# Thematic Services in EOSC-SYNERGY

Ignacio Blanquer
On behalf of WP4

IBERGRID Session 2022

# Agenda

- General summary
  - The virtuous cycle of the Thematic Services.
  - Adoption of services from the EOSC Marketplace.
  - Improvement of Quality through SQAAaS pipelines.
  - Training Material.
  - Usage Metrics.
- Details per TS.
- Conclusions.

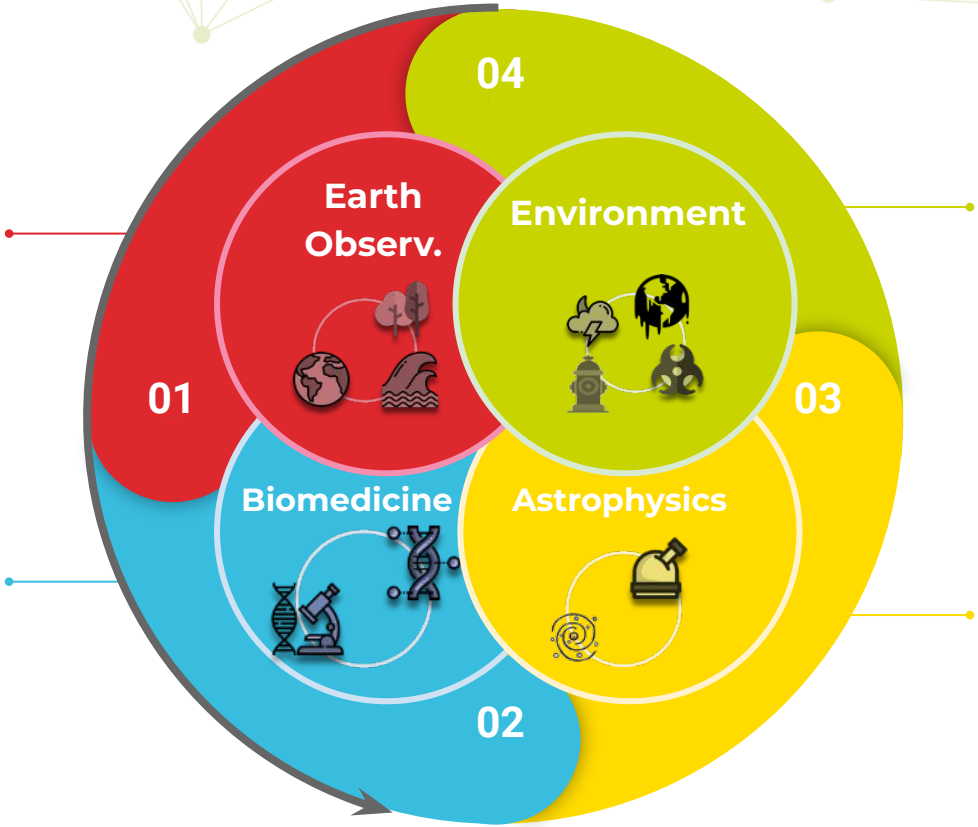
# 0. The Thematic Services Virtuous Cycle

**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.

# Increasing the capacity, performance, reliability and functionality: Adoption of EOSC Services

| | | |
|---|---|---|
| **AAI** | **8 services adopted EGI-Checkin, 1 B2Access and 1 Life Sciences AAI** | • Support of OpenID and external IdPs<br>• Higher trust and convenience for users.<br>• Interoperability among services. |
| **Data** | **2 EGI Datahub, 3 B2*, 2 Dataverse and 6 improved local storages.** | • PID and DOI Assignment.<br>• Long-term data preservation.<br>• Data and metadata archiving. |
| **Resources** | **8 services adopted TOSCA cloud orchestrators.** | • Agnostic to cloud IaaS backends.<br>• Automatic cluster elasticity.<br>• Also using batch queues as a backend. |
| **Workload** | **3 K8s backends, 8 SLURM Based backends, 1 Workflow-based.** | • Migration to container-based solutions.<br>• Improvement from local to backend-based workloads.<br>• Integration in Galaxy environments. |

