European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures

ESCAPE
POSITION
STATEMENT

TABLE OF CONTENTS

European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures (ESCAPE) aims to address the Open Science challenges shared by ESFRI facilities (CTA, ELT, EST, FAIR, HL-LHC, KM3Net, SKA) as well as other pan-European research infrastructures (CERN, ESO, JIVE, EGO) in astronomy and particle physics.

PREAMBLE

This document summarizes the current views and expectations of astronomy and particle physics partners in ESCAPE about EOSC. It is written keeping in mind the various exchanges that are taking place with the EOSC secretariat, governance and working groups. It is also written to follow up the EC chaired EOSC stakeholders' concertation events aiming at gathering the scientific views on the implementation of EOSC.

The ESCAPE Executive Board addresses mainly three questions:

1. What is the research community expecting from EOSC?
2. What added value will EOSC bring to the research?
3. Which main issues/key messages should be considered by both the EOSC executive and governing boards?

WHAT IS THE RESEARCH COMMUNITY EXPECTING FROM EOSC?

1. From the ESCAPE perspective, EOSC will federate existing resources across national data centres, e-infrastructures, and research infrastructures, allowing researchers (and citizens) to access and re-use data produced by the ESFRI projects in Astronomy and Particle/Nuclear Physics for a multi-probe approach to understand the Universe; accelerating the discoveries and increasing scientific value by sharing data and by transferring knowledge within scientific communities.

2. The important volumes of data produced and managed by several of the ESFRI projects that are partners in ESCAPE imply investment in IC technology developments and related increase of costs for services, computing and storage resources to access their data. Therefore, ESCAPE partners aim to contribute to the provision of data services for the benefit of EOSC, while they expect that EOSC provides a sustainable environment in which data and data infrastructures are made interoperable and re-usable. Such a sustainability implies the establishment of a cooperative framework for open data, where research infrastructures, national and pan-European e-infrastructures co-develop and support together a dedicated infrastructure for data-research for the scientific-community-based needs in terms of an EOSC repository of open-source software for analysis, services, computing and storage resources.

3. The research-community-based foundation approach to build a global virtual research environment within EOSC for data interoperability and analytics is a way for ESCAPE partners to commit to building the EOSC and to contributing to the EOSC catalogues and portal. This would imply a combination of data produced from the facilities (namely the ESFRI research infrastructures encompassed within ESCAPE) and other data generated and shared by the concerned community. Therefore, a sharing of responsibilities for quality-certified scientific data is envisaged among the operators of the facilities and researchers in the longer term. ESCAPE partners would expect that EOSC defines and/or supports credit systems, acknowledgement, standard licences, and certification for the results of all researchers engaged in co-operative work that allow all to reap the benefits of open science. EOSC should be able to include and make sustainable the infrastructures and platforms that the community will deploy for such purposes.

4. EOSC is expected to establish a reference federated digital repository for the sustainability of scientific data preservation independently of the lifetime of each individual research infrastructure. Such a repository would leave the management of the responsibility for the respective data and the financing of their share of the repository with the originators so that a commensurate open access plan can be supported.

The following picture summarizes the ESCAPE Work Packages (WP) activities and expected delivered components of what one could call "a thematic cell" of the global EOSC. Such a "cell" in Astronomy and Particle/Nuclear Physics will enable EOSC to adopt transversally some services and e-infrastructures that will be useful also in support of other disciplinary "cells".

The added value that EOSC will bring to our community, through the ESCAPE commitments, is the formal opportunity to build upon the "cell" a "virtual research environment". There, researchers can upload their analyses in a notebook-like, reproducible and shareable style. Through this they will co-develop software and add mining to data, as well as running and improving workflows, in a real-time collaboration.

The following sections describe the main ESCAPE contributions to ESOC through its work programme
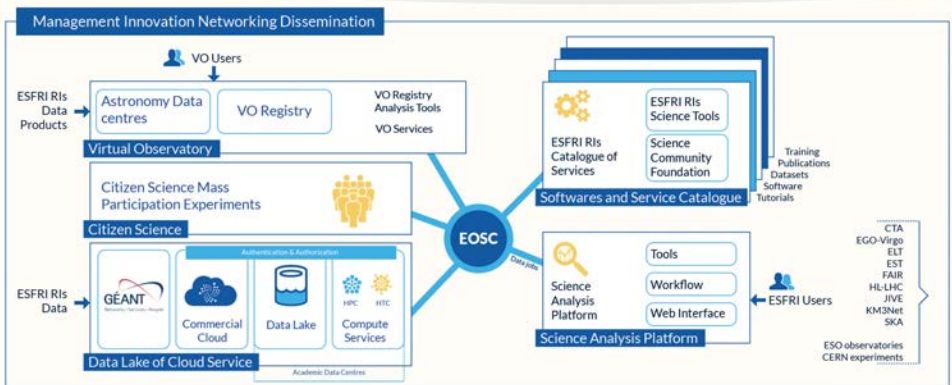
ESCAPE *Data Infrastructure for Open Science*

