EOSC-SYNERGY

EU DELIVERABLE D4.2

D4.2 - First prototype of the EOSC Thematic services

(DEMONSTRATION)

Document Identifier:	EOSC-SYNERGY-MS17
Date:	31/8/2020
Activity:	WP4
Lead Partner:	UPV
Document Status:	Final
Dissemination Level:	PUBLIC
Document Link:	

Abstract:

This document describes the thematic services that expose the applications and data to the scientific community. This is a report that supports the demonstration deliverable D4.2. The 10 thematic services of EOSC-SYNERGY are reasonably different in maturity, requirements, technological needs and approach, which guarantees that the expansion of the EOSC capacity addresses multiple dimensions and challenges. The document describes four of them more in detail, as planned in the DoA.

EOSC-SYNERGY receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 857647

I. Copyright Notice

II. Delivery Slip

	Name	Partner/Activity	Date
From	Ignacio Blanquer	UPV/WP4	
Reviewed by	**Moderator:** Isabel Campos **Reviewers:** Lara Lloret Marcin Plociennik	CSIC/WP1 CSIC external PSNC/WP6	
Approved by	PMB	PO	

III. Document Log

Issue	Date	Comment	Author/Partner
v0.1	16/7/2020	TOC and initial draft version	Ignacio Blanquer / UPV
v0.2	31/7/2020	SAPS thematic service	Amanda Calatrava / UPV
v0.3	13/8/2020	WORSICA thematic service	Alberto Azevedo / LNEC
v0.4	24/8/2020	WORSICA thematic service	Ricardo Martins / LNEC Samuel Bernardo / LIP
v0.5	24/8/2020	LAGO, SCIPION, O3AS thematic services	Antonio Juan Rubio / CIEMAT, Laura del Caño / CNB, Valentin Kozlov / KIT, Borja Esteban / KIT
v0.6	25/8/2020	G-CORE, UMSA, OpenEBench, MSWSS, SDS-WAS thematic services	Jan Astalos / IISAS, Salvador Capella / BSC, Ales Krenek / CESNET, Francesco Benincasa / BSC
v1.0	26/8/2020	Global homogenisation, Summary and conclusions	Ignacio Blanquer / UPV
v1.1	31/8/2020	Review comments	Marcin Plociennik / PSNC, Lara Lloret / IFCA-CSIC, Ignacio Blanquer / UPV

IV. List of Acronyms

Acronym	Description
AAI	Authentication and Authorisation Infrastructure
AEMET	Spanish State Meteorological Agency
AERONET	AErosol RObotic NETwork
B2FIND	EuDat Discovery service based on Metadata
B2SAFE	EuDat Service for distributing and storing large volumes of data
B2STAGE	EuDat service for data ingestion
CCMI	Chemistry-Climate Model Initiative
CEDA	Natural Environment Research Council's Data Repository for Atmospheric Science and Earth Observation
CF	Climate and Forecast
CORSIKA	COsmic Ray SImulations for KAscade
CS	Consortium Spatial Information
CSW	Catalog Service Web
DEM	Digital Elevation Model
DIRAC4EGI	Distributed Infrastructure with Remote Agent Control for the European Grid Initiative
DMP	Data Management Plans
DOI	Digital Object Identifier
DREAM	Dialogue on Reverse Engineering Assessment and Methods
DYNAFED	Dynamic Federations system
ebRIM	Registry Information Model
EC3	Elastic Compute Clusters in the Cloud
EGI	European Grid Initiative
EIRENE	European Environmental Exposure Assessment Network
ELIXIR	Life Sciences ESFRI
EMODNET	The European Marine Observation and Data Network
EMPIAR	Electron Microscopy Public Image Archive
EOSC	European Open Science Cloud
EPA	Environmental Protection Agency's

EPANET	Water distribution system modeling software package from the United States EPA
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum on Research and Innovation
EuDat	Collaborative Data Infrastructure for Data Preservation
FAIR	Findable, Accessible, Interoperable, Reusable
G-CORE	Earth observation data processing software from INDRA
GA4GH	Global Alliance for Genomics and Health
GEANT4	Toolkit for the simulation of the passage of particles through matter
GEE	Google Earth Engine
GPU	Graphics Processing Unit
HDF	Hierarchical Data Format
I2PC	Instruct Image Processing Center
IdP	Identity Providers
IGAC	International Global Atmopsheric Chemistry
IM	Infrastructure Manager
INGENIO	Spanish Earth Observation Satellite
INSTRUCT	Integrated Structural Biology Infrastructure
JSON	JavaScript Object Notation
LAGO	Latin American Giant Observatory
LANDSAT	Earth Resources Technology Satellite
LSDF	Large Scale Data Facility
LSDMA	Large-Scale Data Management and Analysis
MODIS	Moderate Resolution Imaging Spectroradiometer
MSWSS	Modelling Service for Water Supply System
NAMEE	Northern Africa, Middle East and Europe
NASA	National Aeronautics and Space Administration
NCEI	National Centers for Environment Information
netCDF	Network Common Data Form
netCDF	Network Common Data Form
NWP	Numerical Weather Prediction
O3AS	Ozone (O3) Assessment
OGC	Open Geospatial Consortium
OGC SOS	Open Geospatial Consortium Sensor Observation Service

OneData	Distributed Data Management solution from Cyfronet
OPENCoastS	Coastal circulation on-demand forecast
OpenEBench	Benchmarking service for Bioinformatics from ELIXIR
PAZ	Spanish Earth observation and reconnaissance satellite
POSIX	Portable Operating System Interface for X
QFO	Quest for Orthologs
RECETOX	Research Centre for Toxic Compounds in the Environment at Masaryk University
ROOT	Data Analysis Framework from CERN
SAPS	Serviço Automático de Processamento do SEBAL
Scipion	Cryo em image processing framework. Integration, traceability and analysis
SDS-WAS	Sand and Dust Storms Warning Advisory and Assessment System
SEBAL	Surface Energy Balance Algorithm for Land
SGE	Sun Grid Engine
SIC	Satellite Imaging Corporation
SMOS	Soil Moisture Ocean Salinity
SPARC	Stratosphere-troposphere Processes and their Role in Climate
TCGA	Cancer Genome Atlas
UAV	Unmanned Aerial Vehicles
UMSA	Untargeted Mass-spectrometry Analysis
WCD	water-Cherenkov detectors
WebDav	Web Distributed Authoring and Versioning
WMO	World Meteorological Organisation
WORSICA	Water mOnitoRing SentInel Cloud plAtform
ZBGIS	Basic Slovak database for GIS
Zenodo	OpenAIRE repository for Open Science

Table of Contents

Executive Summary

EOSC-SYNERGY aims at expanding the uptake of EOSC by building capacities. Thematic services constitute an important part of EOSC-SYNERGY and are the final layer that is exposed to final users. EOSC-SYNERGY has identified ten thematic services addressing four scientific areas (Earth Observation, Environment, Biomedicine and Astrophysics). Those thematic services gather and expose data and processing services directly to researchers in a convenient interface.

The thematic services are evolving in EOSC-SYNERGY by refactoring its architecture and integrating EOSC services from the EOSC marketplace. This will lead to higher performance and capacity as well as enhanced functionality.

The ten thematic services will be on operation by the end of 2020. Four thematic services (WORSICA, SAPS, OpenEBench, UMSA) have been selected to be prototyped and exposed earlier in the project lifetime to demonstrate the adoption of the selected EOSC services and to be used as best practices. Those thematic services have different and complementary requirements and cover the full spectrum of the key technical services selected (Authentication and Authorization via Check-in and Life Sciences AAI, cloud orchestration through Infrastructure Manager, Elastic batch queues and Kubernetes through EC3 and access to distributed storage through Dataverse and B2SHARE). Moreover, all the thematic services have already progressed and incorporate some additional services.

The new architecture of the thematic services incorporate seamless integration of the Authentication and Authorization for resources, processing and data as well as the embedded resource provisioning and elastic resizing of resources according to workload and efficient access to data storage. Further evolution will improve and consolidate such functionalities.

The ten thematic services have released a first version and have a clear plan for adopting the additional technical services needed by the end of the year, when they will release the first production version. Further improvements are already considered in their plan.

1. Introduction

This report belongs to WP4, “Capacity building for Thematic Services”. This activity aims at expanding the capacity and capabilities of ten thematic services identified in the project. These thematic services have been partially redesigned and adapted to leverage the functionality offered by services in the EOSC marketplace aligning their architecture to the other services in EOSC.

This document describes the status of the ten services, with higher detail on four of the cases that have been considered as more mature at this point in time of the project.

1.1. Scope of the document

This deliverable is of type “DEM” which refers to pilots, prototypes or demonstrators. This document is a short report that organizes and describes the pilot prototypes developed which constitute the actual deliverable. The document should be considered as a guideline to understand the scope of the services and to evaluate the adaptation performed. Demonstrations are planned for the EOSC-hub symposium, if the contribution requests are accepted.

1.2. Target Audience

This document serves the project partners as a summary of the actual progress of the thematic services, as well as an updated description of the architecture and EOSC services involved. This information is relevant for both WP2 (for the provision of services) and WP3 (for the identification of key services whose quality should be evaluated). Finally, this document will serve the evaluators of EOSC-SYNERGY to evaluate the progress of the action with respect to the metrics defined.

1.3. Structure of the document

In addition to this introduction, this document is structured in four main sections. First, section 2 describes in general the service adoption plan for the ten thematic services. Section 3 describes the four prototype cases that have been considered as “mature ones” according to the plan (verification means for milestone MS17 “Pioneer thematic services integrated in EOSC”, which states in the DoA that a reduced set (at least three) of significant thematic services are integrated into EOSC with reduced functionality. Section 4 describes the design and services integrated for the other 6 thematic services. Finally, section 5 draws up the conclusions.

2. Summary of the Thematic Services

The ten thematic services have complementary requirements and features. However, in general they share needs on four different categories:

- Authentication and Authorization Infrastructure (AAI). All cases require users to be properly authenticated and authorised. In some cases, there is a need for delegation from the users that access the platform for accessing data or processing resources. In those cases, it is mandatory to have a coherent single-sign on mechanism. Other cases may require an AAI linked to popular scientific IdPs and implement the authentication via Virtual Organization membership.
- Workload Management. Most of the cases deal with the execution of a set of batch jobs. In those cases, workload managers should be integrated. This will provide the capability to deal with a larger capacity. Options range from using a standard batch queue (SLURM) eventually powered up with automatic elasticity to using Kubernetes for the orchestration of containers.
- Resource Management. Most of the thematic services require deploying a virtual infrastructure where the services that provide the functionality and the processing will take place. In most cases, the use of Infrastructure Manager (IM) or Elastic Compute Clusters in the Cloud (EC3) could provide the capability of defining a virtual infrastructure as code and deploying it on the cloud.
- Data Storage. The services need to have a storage connected to the processing that can be efficiently accessed. In this case, there is a wide range of different solutions, ranging from EGI-DataHub and B2Share to local solutions based on Nextcloud, Datavers, Elasticsearch and WebDav.