# Increasing the capacity, performance, reliability and functionality: Adoption of EOSC Services

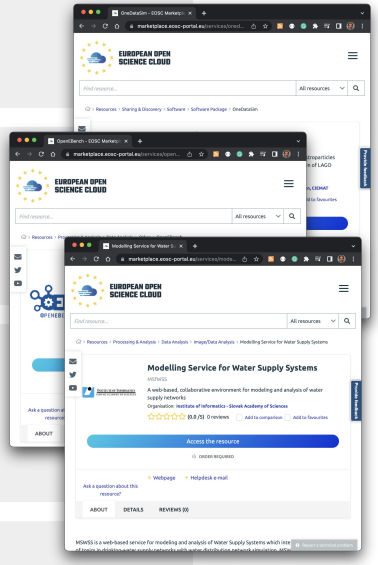
| | | |
|---|---|---|
| **SW** | **28 SW pipelines** | • Documentation, environment, security Coding style, Functional tests, test coverage.. |
| **Service** | **10 Deployment pipelines** | • Testing TOSCA recipes through IM or YAML specifications in K8s.<br>• Downloading, building, installation, Coding style, Docker images |
| **FAIR** | **>4 FAIR Evaluation pipelines + several (TBD) FAIR Evaluation Assesments.** | • Integration in FAIR pipelines.<br>• Direct execution of the FAIR Evaluator or the F-UJI tool. |
| **Assurance** | **>5 Bronze Badges (Check)** | • Directly obtained from the SQAaaS platform.<br>• Working to reach at least one “Silver” Badge. |

# Increase the awareness: Training materials produced

| | | |
|---|---|---|
| **Courses** | **4 full courses on the moodle platform and 9 training objects.** | • Describing the adoption process and outlining the advantages of the adaptation.<br>• Additional information in the usage of the TS. |
| **Catalogue** | **9 thematic services registered in the EOSC Portal Marketplace** | • XXX access requests.<br>• 14+70+17+34 Global visits |
| **WWW** | **10 Separate sections in the web site.** | • Available in the “EOSC for Researchers” section<br>• Grouped by areas.<br>• Including access endpoints. |
| **Videos** | **10 demonstration videos** | • Showing up the advantages of the adaptation.<br>• Useful also as a training item. |

# Increasing relevance: Measuring success - Metrics and KPIs: Linkage to e-Infrastructures

| | | |
|---|---|---|
| **Users** | **8 Thematic VO, with 49+14+3+5=71 users, plus over (TBD) anonymous users.** | • A relevant user community.<br>• Many indirect usage. |
| **Compute** | **Over 8M CPU hours and nearly 8K VM deployed** | • A sustained usage rate with processing burst in data challenges.<br>• Plus other resource consumption not reflected in the accounting portal. |
| **Data** | **Difficulties on collecting information on Data processed** | • This is a weak point to correct. |
| **User Experience** | **Collecting user experience through questionnaires** | • Evaluating Usability, correctness, robustness, completion, efficiency and convenience. |

