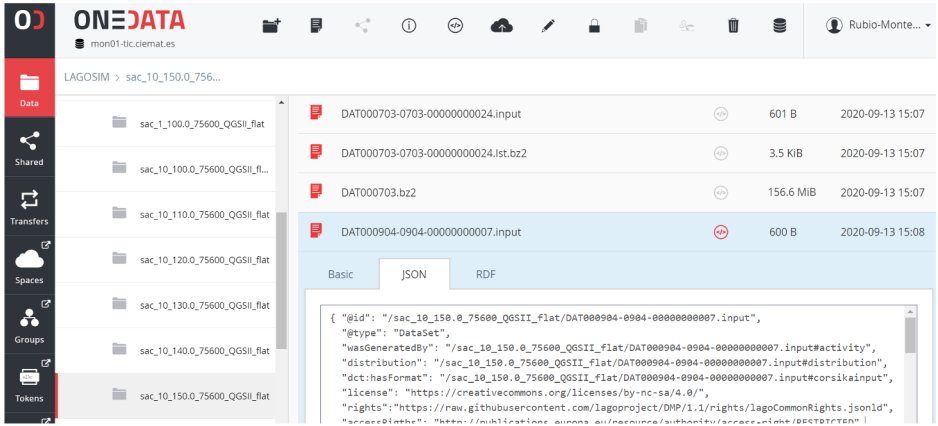
- Regular or containerised jobs.



Persistent storage

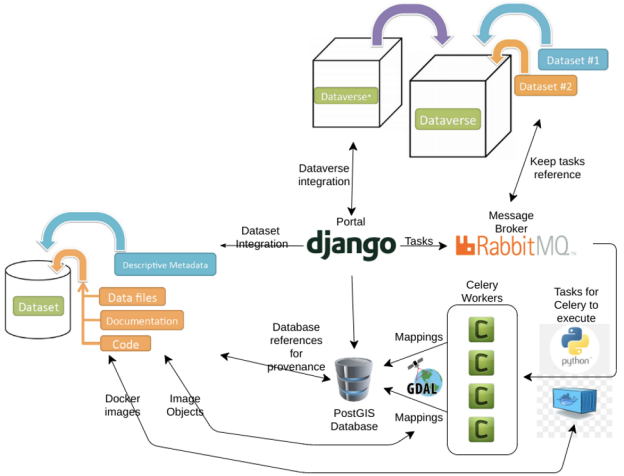
Integrating external Data Infrastructures