Figure 1 shows the thematic services and the technology solutions.

Service	WORSICA	G-Core	SAPS	Scipion	LAGO	SDS-WAS	UMSA	MSWSS	O3AS	OpenEBench
AAI	EGI Check in	Kerberos LDAP & CAS User/pwd	EGI Check in	EGI Check in	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)	Cluster Batch & EC3 / K8s / DIRAC4EGI	Batch (SLURM)	Batch (SLURM) in IM/EC3 (in Galaxy)	Batch (SLURM) in EC3 (in Galaxy)	Cluster batch (SLURM) & Docker Swarm / K8s	GA4GH WES/TES stack
Resource Mng.	IM (TOSCA)	IM / EC3	IM / EC3	IM / EC3	Cluster batch & IM+EC3	EC3	IM / EC3	IM / EC3	IM / Cloudify	ONE
Data Storage	Nextcloud, Dataverse	ElasticSearch for the catalogue	OpenStack Swift	Local	EGI DataHub ONEDATA	B2SHARE /B2SAFE	Local + S3	Local	WebDAV	Local + B2SHARE

Figure 1: Services to manage each one of the four functionality blocks for each thematic service. Green denotes that they have been implemented at the moment of the deliverable, blue boxes denote components that will be integrated by MS18. Grey boxes are functionality that will be included later. Underlined services are services listed

in the EOSC marketplace.

3. Mature Thematic Services

3.1. WORSICA - Water Monitoring Sentinel Cloud Platform

3.1.1. Description

WORSICA is a service for the detection of water using satellites, Unmanned Aerial Vehicles and in-situ data. The main products of the service are: i) coastline detection, which includes coastal inundation areas due to storm-surge events; ii) inland water bodies detection, such as lakes, reservoirs or dams; and iii) water leaks detection on irrigation networks. This thematic service aims at integrating multiple-source remote sensing and in-situ data to determine the presence of water in coastal and inland areas. It is applicable to a range of purposes, from the determination of flooded areas (from rainfall, storms, hurricanes or tsunamis) to the detection of large water leaks in major water distribution networks. It builds on components developed in both national and European projects, integrated to provide a one-stop-shop service for remote sensing information, integrating data from both the Copernicus satellite and drone/unmanned aerial vehicles, validated by existing online in-situ data. The WORSICA service will be available without cost to all european public research groups. The private sector will be able to use the service, but some usage costs may be applied, dependending on the type of resources needed by each application/user.

The integration of the WORSICA service in the EOSC infrastructure will boost the usage of the service at an European level. This service will enable the research communities to generate maps of water presence and water delimitation lines in coastal and inland regions. These products can be useful for emergency and planning methodologies in case of inundations or reservoir leaks. In particular, the service promotes 1) the preservation of lives during an emergency, supporting emergency rescue operations of people in dangerously inundated areas, and 2) the efficient management of water resources targeting water saving in drought-prone areas.

3.1.2. Architecture

The architecture of WORSICA consists of three core components (Fig. 2): i) a frontend component; ii) an intermediate component; and iii) a processing component. The frontend component manages all the interaction of the service with the users through a web portal, such as the configuration of the simulations and requests for the service. The intermediate component is a task orchestrator that manages all the requests that arrive from the frontend and sends them to a processing component, and also deals with input/output storage tasks, such as download/upload of the satellite images and intermediate products and metadata to be sent to the Dataverse repository. The processing component (in purple) is a container with the requirements for image processing (scripts, inputs and software) and sent to be run by a resource manager on the cloud/grid infrastructure.

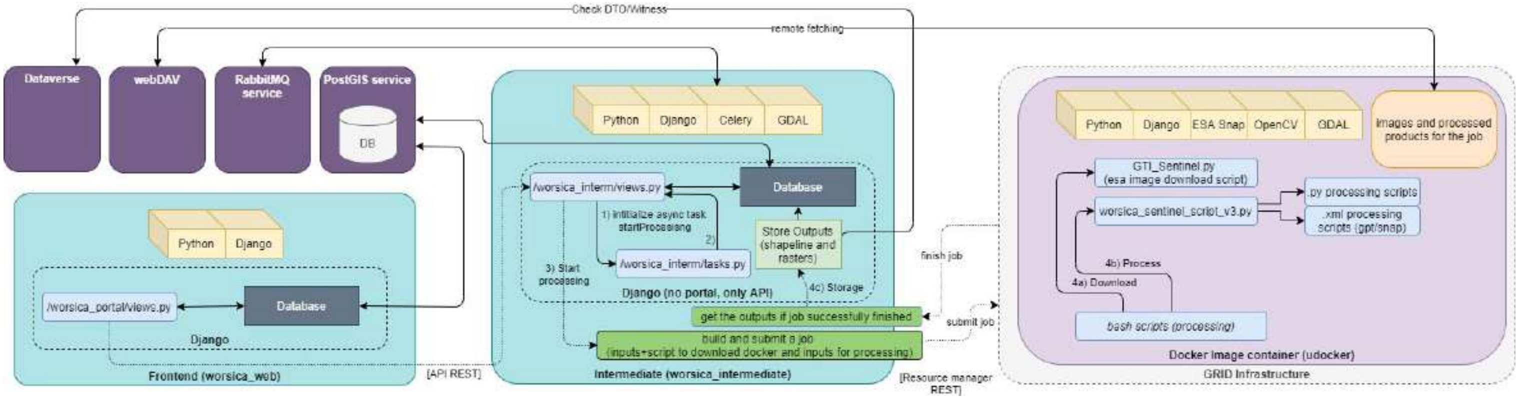

Check DTO/Witness
remote fetching
Dataverse
webDAV
RabbitMQ service
PostGIS service
DB
Python
Django
/worsica_portal/views.py
Database
Django
Frontend (worsica_web)
[API REST]
Python
Django
Celery
GDAL
/worsica_interm/views.py
Database
1) initialize async task startProcessing
2)
3) Start processing
/worsica_interm/tasks.py
Store Outputs (shapeline and rasters)
4c) Storage
Django (no portal, only API)
get the outputs if job successfully finished
build and submit a job (inputs+script to download docker and inputs for processing)
Intermediate (worsica_intermediate)
finish job
submit job
[Resource manager REST]
Python
Django
ESA Snap
OpenCV
GDAL
Images and processed products for the job
GTI_Sentinel.py (esa image download script)
worsica_sentinel_script_v3.py
.py processing scripts
.xml processing scripts (gpt/snap)
4b) Process
4a) Download
bash scripts (processing)
Docker Image container (udocker)
GRID Infrastructure

Figure 2 - Architecture of WORSICA service.

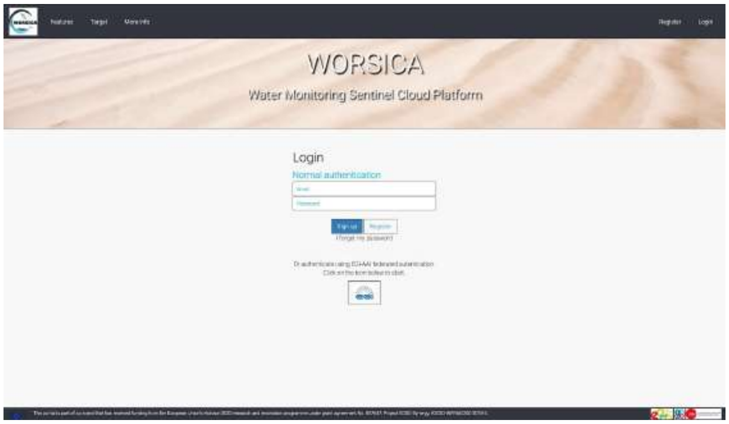
3.1.3. EOSC Services

In the EOSC Synergy, technical aspects of the WORSICA service will be improved considerably, using advanced technologies such as high-performance computing and cloud. The service will also be scaled up to a European level to reach all interested research communities. We will also adapt the service to other European Open Science Cloud (EOSC) services, such as EGI Check-In, and include these in our workflow, such as Dataverse. Therefore, several IT services, available in the EOSC marketplace, will be implemented in WORSICA service.

- **Authentication:** WORSICA uses **EGI Check-In** for the user authentication to the Frontend (portal), and this is a requirement in order to use the available EOSC services.
- **Workload Managers:** Processing jobs submissions are sent by the WORSICA Intermediate service to a GRID infrastructure by using **ArcCE with SLURM.** This allows efficient management of the available resources for HPC in order to speed up the processing jobs.
- **Data Manager: Nextcloud** is used to store processed job submission data input/outputs. **Dataverse** is used to register processed job submission metadata information for data FAIRsFAIR compliance.
- **Ansible and IM tools**: IM is used to deploy the infrastructure required for job processing, repositories and microservices. SLURM and Kubernetes clusters are deployed using TOSCA template over IaaS service and the remaining services will be installed from Docker images. Configurations for SLURM and Kubernetes are set up by ansible playbooks. This will be implemented in the milestone MS18 (All prototype services integrated).

3.1.4. Service Endpoint

The WORSICA web interface manages all the communication with the users. The portal can be accessed using the EGI federated authentication (figure 3) or simply with a verified email.



WORSICA

Water Monitoring Sentinel Cloud Platform

Login

Normal authentication

Email

Password

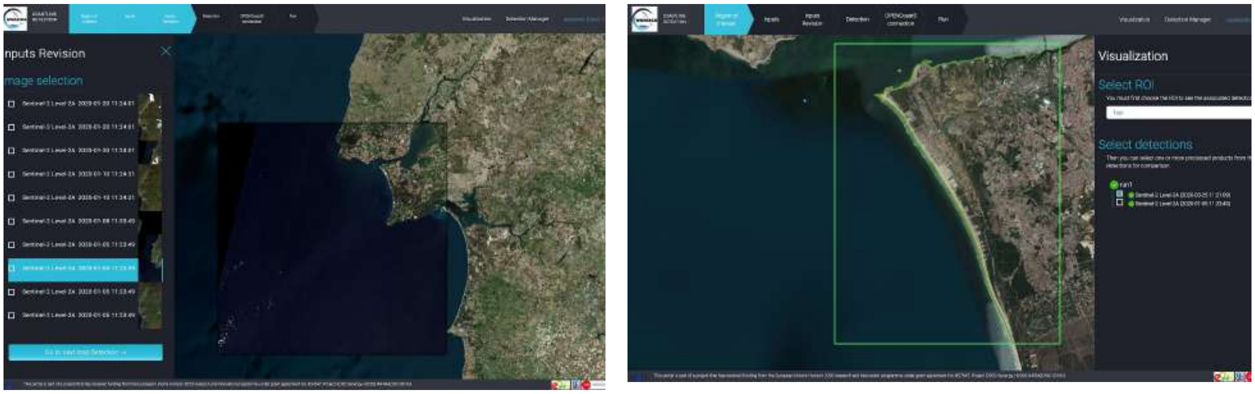
Sign up Register

I forgot my password

Or authenticate using EGI-AAI federated authentication
Click on the icon below to start.