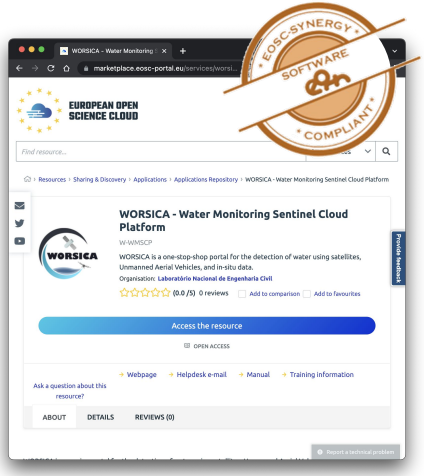


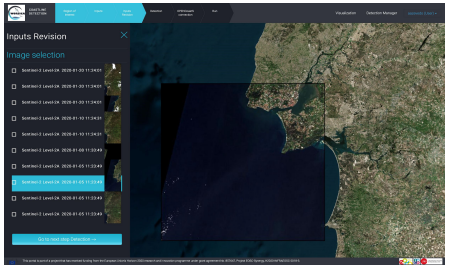
# Temporary slide

- WORSICA
  - Portal (adds/visits): XX/185
  - Users (VO/others): 4/49
  - Anonymous visits: N/A.
- SAPS
  - Portal (visits): 14
  - Users (VO): 18
- gCORE
  - Users (VO/others): XX/XX
  - Anonymous visits: XXX.
- OpenEBench
  - Portal (adds/visits): XX/XX
  - Users (portal): XX
  - Anonymous visits: XXX.
- SCIPION
  - Portal (adds/visits): 4/17
  - Users (VO/others): 3/6
  - Anonymous visits: N/A.
- LAGO
  - Portal (adds/visits): XX/XX
  - Users (VO/others): XX/XX
  - Anonymous visits: XXX.
- SDS-WAS
  - Portal (adds/visits): 0/70
  - Anonymous visits: XXX.
- UMSA
  - Portal (adds/visits): XX/XX
  - Users (VO/others): XX/XX
  - Anonymous visits: XXX.
- MSWSS
  - Portal (adds/visits): 0/34
  - Users (VO/others): 5/2
  - Anonymous visits: n/a.
- O3AS
  - Portal (adds/visits): XX/XX
  - Users (VO/others): XX/XX
  - Anonymous visits: XXX.

# WORSICA: Water Monitoring Sentinel Cloud Platform

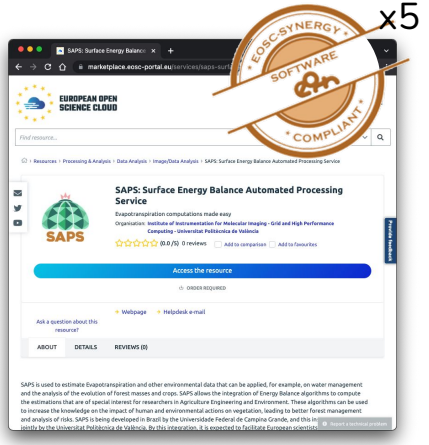
- A service for coastline detection, inland water bodies detection and water leaks detection on irrigation networks.
  - Available in: https://worsica.incd.pt
  - 3 Pipelines developed (1 SW + 1 K8s + 1 FAIR EVA)
  - >140 VMs and >750K CPU·hours
  - Collecting User Experience.
- Future Plans and exploitation
  - Integration of new methodologies for the automatic correction of coastline water levels
  - Implementation of machine learning algorithms for the coastline and Inland services



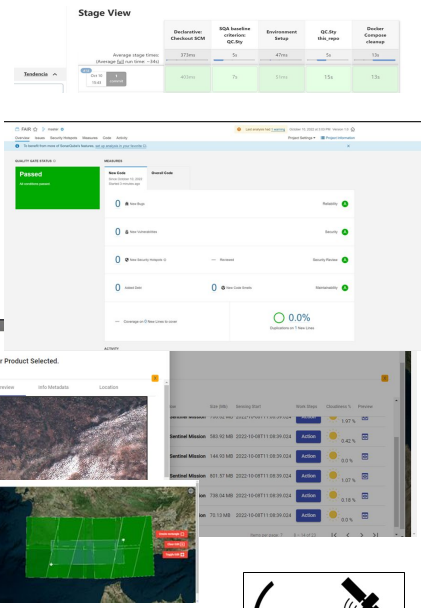
# SAPS: Surface Energy Balance Automated Processing Service

- A service to estimate evapotranspiration and other environmental data that can be applied on water management and the analysis of the evolution of forest masses and crops.
  - Available in http://saps.vm.fedcloud.eosc-synergy.eu/
    - And in the EOSC Marketplace:
  - 8 Pipelines developed (7 SW + 1 Service deploy).
  - 134 VMs, >250K CPU*hours.
  - Collecting User Experience.
- Future Plans and exploitation
  - EGI ACE grant from the 4th Call.
  - Early adopters program to extend the support.
  - Promotion in the EPOS community.