- Both EGIDataHub and EUDAT



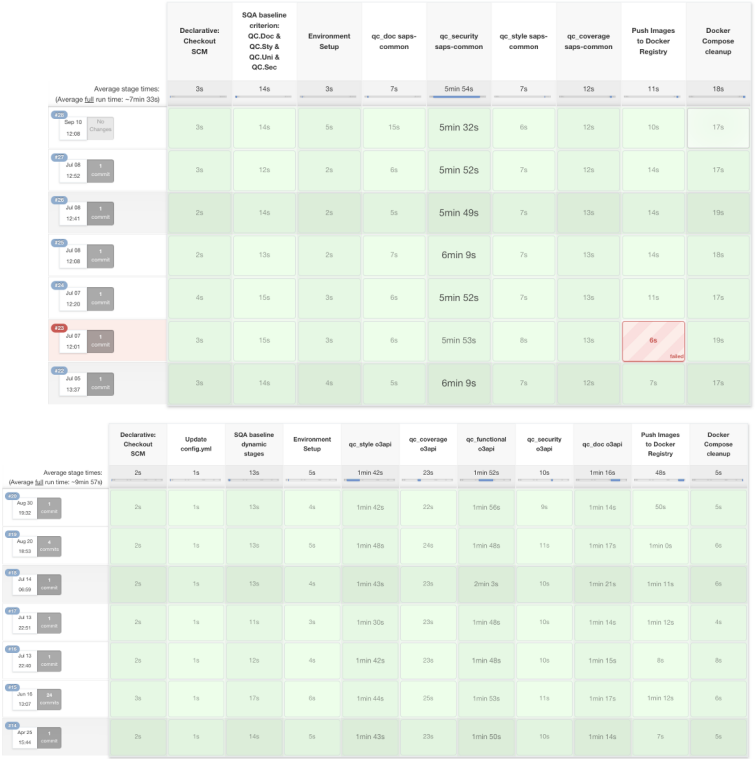
Deploying their own Datastore

- DATAVERSE instance



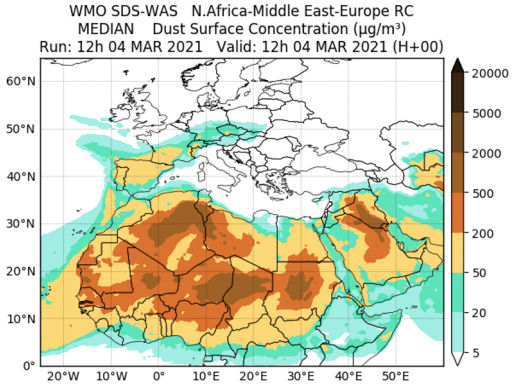
2. Increase service quality: Software Quality Assessment

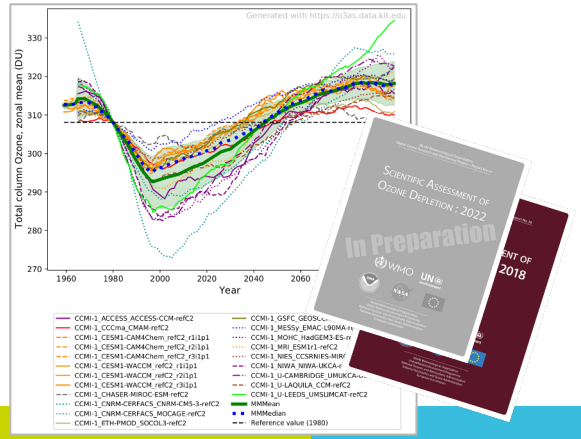
- TS Quality increase through two paths:
  - Evaluation of software quality of the components adopted.
  - Evaluation of the software quality of the adaptation performed in the TS in the frame of EOSC-SYNERGY.
  - Using monitoring and CI/CD pipelines.
    - Code style.
    - Unit test coverage.
    - Container building.
    - Service Deployment
      - *Virtual Infrastructure deployment & configuration + Deployment of K8s applications.*
    - Security checks.
  - Best practices created for Python and Java (https://u.i3m.upv.es/s71hd).



2. Increase service quality: FAIR Data evaluation

- Services consume and produce Data.
  - In some cases the output data is the major outcome.
- Data produced will be evaluated according to the FAIR principles.
  - Checking compliance with the individual DMPs (http://hdl.handle.net/10261/219309).
  - Automatic evaluation of compliance to FAIR principles (RDA / FAIRsFAIR).
- Evolving services to use FAIR repositories, either internal or external.