Figure 3 - Login page for the WORSICA service.

In the WORSICA's portal the user can follow a configuration workflow (upper left row, figure 4-a). The main workflow consists in the following procedure: i) user selects the Region of Interest (ROI); ii) the user chooses the type of images to be processed (Sentinel-2, Pleiades or Drone) and the monitoring period; iii) The user verifies which images should be processed; iv) In the Detection menu, the user can select the water index, the number of classes for the clustering procedure; v) afterwards, the user can specify the informations for the connection to the OPENCoastS service, in order to retrieve the tidal elevation for the same period of the images; vi) in the last step, the user can confirm all the configuration values and submit the simulation. After the submission of the simulation, the user can check the results on the visualization menu (figure 4-b)



a)

b)

Figure 4 - Snapshots of the WORSICA portal. a) Selection of the images that will be processed; b) presentation of the results of the coastline detection product.

3.2. SAPS

3.2.1. Description

SAPS (SEB Automated Processing Service) is a service to estimate Evapotranspiration (ET) and other environmental data that can be applied, for example, on water management and the analysis of the evolution of forest masses and crops. SAPS allows the integration of Energy Balance algorithms (e.g. Surface Energy Balance Algorithm for Land (SEBAL) and Simplified Surface Energy Balance (SSEB)) to compute the estimations that are of special interest for researchers in Agriculture Engineering and Environment. These algorithms can be used to increase the knowledge on the impact of human and environmental actions on vegetation, leading to better forest management and analysis of risks.

3.2.2. Architecture

Figure 5 shows the architecture of SAPS. This architecture is automatically deployed, configured and managed by EC3. All the SAPS components run on a K8s cluster, so the location of each component depends on the K8s scheduler. The only component that needs to run in the front machine of the cluster is the Dashboard, so it can be exposed using the public IP of the front to the users.

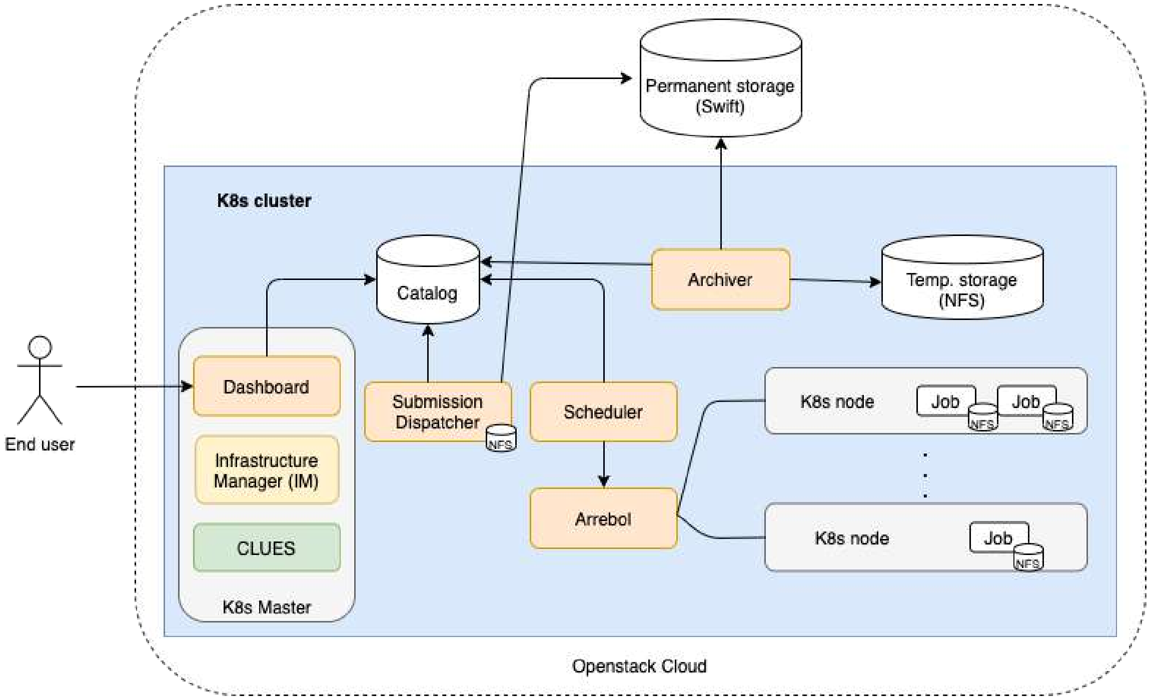


Figure 5 - Architecture of SAPS deployed on a K8s cluster by EC3.

As shown in figure 5, the user interacts with the system through the Dashboard, a web-based GUI that serves as a front-end to the Submission Dispatcher component. Through the Dashboard, the user, after successfully logging in, can specify the region, the period that he/she wants to process, as well as the particular Energy Balance algorithm that should be used. The execution consists of a three-stage workflow: input download, input preprocessing, and algorithm execution. With this data, the Dashboard creates the processing requests and submits them sequentially to the Submission Dispatcher. Each request generated corresponds to the processing of a single scene. The Submission Dispatcher creates a task associated with the request in the Service Catalog database (PostgreSQL). This element works as a communication channel between all SAPS components. Tasks have a state associated with them that is used to indicate which component should act next in the processing of the task.

The Scheduler component is in charge of orchestrating the created tasks through various states until they finish. It uses Arrebol to create and launch the tasks on the K8s cluster as Kubernetes Jobs. A Job downloads the appropriate Docker image from Docker Hub and starts its execution. Input and output files are stored on a Temporary Storage NFS that is accessible to all Jobs running at the cluster. Arrebol monitors all active Jobs to find out the status of the executions, and updates the state of each task in the Service Catalog, accordingly. The Archiver component collects the data and metadata generated by tasks whose processing has either successfully finished or failed. The associated data and metadata are copied from the NFS Temporary Storage, using an FTP service, to the Permanent Storage, which uses the Openstack Swift distributed storage system, where they are made securely and reliably available to the users.

Through the Dashboard, the user can also have access to the output generated by completed requests. The interface to access the output data uses a world map. A heat-map, segmented based on the standard tiles used by the Landsat family of satellites, is superimposed to the world map. The heat-map gives an idea of the number of scenes for each Landsat tile that have already been processed.

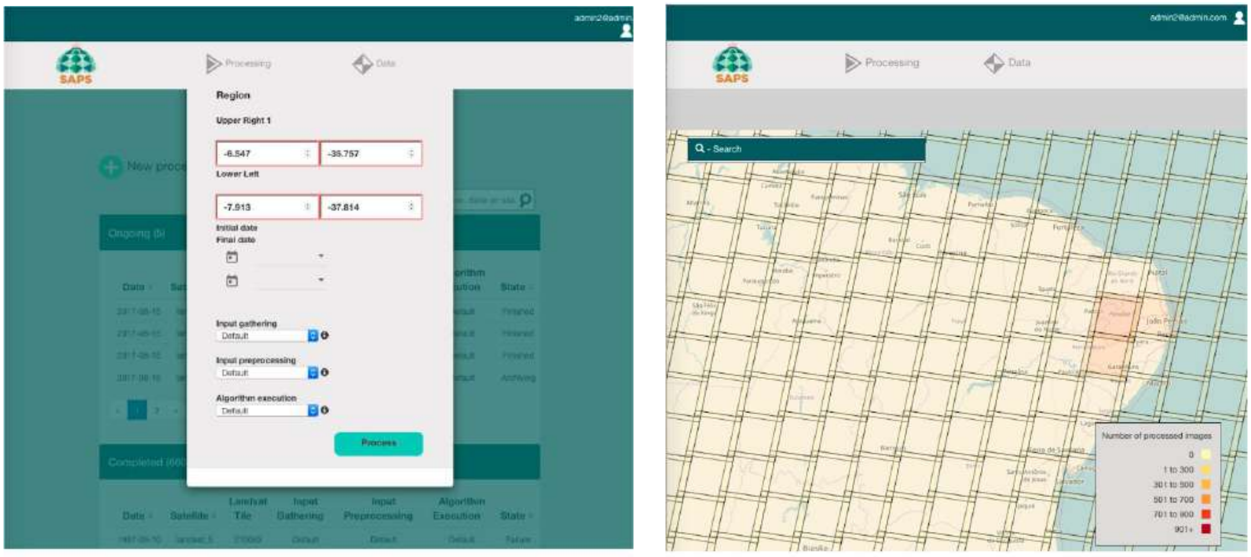
3.2.3. EOSC Services

In the context of EOSC-Synergy, SAPS is being integrated with several services offered by EOSC. This integration will facilitate European scientists to exploit the evapotranspiration estimation services from remote sensing imagery. Currently, the service relies on the next EOSC Services:

- **EC3 and IM** tools: both are services used by SAPS to deploy and configure a Kubernetes cluster automatically with SAPS running on it. Also, EC3 is used to manage the elasticity of the K8s cluster automatically. These tools facilitate the deployment and management of SAPS service.
- **EOSC computing resources**: through EC3 and IM, the SAPS service is deployed on top of a virtual elastic K8s cluster, that may rely on EOSC federated cloud computing resources or in on-premises solutions like Openstack.
- **EGI Check-in**: through EC3 portal. To deploy a cluster with SAPS, we use the EC3 portal of EOSC-Synergy, which is already integrated with EGI Check-in. So, to access a SAPS cluster, you should identify yourself with EGI Check-in. We will also consider integrating EGI Check-in directly on the SAPS dashboard in the next year of the project, for a fixed endpoint of SAPS.

3.2.4. Service Endpoint

The SAPS dashboard is designed to facilitate the deployment and management of Landsat analysis tasks. Figure 6 shows the appearance of it for (a) submission of a new processing request and (b) access to the output data.