# G-Core: Acquisition, cataloguing and processing EOS data

G-Core is a service for the acquisition, storing, cataloguing and processing of data from several Earth Observing System (EOS) missions that can be used as a Data Manager for spatial and non-spatial purposes and a Processing Framework to host external processors developed by third parties to generate added-value products based on Satellite imagery.

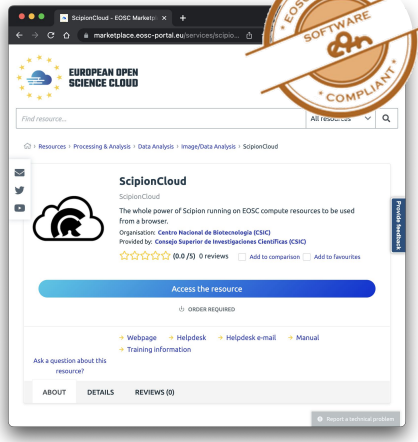


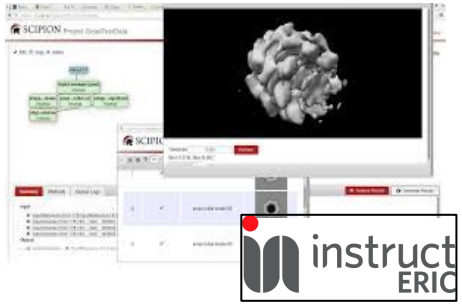

- Available in: https://gcore.vm.fedcloud.eu/eosc-viewer/logging
- 10 Pipelines developed (4 Service, 4 SQaaS, 2 FAIR)
- 77 VMs >130K*CPU hours, 13K GB used
- Collecting User Experience

- Future Plans and exploitation
  - Provide processing services to other applications
  - Create a mirror of Sentinel data offered as a service in EOSC

# ScipionCloud: CryoEM data processing for Structural Biology

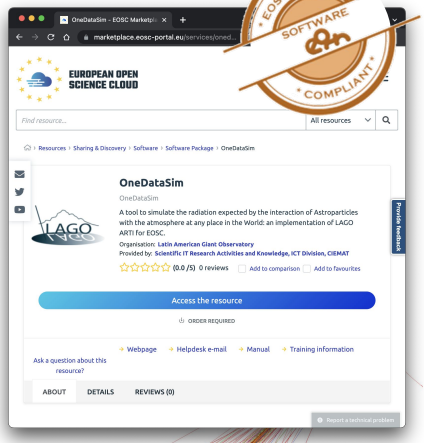
- Service that deploys a virtual infrastructure in the cloud containing CryoEM software for users to process their own microscope data.
  - Available in https://appsgrycap.i3m.upv.es:31443/im-dashboard/configure?selected_tosca=scipion.yaml
    - EOSC marketplace: https://marketplace.eosc-portal.eu/services/scipioncloud
  - 3 Pipelines developed (1 SW + 1 Service deploy + 1 FAIR ongoing)
  - 810K CPU hours, 28K GB RAM
  - User experience survey (4 answers)
- Future Plans and exploitation
  - OneData integration and K8s cluster (EC3)