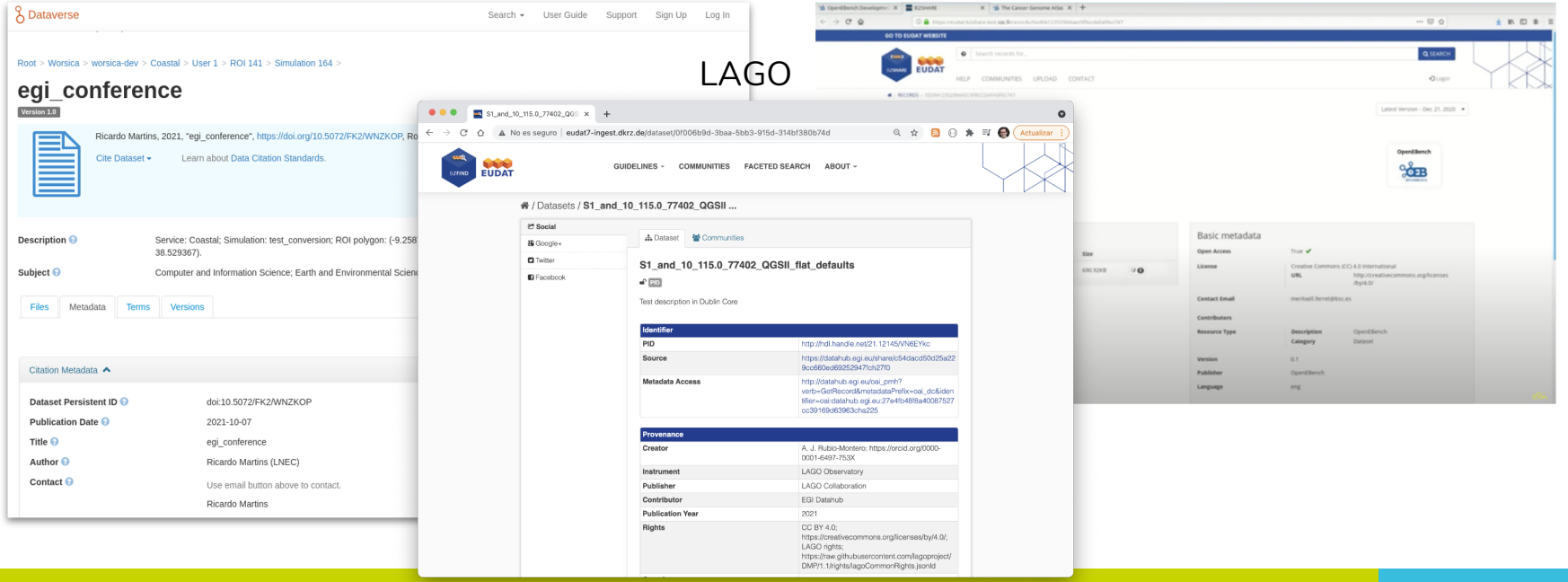


2. Increase service quality: FAIR Data evaluation

WORSICA

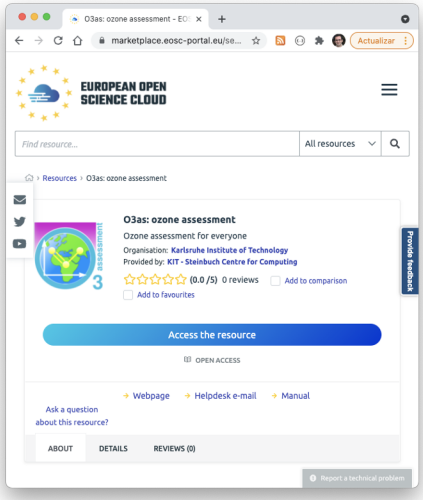
OpenEBench

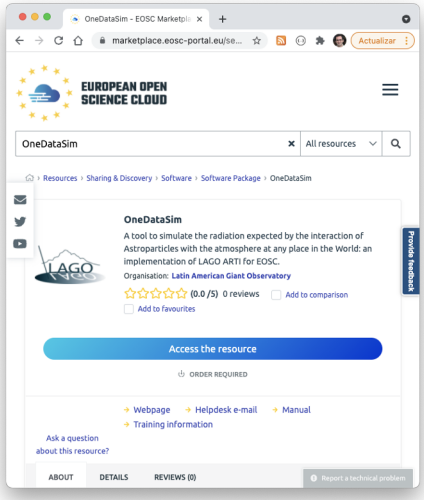
LAGO

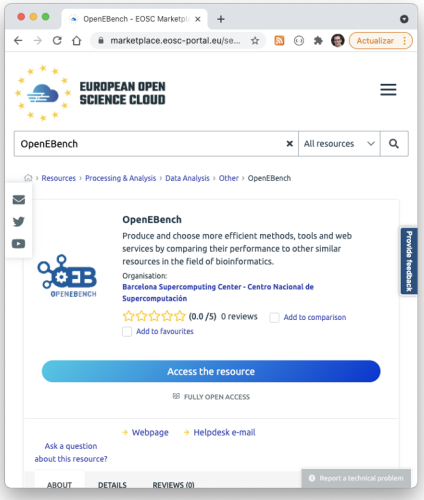


3. Increase the number of users: Integration in EOSC Marketplace

- Services will be registered in the EOSC Marketplace Portal.