Region

Upper Right 1

-6.547 | -35.757

Lower Left

-7.913 | -37.814

Initial date

Final date

Input gathering: Default

Input preprocessing: Default

Algorithm execution: Default

Process

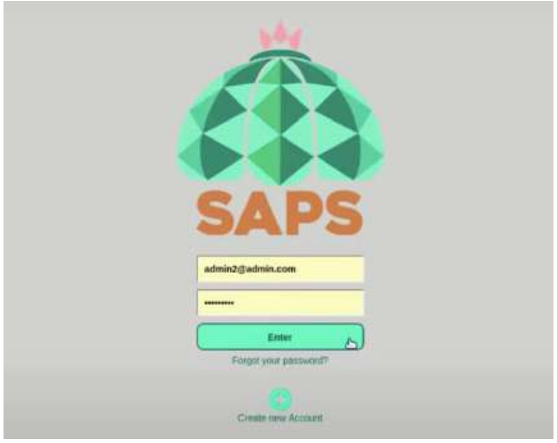
(a) Submission of new processing requests.

(b) Access to output data.

Figure 6 - Snapshot of the SAPS interface.

To access the SAPS Dashboard, a user is requested, as shown in figure 6. Internally, this is managed by local authorisation tokens. This solution is limited to the application, and we plan to study the viability of integrating EGI Check-in.

We do not provide an endpoint of the SAPS service. Instead, you can deploy your own instance easily through the EC3 EOSC-Synergy portal (https://servproject.i3m.upv.es/ec3-synergy/index.php), selecting as LRMS 'Kubernetes' and as Software package 'SAPS', or you can directly use the EC3 recipe and YAML files from the https://github.com/amcaar/saps-docker GitHub repository to deploy your own instance.



admin2@admin.com

••••••••••

Figure 7 - Login screen of SAPS Dashboard.

3.3. OpenEBench

3.3.1. Description

OpenEBench (https://openebench.bsc.es) is the ELIXIR benchmarking and technical monitoring platform for bioinformatics tools, web servers and workflows. The development of OpenEBench is led by the Barcelona Supercomputing Center (BSC) in collaboration with partners across different European projects and Life Sciences communities.

OpenEBench as platform has the overall objectives:

- Provide guidance and infrastructure support for community-led scientific benchmarking efforts.
- Provide an observatory for software quality based on the automated monitoring of FAIR for research software metrics and indicators.
- Work towards the sustainability of the platform by adopting, integrating and promoting principles on Open Software, Open Data and Open Science.
- Adopt community-led standards, protocols and/or including the Global Alliance for Genomics and Health (GA4GH), ELIXIR and the European Open Science Cloud (EOSC).

Building on those objectives, OpenEBench can engage with different end-user profiles across the Life Sciences communities and beyond.

- **Developers**, who have a reference place to identify current challenges and relevant data sets for developing new algorithms and/or measure the impact of new developments. OpenEBench offers the possibility to developers to compare the scientific performance of their solutions with others from the community. Ultimately, it helps to improve their methods and disseminate their results thanks to publications and results spreading.
- OpenEBench assists **Communities** in the organization of their scientific benchmarking activities and the identification of new trends in their concrete area by providing examples of assessment metrics already in use in other communities, contributing to results dissemination and establishing good practices.
- **Researchers** mainly benefit from getting guidance about choosing the best resource for their research needs and be aware of the latest advancements in the area by getting information from trusted experts and staying up to date with new developments.
- **Funders** are able to maximize impact from projects which include the development of new software resources and/or improve the existing ones.

3.3.2. Architecture

As described in figure 8, OpenEBench scientific benchmarking architecture has three different levels that allow communities at different maturity stages to make use of the platform.

- Level 1 is used for the long-term storage of benchmarking events and challenges aiming at reproducibility and provenance. Level 1 makes use of the OpenEBench data model (https://github.com/inab/benchmarking-data-model), which allows organizing any relevant data used and/or generated by community-led scientific benchmarking activities. Data is bundled and deposited in services provided by facilities like B2SHARE from EUDAT, where they receive a DOI. This enabes full data provenance and reproducibility for everyone involved.

- Level 2 allows the community to use benchmarking workflows to assess participants' performance. Those workflows compute one or more evaluation metrics given one or more reference datasets. Workflows for level 2 are organized using software container technologies (e.g. Docker or Singularity), and computational workflows managers like Nextflow. This choice facilitates the deployment and use of level 2 workflows across any computational installation compatible with such technologies.
- Level 3 goes further by getting workflows specifications from participants, and then evaluating them in terms of technical and scientific performance. At this level, the whole benchmarking experiment is performed at OpenEBench; first, the predictions are made using the software provided by the participants; then, those predictions are evaluated with the benchmarking workflows; and, finally, the results are stored and visualized in the web server.

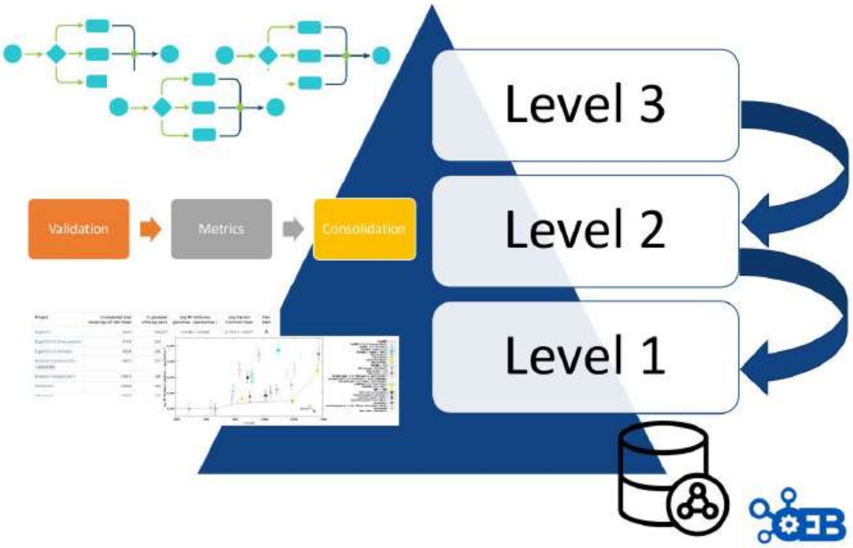


Figure 8. Conceptual diagram of support levels of OpenEBench.

Importantly, each level makes use of the architecture defined in the previous level e.g. participants' data generated by workflows at Level 3 are evaluated using the metrics and reference datasets in Level 2, and the resulting data is stored following the data model in Level 1 for private and/or public consumption.

3.3.3. EOSC Services

OpenEBench already uses ELIXIR AAI, which is intended to evolve together with other services e.g. GEANT; as Life Sciences AAI in the context of the cluster project EOSC Life. OpenEBench has started to incorporate some of EOSC Life services, specifically, WorkflowHub, as a mechanism to facilitate the provenance of the workflows used in the platform as well as a mechanism to monitor the availability and deployability of workflows used in OpenEBench within the community-led scientific benchmarking activities.

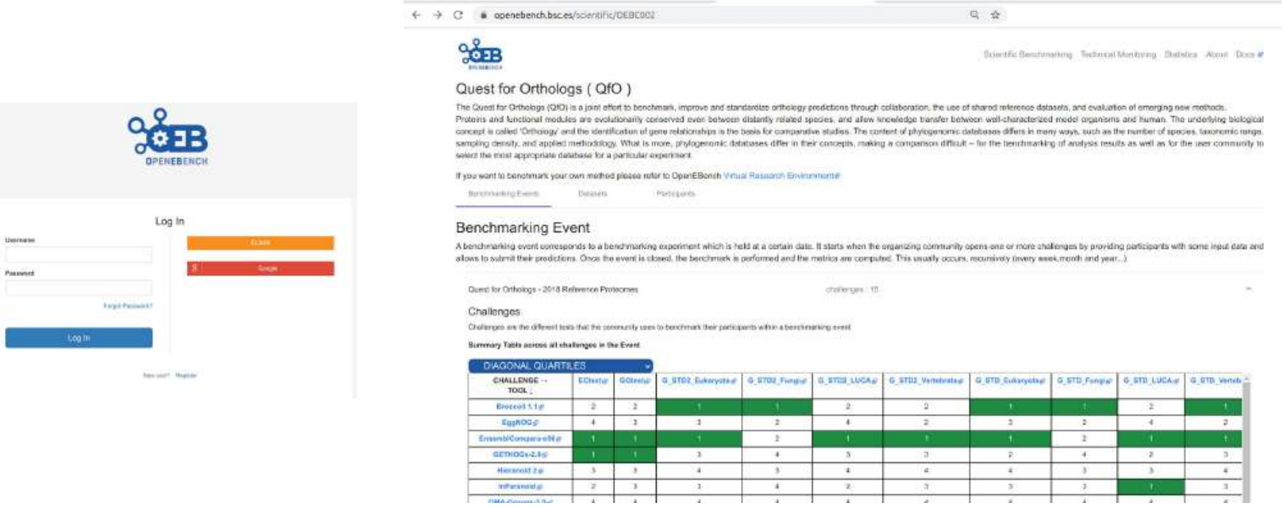
OpenEBench is integrating specific services from the EOSC Portal. Specifically, OpenEBench is integrating EUDAT services for the long-term availability of benchmarking results. To make this possible, EUDAT has created an OpenEBench Community, which will be used to associate any datasets from community-led scientific benchmarking activities. Using EUDAT allows us to assign a unique identifier, e.g. DOI, for those datasets contributed by members of communities at OpenEBench when publishing their results. In this way, it will be possible to reproduce at any time specific published benchmarking results by anyone interested. This integration requires the advanced management of users authorization as data should be deposited on behalf of their original owner rather than using OpenEBench identities.

It is expected that OpenEBench will become part of the EOSC portal portfolio by exposing and deploying the benchmarked analytical workflows as well as extending its capacity through best practices and additional services. As impact, we expect Life Science researchers will have semantically annotated, up-to-date collections of analytical workflows, which can be deployed across heterogeneous systems, organized by scientific communities around specific topics. As it is already happening, OpenEBench is contributing to organize emergent communities around scientific benchmarking activities by providing best practices and success stories of other communities.

3.3.4. Service Endpoint

This section includes some snapshots of the interface including the access procedure. Figure 9 (right) shows the result of a benchmark comparison, figure 9 (left) shows the access through Life Sciences AAI and figure 10 shows the user's workspace.

The service is available at https://openebench.bsc.es/. The Virtual Research Environment (VRE) is accessible at https://openebench.bsc.es/vre/tools/QFO_6/input.php?op=0. In this VRE, the users can upload their own data and applications for the execution of the benchmarks. In the general thematic service, any user can browse the information related to the benchmarks registered in the platform.