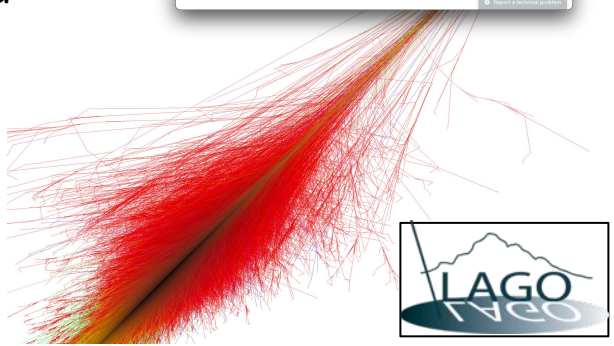


# LAGO: Latin American Giant cosmic ray Observatory

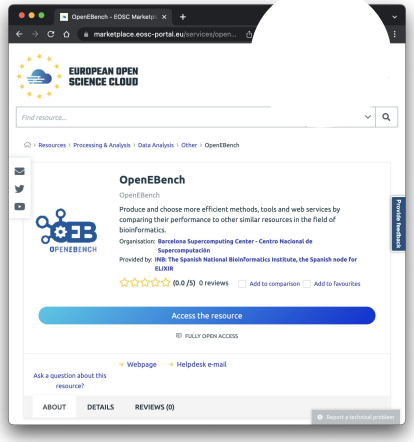
- LAGO is a cosmic ray observatory made of a network of water- Cherenkov detectors (WCD) spanning over different altitudes and latitudes making research on High Energy Physics, Space weather, etc.
  - Available in https://marketplace.eosc-portal.eu/services/onedatasim
  - 3 Pipelines developed
  - ~ 2M CPU·h of air shower simulations. Current and future LAGO sites are now fully characterized
  - +15 Users Experience
- Future Plans and exploitation
  - Space weather impact, mining prospecting, underground labs, volcanic risks assessments

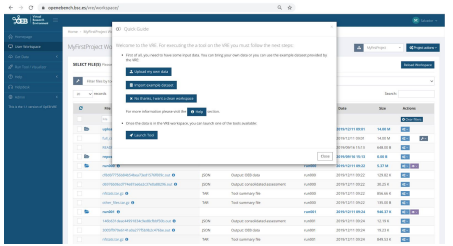


# OpenEBench: ELIXIR benchmarking and technical monitoring platform

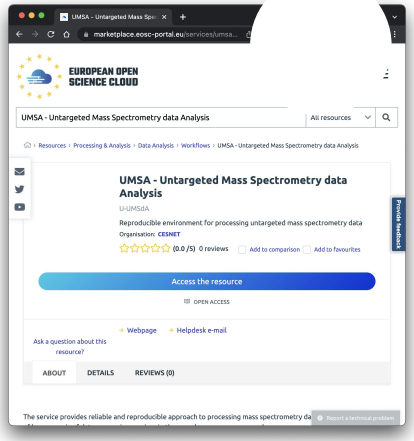
- Used to evaluate bioinformatics tools, OpenEBench is an observatory for SW quality based on the automated monitoring of FAIR for research software metrics and indicators.
  - Available in https://openebench.bsc.es/
  - 1 Pipelines developed
  - Over thousands of monthly visits.
  - User Experience evaluated ~4 out of 5 by users .
- Future Plans and exploitation
  - Automating some processes adding functionality
  - 8 Scientific groups using OEB, 6 with public benchmarking results and 2 in preparation
  - Used in current and new research projects (EuCanImage, 2020-2024; IMPaCT-Data, 2021-2023, DataTools4Hearth, 2022-2026; EUCAIM 2023-2027)

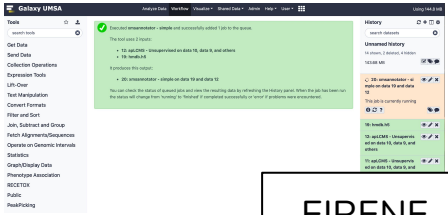


# UMSA: Untargeted Mass- Spectrometry Analysis

- 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.
  - Available in https://umsa.cerit-sc.cz/
  - Pipelines developed: Github CI/CD for individual tools, Ansible service deployment
  - Usage: 4143 jobs in 1-9/2022
  - 3 User Experience (informal yet)
- Future Plans and exploitation
  - extend user community (collaborators from newly spawned projecte)
  - go beyond limits of current tools (detection of low-abundance compounds) with new aproaches (ML, ...)