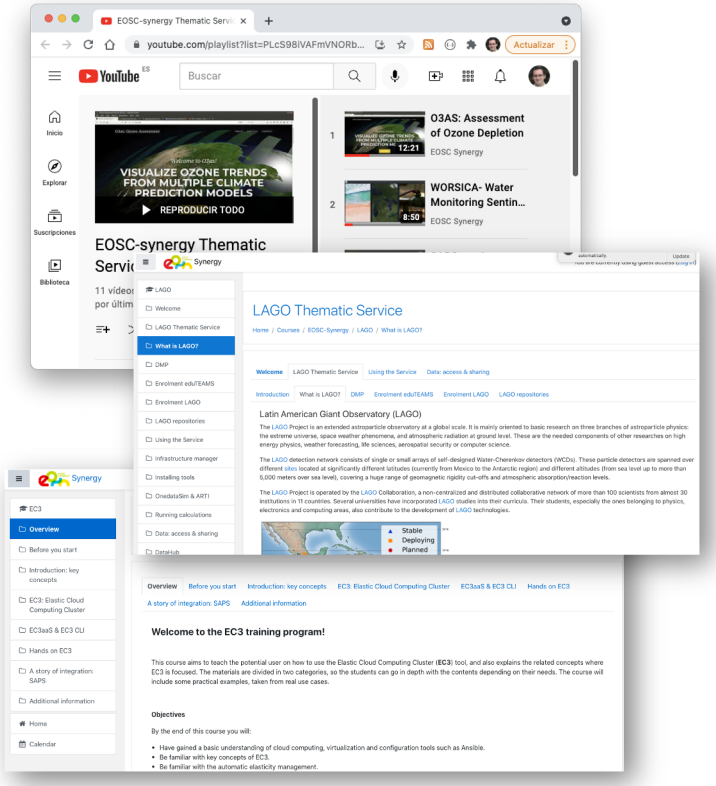






3. Increase the number of users: Training

- Demonstration videos available for explaining the improvements to the Thematic Services and the new usage models (https://bit.ly/3bXJZq2)
  - Integration of standardized AAI IdPs (UMSA 1:08)
  - Improvement of processing backends (SAPS 0:50).
  - Publishing the output results in persistent repositories. (SDS-WAS 4:04)
  - PID annotation of output data & integration in official harvesters (OpenEBench 7:20)
  - Improving repeatability and platform-agnosticism through standard TOSCA descriptions (SCIPION 3:15).
  - Self-management of resources to reduce maintenance costs (MSWSS 1:55).
- Training materials in development
  - https://learn.eosc-synergy.eu/courses/