Quest for Orthologs (QfO)

The Quest for Orthologs (QfO) is a joint effort to benchmark, improve and standardize orthology predictions through collaboration, the use of shared reference datasets, and evaluation of emerging new methods. Proteins and functional modules are evolutionarily conserved even between distantly related species, and allow knowledge transfer between well-characterized model organisms and human. The underlying biological concept is called 'Orthology' and the identification of gene relationships is the basis for comparative studies. The content of phylogenomic databases differs in many ways, such as the number of species, taxonomic range, sampling density, and applied methodology. What is more, phylogenomic databases differ in their concepts, making a comparison difficult – for the benchmarking of analysis results as well as for the user community to select the most appropriate database for a particular experiment.

If you want to benchmark your own method please refer to OpenEBench Virtual Research Environment

Benchmarking Event

A benchmarking event corresponds to a benchmarking experiment which is held at a certain date. It starts when the organizing community opens one or more challenges by providing participants with some input data and allows to submit their predictions. Once the event is closed, the benchmark is performed and the metrics are computed. This usually occurs, recursively (every week,month and year...)

Quest for Orthologs - 2018 Reference Proteomes

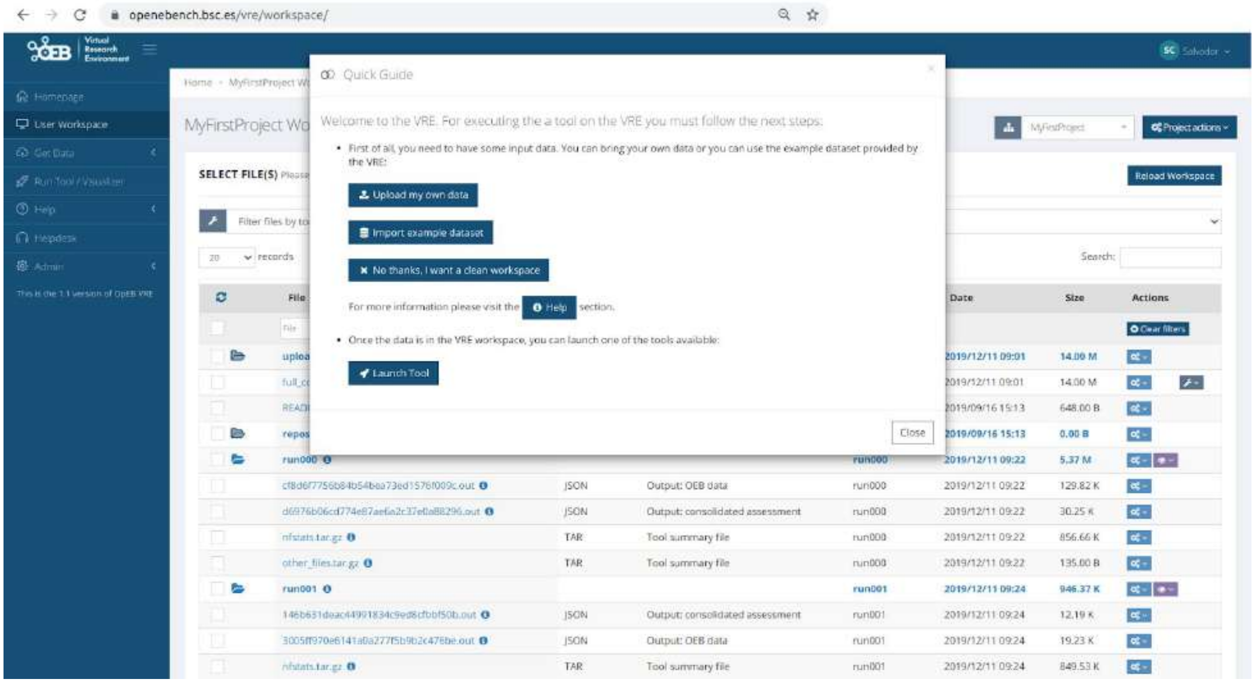
Challenges

Challenges are the different tests that the community uses to benchmark their participants within a benchmarking event

Summary Table across all challenges in the Event

CHALLENGE → TOOL ↓	ECtest	GOtest	G_STD2_Eukaryota	G_STD2_Fungi	G_STD2_LUCA	G_STD2_Vertebrata	G_STD_Eukaryota	G_STD_Fungi	G_STD_LUCA	G_STD_Verteb
Broccoli 1.1	2	2	1	1	2	2	1	1	2	1
EggNOG	4	3	3	2	4	2	3	2	4	2
Ensembl/Compara e94	1	1	1	2	1	1	1	2	1	1
GETHOGs-2.0	1	1	3	4	3	3	2	4	2	3
Hieranoid 2	3	3	4	3	4	4	4	3	3	4
InParanoid	2	3	3	4	2	3	3	3	1	3
OMA Groups 2.3	4	4	4	4	4	4	4	4	4	4

Figure 9. Screenshots of OpenEBench with the Sciences AAI access (left) and with the comparison of different methods for a specific benchmark using OpenEBench (right).



openebench.bsc.es/vre/workspace/

This is the 1.1 version of OpEB VRE

Home › MyFirstProject W

MyFirstProject Wo

SELECT FILE(S) Please

Filter files by to

20 records

Search:

Quick Guide

Welcome to the VRE. For executing the a tool on the VRE you must follow the next steps:

- First of all, you need to have some input data. You can bring your own data or you can use the example dataset provided by the VRE:

For more information please visit the Help section.

- Once the data is in the VRE workspace, you can launch one of the tools available:

File				Date	Size	Actions
uploa				2019/12/11 09:01	14.00 M	
full_c				2019/12/11 09:01	14.00 M	
READ				2019/09/16 15:13	648.00 B	
repos				2019/09/16 15:13	0.00 B	
run000			run000	2019/12/11 09:22	5.37 M	
cf8d6f7756b84b54bea73ed1576f009c.out	JSON	Output: OEB data	run000	2019/12/11 09:22	129.82 K	
d6976b06cd774e87ae6a2c37e0a88296.out	JSON	Output: consolidated assessment	run000	2019/12/11 09:22	30.25 K	
nfstats.tar.gz	TAR	Tool summary file	run000	2019/12/11 09:22	856.66 K	
other_files.tar.gz	TAR	Tool summary file	run000	2019/12/11 09:22	135.00 B	
run001			run001	2019/12/11 09:24	946.37 K	
146b631deac44991834c9ed8cfbbf50b.out	JSON	Output: consolidated assessment	run001	2019/12/11 09:24	12.19 K	
3005ff970e6141a0a277f5b9b2c476be.out	JSON	Output: OEB data	run001	2019/12/11 09:24	19.23 K	
nfstats.tar.gz	TAR	Tool summary file	run001	2019/12/11 09:24	849.53 K	

Figure 10. Screenshot of the workspace in OpenEBench.

3.4. UMSA

3.4.1. Description

UMSA is an untargeted mass-spectrometry analysis service from RECETOX (Research Centre for Toxic Compounds in the Environment at Masaryk University) in the Czech Republic. The service is evolving to a key component of the emerging EIRENE ESFRI. By means of the integration in EOSC, uniform access to data and computing resources are provided, scaling the service to the target European-wide user community. Typically, mass spectrometry is done in a targeted way to confirm or disprove the presence of a specific compound in a sample. On the contrary, we aim at processing data to correlating the whole spectra (ie. all the present compounds) with other data (social, medical, other sample analyses, etc.) to work with more complex hypotheses of environmental impacts on human health.

The data are unrecoverable, original samples cannot be re-acquired, therefore long-term data storage (even decades) is required, together with appropriate data curation. Tracking provenance of the secondary (derived) datasets (what was the exact process of generating them from the original source data), is fairly critical, as the results may differ dramatically with different settings.

The current release provides a Galaxy workflow based on re-factored tools originating from Emory university (apLCMS and xMSAnnotator), which detect peaks in the input spectra and matches those to metabolite and pathway databases. Extending the workflow with auto-tuning peak picking parameters (based on the original xMSAnalyzer tool) and false-positive filtering by retention time prediction is in progress.

3.4.2. Architecture

The service is deployed as a virtual cluster using an Infrastructure Manager RADL recipe. The cluster consists of a single head node, running Galaxy frontend and the Slurm server, and an arbitrary number of Slurm worker nodes. Data is shared among the nodes over NFS.

The deployment of the head node registers a configurable dynamic DNS name (umsa.dyn.cerit-sc.cz currently) to point to its assigned IP address. The well known service endpoint (umsa.cerit-sc.cz) is expected to be an alias pointing to the dynamic one. In the next step a *Let's encrypt* certificate is acquired to allow smooth https connection.

The RADL recipes also require simple provider-specific customization (cloud network names and base image identifiers in particular).

Authentication of the end users is managed by ELIXIR AAI via its corresponding Galaxy module. Configuration of Elixir AAI is the only manual step in the service deployment; the policy of introducing a new service to Elixir AAI requires human approval and exchange of secrets which cannot be automated so far. In the current version, the service can be accessed *bona-fide* -- all users who pass ELIXIR AAI authentication are allowed. However, we plan a trivial registration procedure (using the ELIXIR tools) to restrict the access.

The payload of the service are several tools in Galaxy. In order to keep strict control on the complex software dependencies, we wrap all the software to Docker containers; Galaxy and Slurm are configured to execute them in this way only. The tools themselves are installed from standard Galaxy toolshed to follow common procedures.

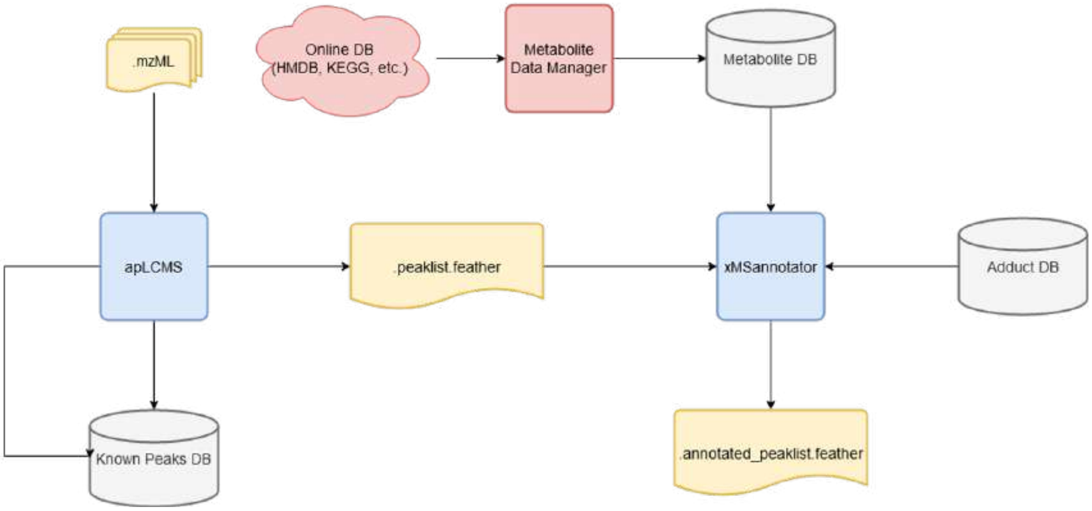


Figure 11. UMSA Data processing Workflow.