ESCAPE will design, implement, and operate a prototype data lake – a federated data infrastructure that will form the basis of an open access data service and science analysis environment for the ESFRI projects within the ESCAPE cluster covering astronomy and particle physics. It will propose such a solution as a key component of a future EOSC framework. The data lake concept enables the large, reliable, national research data centres to work together to build a robust cloud-like service to curate and serve data of CTA, FAIR, EGO-Virgo, HL-LHC, JIVE, KM3NeT and SKA at all scales up to the multi-Exabyte needs of such projects. These data centres (CERN, CNRS-CCIN2P3, DESY, GSI, IFAE-PIC, INFN, CNRS-LAPP-MUST, Nikhef, SURFSara, UG), partners in ESCAPE, have experience built up over a decade in particle physics within WLCG and in supporting major astronomy and astroparticle physics precursors pan-European research infrastructures (such as AMS, ANTARES, HESS, LOFAR, MAGIC, etc.).

Key outputs: *A mechanism required for large research infrastructures such as HL-LHC and SKA that in future will manage multi-Exabyte data sets, that will need to be served to global user communities in a scalable and performing way. A federated storage infrastructure implementing the FAIR data management principles at the base level, and form the basis for higher-level data preservation and access services delivered in the other work packages.*

Integration in EOSC. The Data Lake development leverages collaboration and integration of work and results from previous and ongoing frameworks (e.g. EOSC-hub). It will build on and integrate existing work from a variety of areas – the Research Infrastructures, previous EU projects, as well as using the current state of the art solutions in the appropriate areas and collaborating with ongoing work from GEANT, PRACE, and other proposed H2020 projects specifically addressing the European Open Science Cloud. The Data Lake will consist of an ecosystem of tools and services integrated into a reference implementation, while still providing, to the different science projects in ESCAPE, the flexibility to decide which ones to use.

ESCAPE Open-source scientific Software and Service Repository

ESCAPE deals with software services for open science data-analysis of the ESFRI facilities. The development of multi-messenger data analysis practices promotes activities for innovative methods, to maximise software re-use and co-development, to identify open standards for software release, to investigate data mining tools and new analysis technique. ESCAPE supports an open environment to guarantee cross-fertilisation and to develop community-specific data services that will be exposed under the EOSC catalogue of services under the FAIR principles.

Key outputs: *A sustainable open-access repository to share scientific software, digital libraries for data analysis, data-sets, data products including related user-support documentation, tutorials and training activities, which will be dynamically enhanced and maintained by ESCAPE ESFRI projects and included in the EOSC catalogue.*

Integration in EOSC. ESCAPE will make the software and service developed by the ESFRIs available to the scientific community via an open-source scientific software and service repository, which will be fully integrated into the EOSC environment. The services from the EOSC catalogue will be used where possible and thin-layer interfaces generated where needed. Central regulations compliant with the EOSC global ones will ensure the software and service quality via audits. Training activities and a help desk will provide support to the ESFRIs to devise regulations and adhere to them. At the same time, an EOSC-based approach allows for new innovative models of data exploitation, including data mining and deep learning techniques, making these accessible beyond the current expert communities. A competence group for innovations will be formed to steer these developments.

ESCAPE Connecting ESFRI projects to EOSC through VO framework

ESCAPE plans to: a) make the seamless connection of ESFRI and other astronomy and astroparticle research infrastructures to the EOSC through the Virtual Observatory (VO), b) refine and further pursue implementation of FAIR principles for astronomy data, and c) establish stewardship practices for adding value to the scientific content of ESFRI data archives. With the VO, Astronomy has built an operational interoperability infrastructure that has proven to be a great success for many aspects of astronomy data interoperability. The VO is an essential component of the astronomy data landscape, as has been strongly stressed in the ASTRONET Infrastructure Roadmap since its first publication in 2008. International astronomy data providers, in particular ground- and space-based telescopes, publish their data using the IVOA standards, and compliant scientific tools and services enable discovery, access and use of the data by the whole astronomy research community. Integrating the VO in EOSC implies the need to scale its framework to the biggest data sets that will be produced by the ESFRI and other projects.