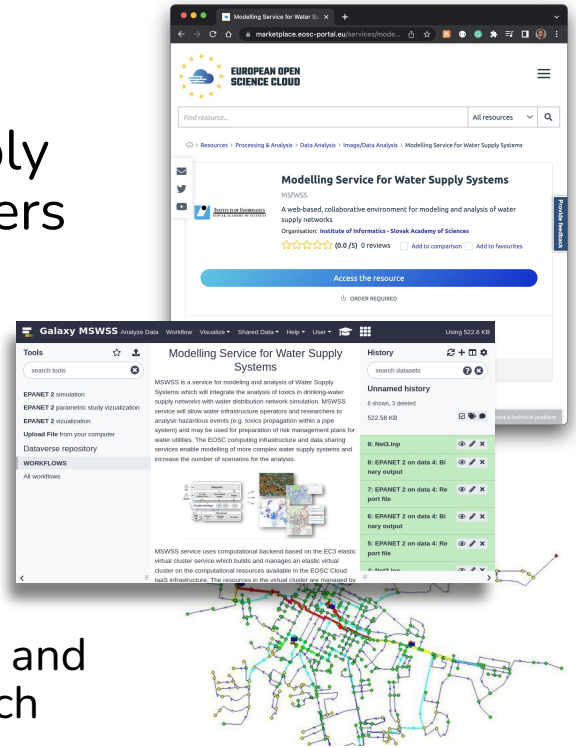



EIRENE

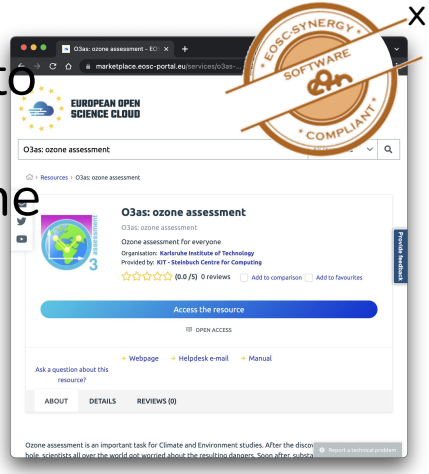
# MSWSS: Water Supply Systems modeling and analysis

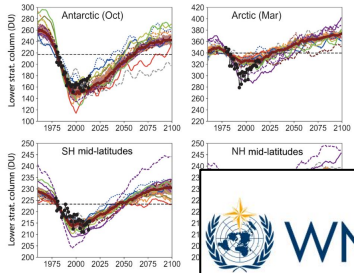
- MSWSS integrates the analysis and simulation of toxics in drinking water supply networks to allow operators and researchers to analyse hazardous events.
  - Available in https://mswss.ui.savba.sk/ and in EOSC Marketplace
  - 1 Deployment pipeline developed
  - ~200k CPU hours usage
  - +2 Users Experience
- Future Plans and exploitation
  - Improve users experience by adding more tools and features, the service will be used in new research projects



# O3AS: Ozone Analysis Service

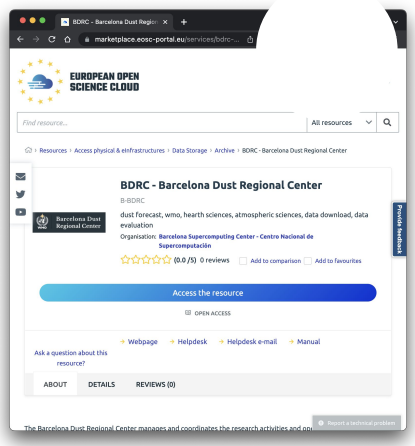
- The O3AS service shall provide an invaluable tool to extract O3 trends from large climate prediction model data to produce figures of stratospheric ozone trends.
  - Available in https://o3as.data.kit.edu
  - 4 Pipelines developed (3 SW + 1 Deploy)
  - Usage: 40 Unique Visitors/day, ca.200 API calls/day
  - 5 User Experiences collected
- Future Plans and exploitation
  - We intend to extend and use this tool for the upcoming ozone assessment exercises, publishing ozone return rates for climate model projections.