The diagram in figure 11 shows the essential mass-spec data processing workflow. Besides the user input (.mzML spectra files) the workflow needs to query a metabolite database. In order to achieve reproducibility, timestamped snapshots of the online databases are used. They are downloaded and processed only from time to time, and the snapshots are available to all service users. Most of the intermediate files passed between the tools are some kind of tabular data, using the Apache feather format consistently.

3.4.3. EOSC Services

UMSA leverages the following EOSC services

- Infrastructure Manager (IM) and Elastic Compute Clusters in the Cloud (EC3): except from minor provider-specific customization the service is deployed with a generic RADL recipe submitted to IM, and IM takes care of its lifecycle.
- EOSC computing resources: the deployment relies on cloud computing resources; due to non-trivial computational demand of the workflow, "HPC" flavors of cloud virtual machines (with no CPU overprovisioning) are prefered. The production service runs at the CESNET/Masaryk University cloud site, we managed to deploy at CESGA successfully as well.
- EGI CheckIn: used to authenticate to IM as well as to deploy the VMs to the cloud sites. The service is currently using the "catch all" eosc-synergy.eu[1] virtual organization.

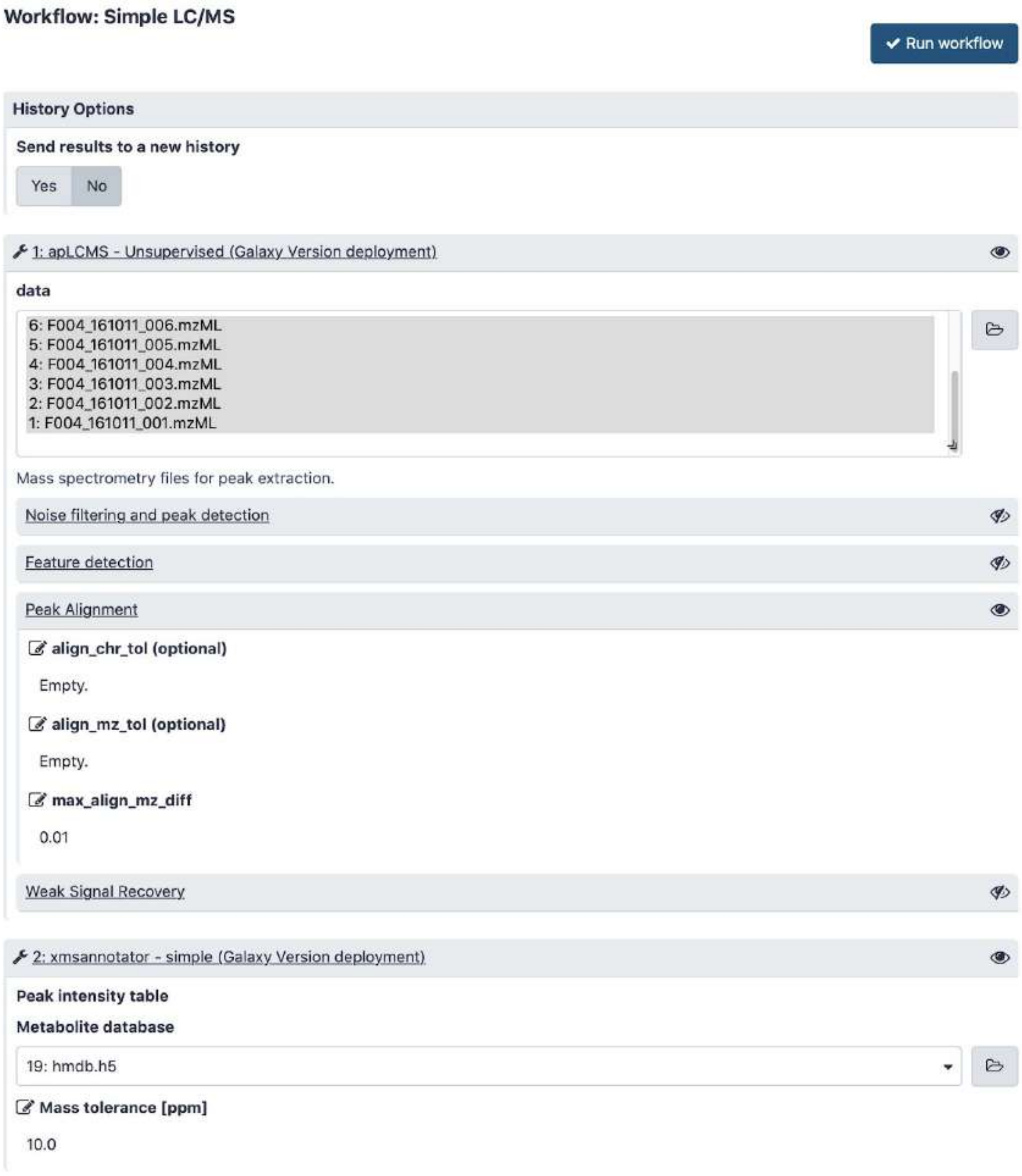
3.4.4. Service Endpoint

The principal service endpoint is https://umsa.cerit-sc.cz/. The user interface is standard Galaxy with minimalistic visual branding. Login with Elixir AAI credential is required as described above.

The individual tools are available in PeakPicking and Annotation sections, both exist in simple and advanced forms to address the needs of different user experience levels. The in-line documentation provides extensive description of the meaning of numerous input parameters, as well as appropriate references to web pages with further documentation of the tools, and the essential journal papers.

[1] http://operations-portal.egi.eu/vo/view/voname/eosc-synergy.eu

Figure 12 shows a typical input form of a simple workflow connecting the tools together. Figure 13 shows execution of this workflow in progress, generating its output files in Galaxy history.



Workflow: Simple LC/MS

History Options

Send results to a new history

1: apLCMS - Unsupervised (Galaxy Version deployment)

data

6: F004_161011_006.mzML
5: F004_161011_005.mzML
4: F004_161011_004.mzML
3: F004_161011_003.mzML
2: F004_161011_002.mzML
1: F004_161011_001.mzML

Mass spectrometry files for peak extraction.

Noise filtering and peak detection

Feature detection

Peak Alignment

align_chr_tol (optional)

Empty.

align_mz_tol (optional)

Empty.

max_align_mz_diff

0.01

Weak Signal Recovery

2: xmsannotator - simple (Galaxy Version deployment)

Peak intensity table

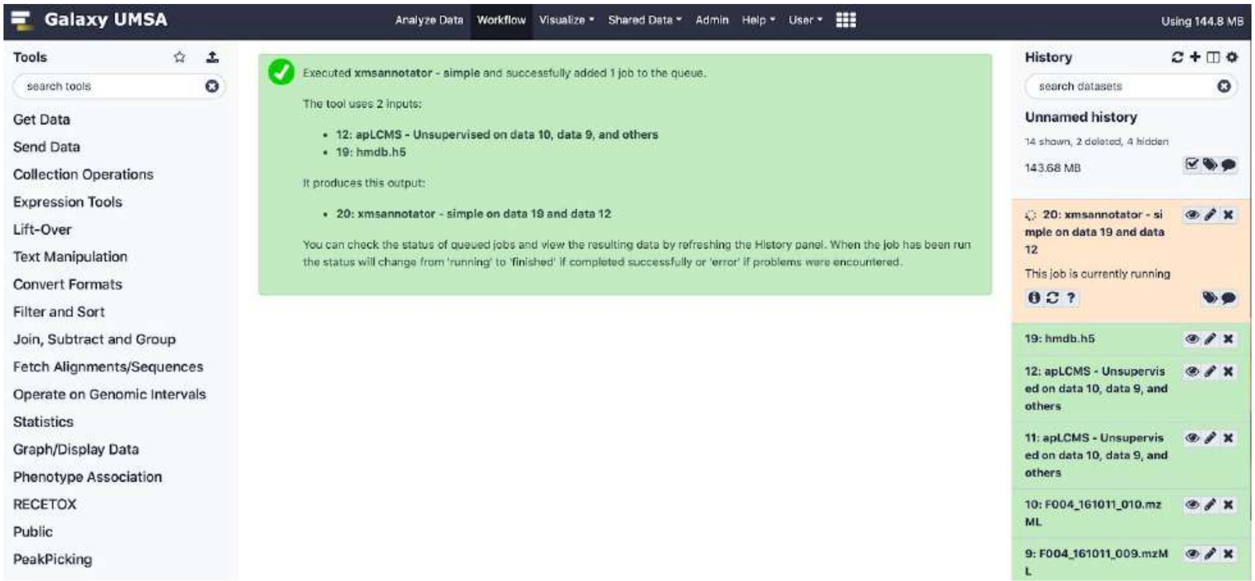
Metabolite database

19: hmdb.h5

Mass tolerance [ppm]

10.0

Figure 12. Screenshot of an input form of a simple workflow in UMSA.



Executed **xmsannotator - simple** and successfully added 1 job to the queue.

The tool uses 2 inputs:

- **12: apLCMS - Unsupervised on data 10, data 9, and others**
- **19: hmdb.h5**

It produces this output:

- **20: xmsannotator - simple on data 19 and data 12**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

Unnamed history

14 shown, 2 deleted, 4 hidden

143.68 MB

20: xmsannotator - simple on data 19 and data 12

This job is currently running

19: hmdb.h5

12: apLCMS - Unsupervised on data 10, data 9, and others

11: apLCMS - Unsupervised on data 10, data 9, and others

10: F004_161011_010.mzML

9: F004_161011_009.mzML

Figure 13. Screenshot of the execution of the previous workflow in progress. The peak-picking step (apLCMS) has already finished, annotation is running.

4. Other Thematic Services

All the thematic services in EOSC-SYNERGY have progressed in the adoption of EOSC services. This section includes a shorter description of the services not included in the first part of the document. For each service, we include an updated description, the revamped architecture that includes the EOSC services and their description.

