



Contribution ID: 60

Type: **Presentation short (15 min)**

The ESCAPE Data Lake as the bridgehead for the EOSC

Tuesday, 19 October 2021 14:45 (15 minutes)

Experiments and scientists, whether in the process of designing and building up a data management system or managing multi-petabyte data historically, gather in the European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures (ESCAPE) project to address computing challenges by developing common solutions in the context of the EOSC.

A modular ecosystem of services and tools constitutes the ESCAPE Data Lake, which is exploited by flagship ESFRIs in Astro-particle Physics, Electromagnetic and Gravitational-Wave Astronomy, Particle Physics, and Nuclear Physics to pursue together the FAIR and open-access data principles.

This infrastructure fulfils the needs of the ESCAPE community in terms of data organisation, management, and access, and dedicated assessment exercises demonstrated its robustness.

As a result, collaborating sciences are choosing their reference implementations of the various technologies among the proposed solutions.

A variety of challenges and specific use cases boost ESCAPE to carefully take into account both user and infrastructure perspectives, and contributed to successfully conclude the pilot phase beyond expectations, embarking on a like-production prototype stage.

The ongoing phase of the project aims at consolidating the functionalities of the services, e.g. integrating token-based AuthN/Z or deploying a tailored content delivery and caching layer, and at simplifying the user experience. Specifically for this reason, a considerable effort is being devoted towards a DataLake-as-a-Service whose goal is to provide the end-user with a Notebook ready-to-be-used and fully integrated with the Data Lake.

ESCAPE milestones achieved during the length of the project represent a fundamental accomplishment under both sociological and computing model aspects for different scientific communities that should address upcoming data management and computing challenges in the next decade.

Speaker bio:

Riccardo Di Maria is leading the effort on novel data access technologies for future distributed storage infrastructures, referred to as the Data Lake, being prototyped in the ESCAPE European-funded project and leveraging Worldwide LHC Computing Grid (WLCG) technologies. ESCAPE aims to integrate research facilities of Astro-particle and Particle Physics, Electromagnetic and Gravitational-Wave Astronomy, and Nuclear Physics into a common data infrastructure in the context of the European Open Science Cloud (EOSC), pursuing the FAIR and open-access data principles.

Most suitable track

Delivering services and solutions

By submitting my abstract, I agree that my personal data is being stored in accordance to conference Privacy Policy

Primary author: DI MARIA, Riccardo (CERN)

Co-author: Mr DONA, Rizart (CERN)

Presenter: DI MARIA, Riccardo (CERN)

Session Classification: EOSC - Presentations

Track Classification: EOSC