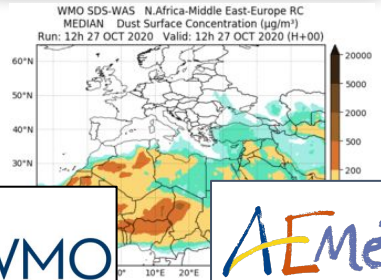


# SDS-WAS: A Service related to the mineral dust forecast

- The Barcelona Dust Regional Center (SDS-WAS) manages and coordinates the research activities and operations of the World Meteorological Organization (WMO) related to sand and dust storms, provides access to available dust products and coordinates a network of collaborators (researchers, data providers and users' communities) in Northern Africa, the Middle East and Europe.
  - Available in https://dust.aemet.es/ and BDRC - Barcelona Dust Regional Center - EOSC Marketplace
  - 5 SW Pipelines developed
  - Usage: ~100 to ~700 average daily users
  - Collecting User Experience
- Future Plans and exploitation
  - Improvement of current products (3D fields, mobile version, better data download experience), development of a climate services dashboard, training courses and webinars



# D4.4. EOSC Thematic Services Validation Report

1. Executive Summary (UPV)
2. Introduction (UPV)
3. Evaluation using the SQAaaS Pipelines
    - 3.X. Thematic Service XX (One section per TS)
        - 3.1.1. Software and Service Deployment Pipelines
            - Include a list of the SW Quality pipelines created manually or through the portal and the metrics used.
            - If service evaluation is implemented, include a list.
        - 3.1.2. FAIR Evaluation
            - Include the pipelines for the FAIR evaluation.
        - 3.1.3. Quality Assurance
            - Badges obtained from the SQAaaS platform.
4. Metrics
    - 4.1 Metrics for the Impact on Users
        - Number of users registered.
    - 4.2. Metrics for the Impact on Capacity and Capability
        - Number of service Accesses, Data Consumed / Produced, Processing and Memory from the Accounting
    - 4.3. Metrics for the Impact on Scientific Outreach
        - Publications, sessions and training hours.
    - 4.4. Metrics for the Impact on Usability
        - Usability questionnaire.
    - 4.5. Metrics for the Impact on Cross-fertilization
        - Cross-collaboration.
5. Conclusion

# Joint Dissemination actions

- EGI Conference [EOSC Synergy Session]
- Materials from “Thematic Service integration in EGI Fedcloud” are available here: https://docs.google.com/presentation/d/1zJF_fjdxTSeNl2rQ0rlbzcQBrvI7AQwP_se7O8KlObc/edit?usp=sharing

- A joint paper submitted to Computer Science Review.

A survey of the European Open Science Cloud services for expanding the capacity and capabilities of multidisciplinary scientific applications

Under Review

Last review activity: 4th July 2022

- Reviews completed: 1
- Review invitations accepted: 1
- Review invitations sent: 2

- Still under review.

# Conclusions

- The ten thematic services have been improved using services from the EOSC marketplace
  - The adaptation, improvement and quality assessment of those services on a Federated Data Infrastructure strongly aligns with the objectives of EOSC.
- Software quality improves legibility and maintainability and it is a path for creating reusable software research artifacts.
- Assurance
- The Thematic Services have served as the “guinea pig” and the source of requirements for many of the EOSC-SYNERGY developments.
  - The developments constitute best practices for other services to develop.