4.1. G-CORE

G-Core is a production-ready technology used as a service at ESA's and national programs led by INDRA for the acquisition, storage, cataloguing and processing data from several EOS missions. G-Core provides two main functionalities:

- A Data Manager for spatial and non-spatial purposes.
- A Processing framework to host external processors developed by third parties to generate added value products based on Satellite imageries.

The objective of the adaptation of the thematic service is to explore the sustainability of the EOS services exposed through the creation of added-value products through the integration of G-Core as a data manager.

Figure 14 represents a typical GCORE structure with a marketplace where different services are offered from third parties. The GCORE will offer different GCORE Instances types that can be deployed according to the user requests. Each user can request a specific service that implies the deployment of a GCORE Instance to perform the activities for the services. The capability of GCORE to integrate different processors and to deliver its results allows creating new services in an easy way.

In addition, all services available would benefit from the elastic and scalable up and down capacity of GCORE in order to reduce the operational costs of a third-party project.

This SW will be offered as a component to be included in complex system to implement the processing chain or as a processing framework to create hybrid infrastructures for processing activities on demand or could be published as SaaS to publish the services in the market place to be used by companies/entities or institutions focused to create added value services/products over satellite imageries.

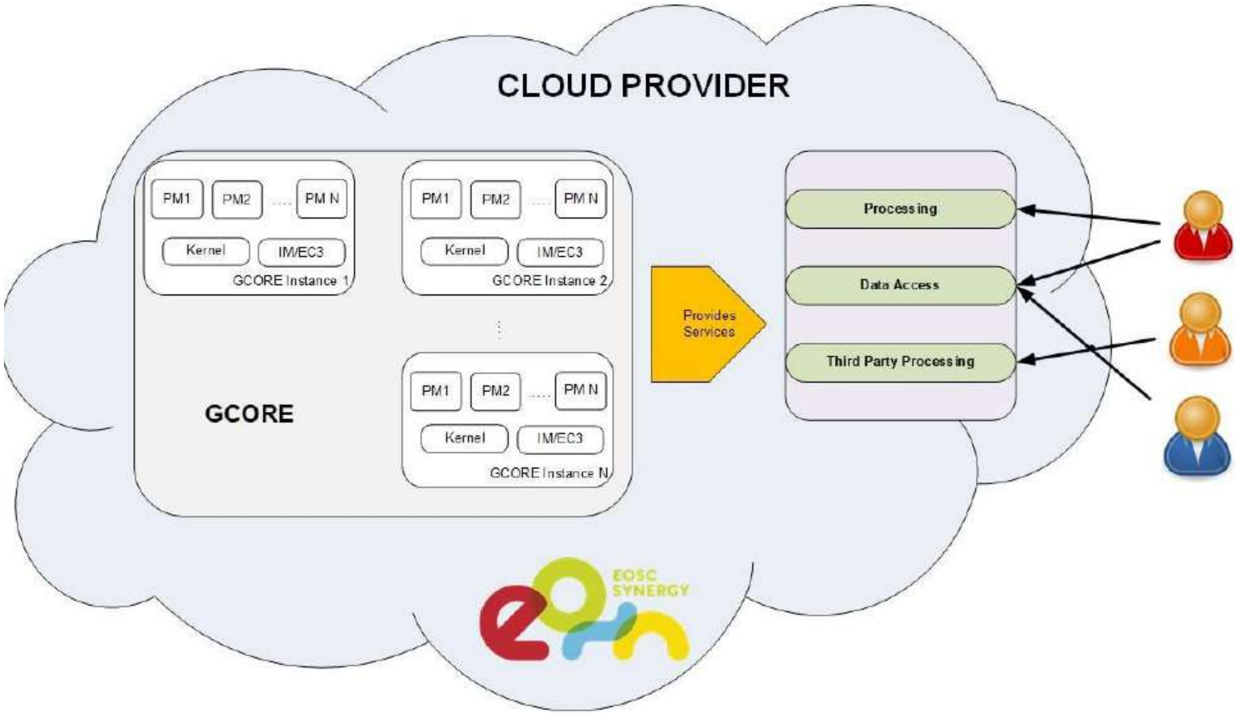


Figure 14. High level architecture for GCORE as EO thematic services

The G-Core service targets the following three user profiles:

- EO data for the science community to use the satellite data in the scientific studies.
- EO data for public organizations to use the satellite imageries as background data.
- EO data for value adders to create added value products from satellite images.
- It will help to define new products and services mixing Earth Observation data with other types of data for scientific and social environments

The expected impact of the adaptation of the service is to democratize the usage of EO data out of the scope of nominal fields. It will help to define new products and services mixing Earth Observation data with other types of data for scientific and social environments.

4.2. SCIPION

Scipion is an application framework developed by I2PC in Madrid to help the Structural Biology community to process CryoEM data. Scipion is developed as a plugin-based workflow management system that integrates many important software packages available in the field. ScipionCloud, currently an EGI AppDB image, is being modified to use standard EOSC services to enhance service deployment and user access as well as an optimized usage of cloud resources. The ScipionCloud service will allow Instruct users to deploy a dynamic cluster in the cloud to keep processing the data acquired at the facility.

The architecture of SCIPION service is shown in figure 15. User acquires images at a microscope facility where some automatic preprocessing is done using Scipion. Raw data and projects are stored locally. Later, a user can access the EC3 portal (or a customized portal embedding EC3 client) using her ARIA credentials on the EGI check-in service and deploy an elastic cluster.

The cluster comprises a front-end node with Scipion, a shared storage containing data and project (raw data might be skipped due to the size) and the required components to spawn worker nodes when a job is launched through SLURM. Users will interact with the front-end server via noVNC from a web browser.

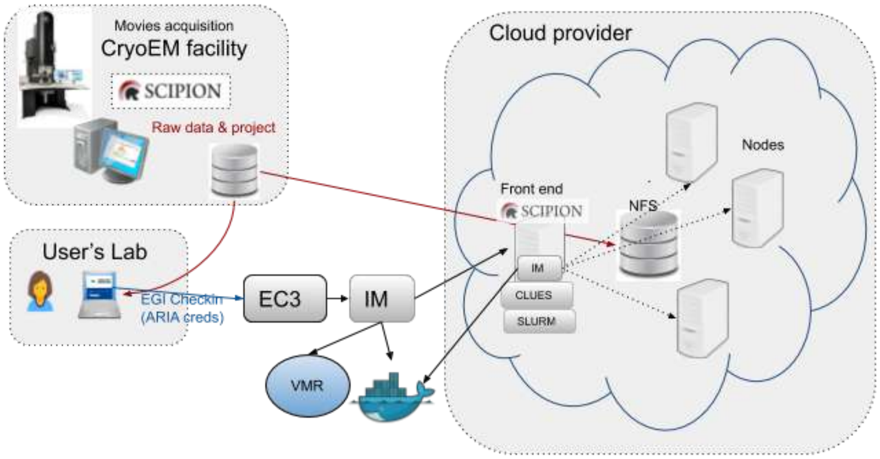


Figure 15. Architecture of the SCIPION Service.

SCIPION service includes or will include the following services listed in the EOSC marketplace:

- EGI check-in: Using ARIA IdP (INSTRUCT users). Needed to access the EC3 portal of EOSC-Synergy and to deploy the cluster in EOSC resources.
- EC3 and IM: These services are used to deploy an elastic cluster on EOSC cloud resources or public clouds such as AWS EC2.
- EOSC cloud resources: The cluster might be deployed on EOSC federated cloud if credentials permit it.

4.3. LAGO

LAGO is a cosmic ray observatory composed of a network of water-Cherenkov detectors (WCD) spanning over different altitudes and latitudes making research on High Energy Physics, Space weather, Life Sciences, Aerospatial security, Computer Science, etc. The measurements collected from these detectors are posteriorly processed and analysed. Additionally, scientists continuously generate simulated data. The final purpose is to enable the long-term curation and re-use of data within and outside LAGO through a Virtual Observatory.

The architecture of the thematic service is shown in figure 16. LAGO's thematic service is focused on providing a standardised way to curate and reuse measurements, analysis and simulations. To achieve this task, it follows the basic design recommended by EGI/EOSC for cloud: core intelligence packed in Docker images, being able to automatically check, store and publish their results in DataHub, with enough metadata to be used by official harvesters (B2FIND), which will act as virtual observatories.

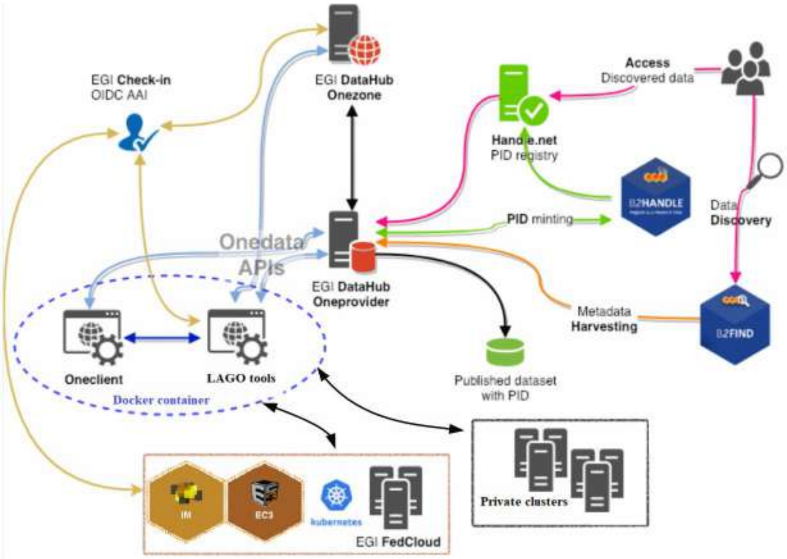


Figure 16. Architecture of the LAGO thematic service.

As the whole computation is self-contained in the image, the production can be easily performed by services such as EC2/IM or even manually in private clusters.

LAGO thematic service includes or will include the following services listed in the EOSC marketplace:

- EGI Check-in (through EduTeams Perun at GEANT): it is needed for accessing any EOSC service, in particular for obtaining a OneData token. Managing the VO with Perun at GEANT was considered because of flexibility and their long-term support to Latin American users.
- EGI DataHub: OneData allows researchers several ways to access the data and metadata of their interest. Collaboration members can directly explore the directory tree at https://datahub.egi.eu or mount it on their PC's. Meanwhile, the general public will get published data through B2FIND. On the other hand, OneData eases storing results without modifying simulation/processing codes, as well as maintaining usable replicas around the world.
- The EOSC Cloud services (IM and EC3) will be explored in the coming months to validate the deployment of batch or Kubernetes clusters.