Key outputs: *Assess and implement the connection of the ESFRI and other astronomy and astroparticle RIs to the EOSC through the Virtual Observatory framework, actively contributing to the setting up of the EOSC services and the provision of trusted data adhering to FAIR principles.*

Integration in EOSC. By leading the connection to the EOSC with the ESFRI facilities we will set the path for a new era of cross-disciplinary interoperability, and connections to the necessary computing resources. EOSC will facilitate the next step for the VO framework to realise its potential to scale to the biggest data sets that will be produced in particular by the ESFRI projects, and will enable use of VO data in scientific analysis platforms. ESCAPE aims to map the VO framework to the EOSC so that the VO enabled archive services from ESFRI will be interoperable. The integration of the VO registry and other standard maps will be key for discovery and reuse, access, deposition and sharing of data, as well as for data management curation and preservation.

ESCAPE will focus on defining and implementing a flexible science platform for the analysis of open access data available through the EOSC environment. These tasks will define and implement a platform that will allow EOSC researchers to identify and stage existing data collections for analysis, tap into a wide-range of software tools and packages developed by the ESFRIs, bring their own custom workflows to the platform, and take advantage of the underlying HPC and HTC computing infrastructure to execute those workflows.

Key outputs: *A platform-service for data analysis into EOSC and tailored to the requirements and the user needs of each of the ESFRI and other RI member of ESCAPE. It will be part of the EOSC catalogue.*

Integration in EOSC. Once data for analysis has been located and staged, and workflows have been defined, either by accessing the EOSC software repository or by the user directly, the next step is to deploy those workflows on the underlying processing infrastructure. For many of the involved ESFRIs and RIs, the data

scales involved require significant computational resources (storage and compute) to support additional processing and analysis. The EOSC-ESFRI science platform therefore must interface to an underlying HPC or HTC infrastructure. Consequently, it is important to make efficient use of the full performance potential of the HPC centres, e.g. by optimizing the access to file systems from the data processing layer and by ensuring the portability of science applications with container solutions. As with the data, this infrastructure is likely to be large, widely distributed geographically, and definitely heterogeneous. Deploying user-initiated processing and analysis tasks on this HPC infrastructure while simultaneously maintaining interactivity and responsiveness in the analysis system will be a challenge and requires a mixture of dynamic resource allocation and optimization. ESCAPE will draw upon the "data-lake" design and implementation and will build upon existing EOSC-hub activities such as Federated High Throughput Computing, Scientific Workflow Management and Orchestration, and EOSC-hub AAI.

ESCAPE *Engagement & Communication*

ESCAPE develops and manages a programme of crowdsourced data mining via Citizen Science mass participation experiments, with additional goals of public engagement and education in parallel.

Key outputs: *Involve citizens directly in knowledge discovery with ESCAPE and the ESFRI facilities, improving transparency of the scientific process. A harmonised suite of Citizen Science mass participation experiments and online video material, also deploying machine learning to accelerate volunteer classifications.*

Integration in EOSC. Central to ESCAPE's implementation of the Open Science Cloud vision is innovation for the society at large, by improving access to all ESCAPE results and through ESCAPE to the EOSC more widely. The ESCAPE EOSC "cell" prototype will allow researchers to stage existing data collections for analysis including citizen science.

WHICH MAIN ISSUES OR KEY MESSAGES SHOULD BE CONSIDERED BY BOTH THE EOSC EXECUTIVE & GOVERNING BOARDS?

ESCAPE as a large thematic cluster has gathered the formal commitment of ESFRI, RIs and pan-European research organizations in building a prototype of this cell that should be seamlessly integrated with the EOSC core services. The consequence of this is an increasing attention and involvement of scientists from our community to be part of the endeavor. Our success and our results depend on the sustainability and the acknowledgement of the cluster actions. The EOSC executive and governing boards aim at exploring and establishing the most pertinent and effective format for the EOSC implementation. However, the bottom-up approach where the researchers are central in the implementation choices for the effectiveness of open-data-research through EOSC is fundamental and must be pursued. A cross-disciplinary approach is critical to guaranteeing that EOSC will be inclusive. The organization, orchestration and accessibility of underpinning heterogeneous e-infrastructures for data archive, access and analysis are important structural EOSC issues.

However, EOSC should be more than that and should not be limited to a marketplace provision, but be a continuous evolving working space where researchers should routinely find out their daily dashboard for trusted data research and continuous networking opportunities within a large scientific community. ESCAPE partners commit to that and naturally advocate the cluster as a sustainable ecosystem as well as a successful approach for effectiveness and inclusiveness. The combination of partners in the project naturally leverages economies in supporting a joint repository and favors longevity of the approach. In addition, a subsidiarity principle could be applied towards the scientific communities; in order to leverage the work of ESCAPE and other clusters, the EU is invited to facilitate Member States to guarantee core and sustained funding for the cluster role in EOSC and to set an example that may be implemented in other research contexts and/or new focus.

projectescape.eu