Contribution ID: **25**   Type: **Presentation (15' + 5' for questions)**

Reproducible Open Science with EGI Services

Thursday, 28 September 2023 13:15 (20 minutes)

Reproducibility is a cornerstone of scientific research, ensuring the reliability and validity of results by allowing independent verification of findings. During the EGI-ACE project, EGI has developed EGI Replay, a service that allows researchers to reproduce and share custom computing environments effortlessly. With Replay, researchers can replicate the execution of your analysis in a notebooks-based platform, ensuring that others can easily access and interact with the content.

EGI Replay is based on Binder technology, which builds computing environments on the fly from a code repository that contains the code you'd like to run, as well as a set of configuration files that determine the exact computing environment to run it. Replay also generates shareable links for others to interact with the content from any browser. This means other researchers can easily reproduce the analysis and access data available in EGI's infrastructure, making it easier than ever to collaborate and share your work with others.

EGI Replay is powered by the EGI distributed infrastructure and integrated with additional EGI services:

- Check-in provides federated authentication to Replay and integration with EOSC AAI. Replay automatically generates and refreshes access tokens for accessing any Check-in enabled service from EGI or third party providers
- DataHub provides simple and scalable access to distributed data for Replay. User's DataHub spaces are automatically mounted and visible on Replay's interface, making it simple to perform analysis of available datasets. DataHub currently hosts more than 1,000 public and private datasets, reaching more than 1.3 PB.
- Software Distribution provides access to software based on CVMFS. Replay mounts selected CVMFS repositories on the environment for even simpler access to community-specific software.

With the recent introduction of the EOSC Data Transfer in the EOSC portal, new workflows for data analytics are enabled in Replay: users can trigger the transfer of available EOSC datasets to EGI infrastructure and then perform their reproducible analysis with Replay accessing the data.

This presentation will provide an overview of the EGI Replay service and how it can support the reproducibility of Open Science using data from the EOSC portal.

Primary authors: FERNÁNDEZ DEL CASTILLO, Enol (EGI Foundation); LA ROCCA, Giuseppe (EGI Foundation); SALAZAR, Xavier (EGI Foundation)

Presenter: SALAZAR, Xavier (EGI Foundation)

Session Classification: IBERGRID

Track Classification: Enabling and fostering Open Science adoption