Additionally, LAGO plans to explore other services such as B2FIND, B2HANDLE and DIRAC4EGI.

4.4. SDS-WAS

SDS-WAS provides a set of services related to the mineral dust forecast. It collects numerical model and observational data from a wide set of partners plus internally developed, generates post-processed analysis and statistics, disseminates results to users and organizes training courses and events. The aim is to give support to institutional entities (e.g. National Meteorological Agencies) to warn about possible dust events and to foster the study of dust-related phenomena into the academic and research communities. The EOSC infrastructure will give the possibility to increase the quantity of data hosted and processed and reach a wider set of users.

As shown in figure 17, the service is configured as an High Availability Cluster of two duplicated instances with a frontend of a web portal (which shows and disseminate all services) and a backend of a local storage and a set of libraries for post-processing/data analytics. The authentication service will be the entry point, and all products will be available through the B2SHARE service. The B2SAFE service will federate all storages running as a backup service.

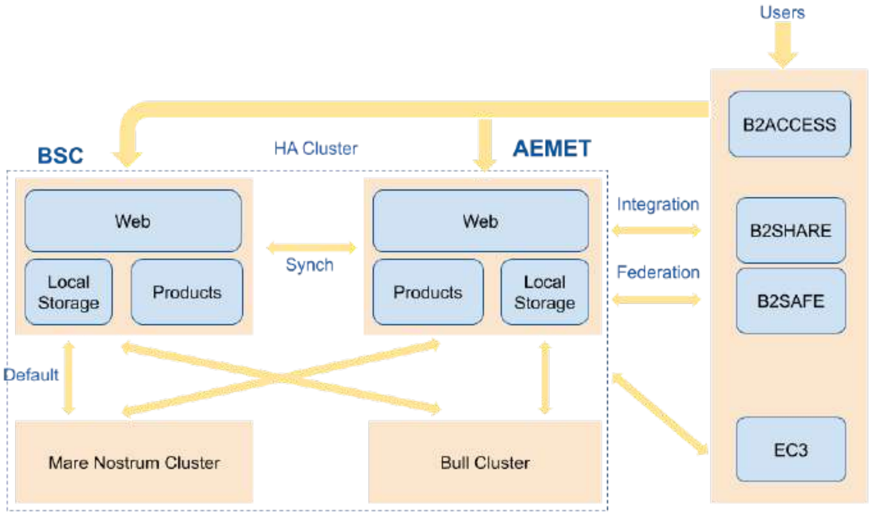


Figure 17. Architecture of the SDS-WASS thematic service.

SDS-WASS thematic service includes or will include the following services listed in the EOSC marketplace:

- B2ACCESS used to authenticate users into the services.
- EC3 to run dynamically the in-house numerical model simulation plus some post-processes.
- B2SHARE will be the service to expose and disseminate developed products.
- B2SAFE will run as a BACKUP service.

4.5. MSWSS

MSWSS is a service for modeling and analysis of Water Supply Systems which integrates the analysis of toxics in drinking-water supply networks with water distribution network simulation. MSWSS service will allow water infrastructure operators and researchers to analyse hazardous events (e.g. toxics propagation within a pipe system) and may be used for preparation of risk management plans for water utilities. The EOSC computing infrastructure and data sharing services enable modelling more complex water supply systems and to increase the number of scenarios for the analysis.

The architecture of the MSWSS service is depicted in figure 18. The MSWSS service uses Galaxy portal where users can share and reuse their workflows with data and prepare their simulation jobs.

MSWSS uses as a computational backend based on the EC3 elastic virtual cluster service which provides the MSWSS service with the building and management of an elastic virtual cluster on the computational resources available in the EOSC IaaS infrastructure. The resources in the virtual cluster are managed by the Slurm workload management system. The data produced by computational jobs are stored locally within the MSWSS service and are available to users via Galaxy portal.

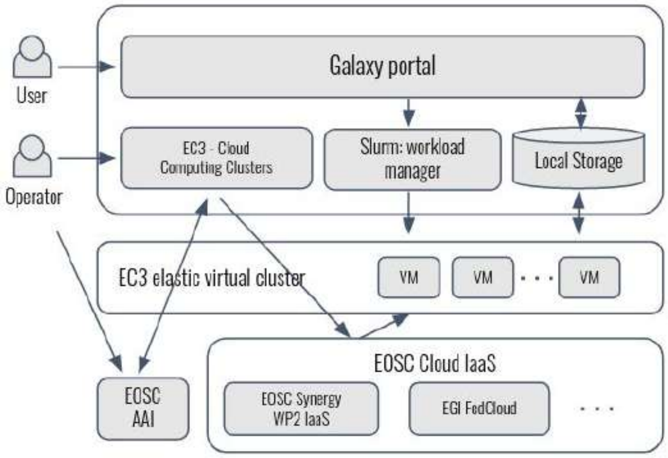


Figure 18. Architecture of the MSWSS Thematic Service.

MSWSS thematic service includes or will include the following services listed in the EOSC marketplace:

- EC3 (Infrastructure Manager, CLUES): is used for creation and management of computational backend based on elastic virtual cluster built from virtual worker nodes
- EOSC Cloud computing resources: are used to build the elastic virtual cluster for MSWSS service
- EGI Check-in: it is used by EC3 to authenticate the MSWSS service to EOSC Cloud computing resources

4.6. O3AS

O3AS is a service mainly for scientists working on the CCMI (http://blogs.reading.ac.uk/ccmi/ccmi-phase-two/) and writing the quadrennial global assessment of ozone depletion (https://www.esrl.noaa.gov/csl/assessments/ozone/2018/). The recent ozone assessment report consists of six chapters and five appendices with about 25 people actively working on each chapter and a multitude of people working in support of the preparation of the document. The O3AS service shall provide an invaluable tool to extract ozone trends from large climate prediction model data to produce figures of stratospheric ozone trends in publication quality, in a coherent way.

Figure 19 shows the architecture of the O3AS technical service. O3as service is split in several actions and components:

1. A user configures his/her request in the Web Application.

2. This request is passed to O3as service via the O3as REST API call.

3. O3as service processes the request, where the pre-processed data (aka skimmed data) is accessed via WebDav and OIDC.

4. In order to produce skimmed data, regular tasks run on HPC to copy primary data and perform data preparation (e.g. data reduction and parameter unification).

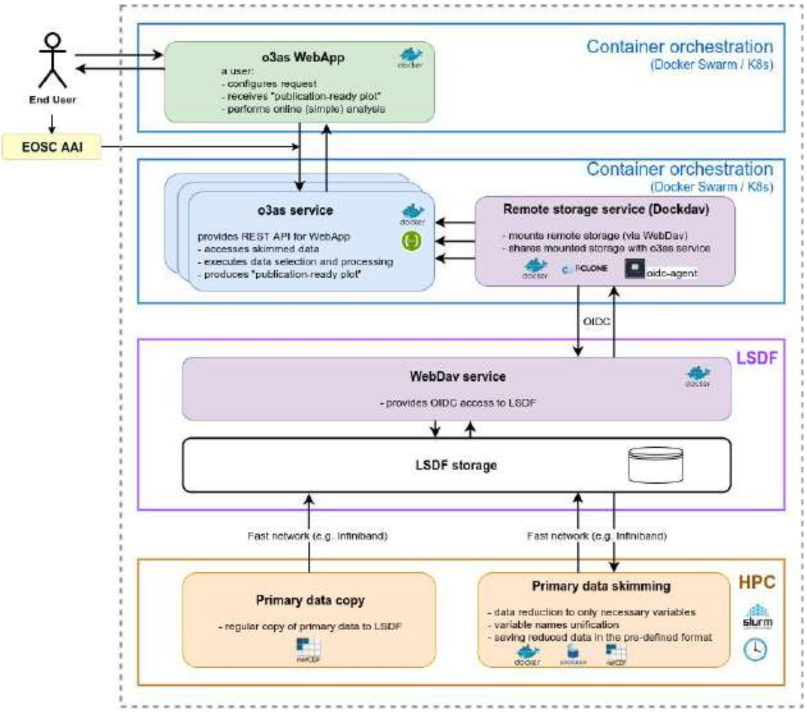

End User
EOSC AAI
o3as WebApp
a user:
- configures request
- receives "publication-ready plot"
- performs online (simple) analysis
docker
Container orchestration
(Docker Swarm / K8s)
Container orchestration
(Docker Swarm / K8s)
o3as service
provides REST API for WebApp
- accesses skimmed data
- executes data selection and processing
- produces "publication-ready plot"
Remote storage service (Dockdav)
- mounts remote storage (via WebDav)
- shares mounted storage with o3as service
RCLONE
oidc-agent
OIDC
LSDF
WebDav service
- provides OIDC access to LSDF
LSDF storage
Fast network (e.g. Infiniband)
Fast network (e.g. Infiniband)
HPC
Primary data copy
- regular copy of primary data to LSDF
netCDF
Primary data skimming
- data reduction to only necessary variables
- variable names unification
- saving reduced data in the pre-defined format
udocker
slurm

Figure 19. Architecture of the O3AS Thematic Service.

O3AS thematic service includes or will include the following services listed in the EOSC marketplace:

- EOSC OIDC providers, e.g. EGI Check-In: to access certain functionalities of the service.
- OIDC-Agent: To mount data servers using WebDAV - HTTP authentication.
- udocker: To perform data skimming in the HPC environment.
- Infrastructure Manager: To deploy service resources (e.g. K8s).

5. Conclusion

This report supports the demonstration deliverable D4.2 - First prototype of the EOSC Thematic services which shows four operational EOSC-SYNERGY thematic services although with limited functionality. The deliverable shows additional information about the status and plans of all the thematic services.

All thematic services have identified several EOSC technical services to address some of the challenges and requirements that were not properly fulfilled in the initial versions. Each thematic service has differences that led to the adoption of one or another thematic service, which enriches the catalogue of experiences, best practices and solutions. As a summary, three different (although compatible among them) AAI methods have been integrated (EGI Checkin, B2ACCESS and Life-Sciences AAI). Job scheduling ranges from solutions based on containers (using Kubernetes) to solutions using batch queues (mainly based on SLURM), supported in some cases by workflow frameworks such as Galaxy and instantiated through EC3. For the interaction with cloud resources, TOSCA and RADL recipes have been developed for Infrastructure Manager. Data access is performed through different solutions such as Dataverse, WebDav, EGI Datahub OneData, B2SHARE and B2SAFE, which clearly states the complexity of the data management issue and the wide range of solutions.

The experience among the thematic services will be extremely useful for new services to be developed, so an important effort on communication will be performed. Opportunities such as the EOSC-hub week will be used to showcase demonstrations and presentations.