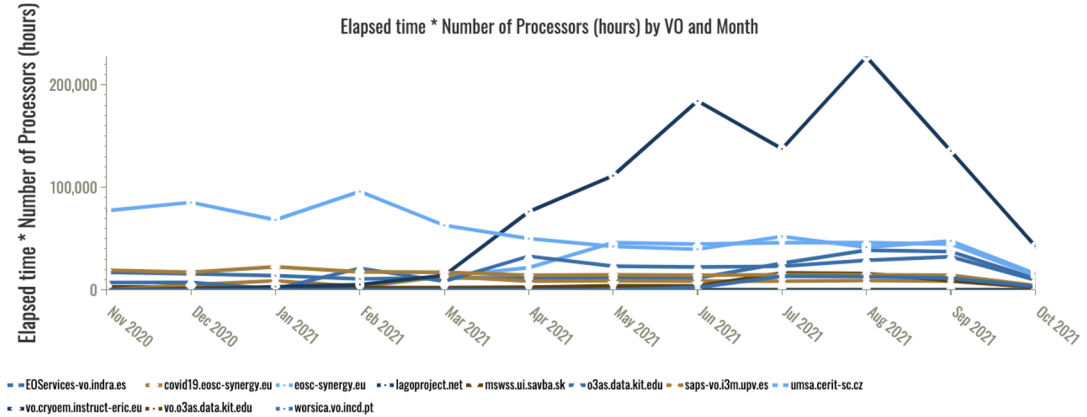


4. Increase relevance: Measuring success - Metrics and KPIs

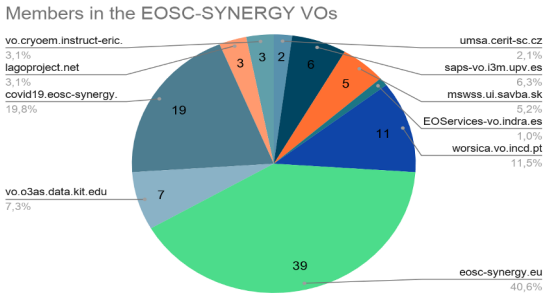
01	User Community	• Number of direct/indirect users in a given period/ accumulative • Number of centres/countries • Number of recurrent users	VOs operations-portal.egi.eu/vo/
02	Service Usage	• Number of service accesses in a given time / accumulative • CPU hours / RAM Size in a give time / accumulative • Max. capacity / capability experimented (vcpus & Memory). • Max. throughput in service accesses.	Accounting https://bit.ly/2LGpiFp & Deployment Services
03	Usability	• Performance, Scalability. • Learning curve, Error management, Robustness. • Completion, Interoperability, Convenience.	Questionnaires
04	Scientific Impact	• Number of publications acknowledging the service. • Number of communications (talks, panels, posters, etc.).. • Number of individual training hours on the service.	Publication archives
05	Cross Fertilization	• Number of code transfers. • Number of joint dissemination actions. • Number of synergies among thematic services not reflected above.	

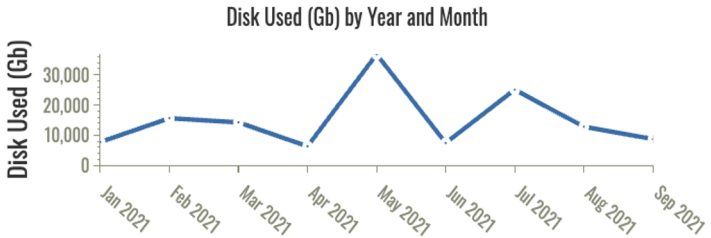
4. Increase relevance: Measuring success - Metrics and KPIs: Linkage to e-Infrastructures

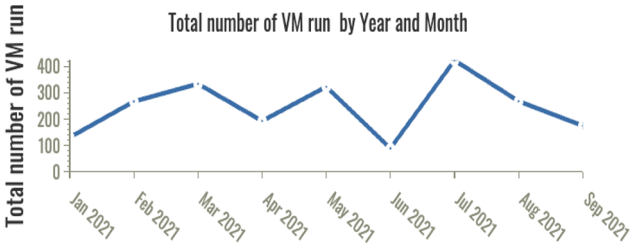
Near 3 Million CPU core hours in the last year.
https://bit.ly/2LGpiFp



90 registered users in VOs







Conclusions

- Identification of the needs and requirements of the Thematic Services (TS) for the increase the capacity, performance, reliability and/or functionality and Identification of key EOSC services that address those issues (http://dx.doi.org/10.20350/digitalCSIC/12609).
- Integration of those EOSC services into the TS to create Best Practices and deployment of the TS in EOSC resources (http://dx.doi.org/10.20350/digitalCSIC/12610, supported by 8 VOs).
- Integration of Quality evaluation for Software, Services
- Publishing services on the EOSC Marketplace Catalogue.
- Training and demonstration objects created.

Questions & Contact

Ignacio Blanquer Espert

iblanque@dsic.upv.es

Instituto de Instrumentación Para la Imagen Molecular

Universitat Politècnica de València

www.eosc-synergy.eu

@EOSC_synergy