

Quality and Capacity expansion of Thematic Services in EOSC-SYNERGY

Tuesday, 3 November 2020 10:05 (20 minutes)

EOSC-SYNERGY Thematic services aim to increase the acceptance of EOSC by building capacities and introducing improved platform and infrastructure services. EOSC-SYNERGY has identified ten thematic services addressing four scientific areas (Earth Observation, Environment, Life Sciences and Astrophysics). Those thematic services are heterogeneous, address a wider range of requirements and have different maturity levels, targets and usage models. In the field of Earth Observation, the services deal with monitoring coastal changes and inundations, processing satellite image data and estimating forest mass. In the field of environment, they include stratospheric ozone monitoring and the protection and recovery of the ozone layer, the forecast of sand and dust storms, the simulation of water network distribution and untargeted mass-spectrometry analysis for toxics. In Astrophysics, the project will set up a European service for the Latin American Giant Observatory, and in Life Sciences, EOSC-SYNERGY covers both a platform for supporting community-led scientific benchmarking efforts and the processing of Cryo-electron microscopy imaging.

These thematic services will be improved in terms of authentication and authorisation, resource management, job scheduling, data management and accounting. Not all services have identified gaps in all aspects, so each thematic service will focus on those that are relevant according to their bottlenecks.

The thematic services have several technical similarities and differences. Common to all thematic services is the need for robust authentication and authorisation infrastructure compatible with those used by the users' institutions. The EGI Check-in has a widely accepted choice although services like ELIXIR AAI - soon to be upgraded to Life Sciences AAI - are also important assets. With respect to resource management, all services have an interest in providing processing resources dynamically. The Infrastructure Manager and the Elastic Compute Clusters in the Cloud have been identified by most of them as candidate technologies for this gap. Regarding job management, most services use batch queues, which could be extended to support containerised jobs. The use of Kubernetes to orchestrate microservices and containerised job queues are also being considered. The most challenging part is data management. Thematic services have identified issues in transferring and accessing large amounts of data requiring smart caching, advanced data transfer and persistent massive data storage.

The thematic services expect a workload between 400 and 46.500 CPU hours per week (a cumulative 71K CPU hours per week), consumed by up to 10k jobs per week requiring a median of 16 GB RAM and 15 GB of storage per job. The persistent storage requirements range from 2 GB to 500 GB (a median of 100GB and a total of 1 PB).

The thematic services have also defined a set of performance metrics grouped into five impact categories (users, service Capacity and Capability, Scientific Outreach, service usability and Cross-Fertilization). These metrics can provide quantitative indicators of the performance and improvement of the thematic services.

Thematic services constitute a key activity to evaluate the impact of the capabilities in EOSC-SYNERGY with respect to adopting mature and scalable services, software and service quality assurance, increased resource capacity and improved user skills.

Primary authors: BLANQUER, Ignacio (UPVLC); AZEVEDO, Alberto (LNEC - Laboratório Nacional de Engenharia Civil)

Co-authors: Dr PEREIRA, Thiago Emmanuel (Universidade Federal de Campina Grande); Mr PAVESIO-BLANCO, Manuel (INDRA); Dr CAPELLA-GUTIERREZ, Salvador (BSC); DEL CANO, Laura (CSIC); Dr ANTONIO JUAN, Rubio-Montero (CIEMAT); ASTALOS, Jan (Institute of Informatics, Slovak Academy of Sciences); Dr KERZENMACHER, Tobias (KIT)

Presenter: BLANQUER, Ignacio (UPVLC)

Session Classification: Data analytics and thematic services - part 2