



Project Title	European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructure
Project Acronym	ESCAPE
Grant Agreement No	824064
Instrument	Research and Innovation Action (RIA)
Topic	Connecting ESFRI infrastructures through Cluster projects (INFRA-EOSC-4-2018)
Start Date of Project	04.02.2019
Duration of Project	42 Months
Project Website	https://projectescape.eu/

ESCAPE DECEMBER 2021 - PROJECT BRIEF

Work Package	WP1, WP2, WP3, WP4, WP5, WP6
Lead Author (Org)	Kay GRAF (FAU), Mark ALLEN (CNRS-INSU), John SWINBANK (ASTRON), Stephen SERJEANT (OU), Xavier ESPINAL (CERN), Giovanni LAMANNA (CNRS-LAPP)
Contributing Author(s) (Org)	Julie CHAUDAN (CNRS-LAPP), Mathilde HUBERT (CNRS-LAPP)
Due Date	10.12.2021, M35
Date	22.11.2021
Version	1.4

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

Disclaimer

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 210506816.

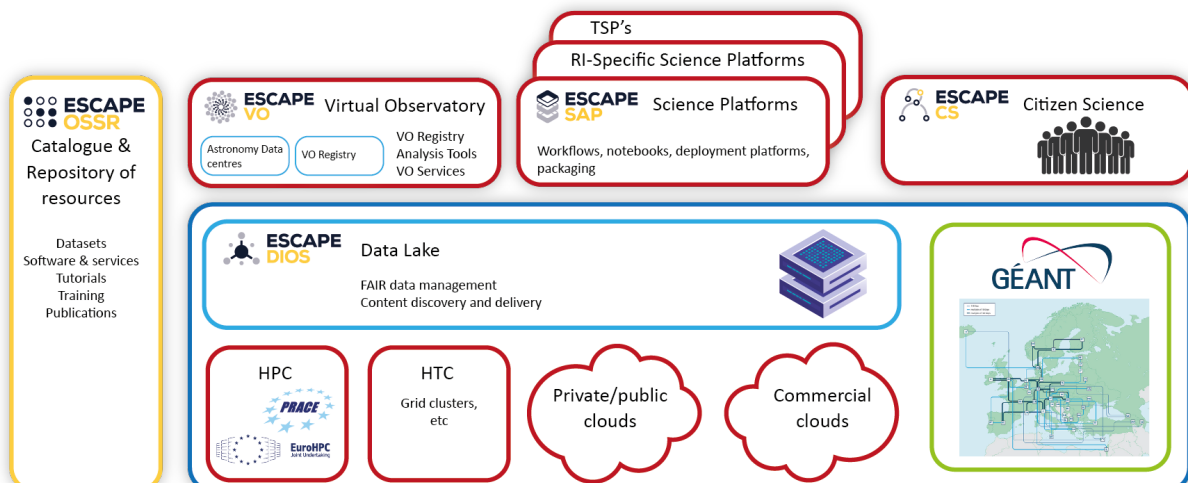
ESCAPE, a preamble

ESCAPE brings together a large fraction of the European research infrastructures (RI) in Astronomy, Astrophysics, Particle and Nuclear Physics.

These RIs are ESFRI facilities and landmarks such as CTA, ELT, EST, FAIR, HL-LHC, KM3Net and SKA as well as other pan-European research infrastructures such as CERN, ESO, JIV-ERIC and EGO. The ESCAPE project is supported and supervised by national universities and institutes of the European Union member states that are organized in thematic consortia such as ApPEC, ASTRONET, ECFA and NuPECC. The large professional community engaged in ESCAPE-related science extends to tens of thousands of scientists.

The partners in ESCAPE recognise the strong synergies and potential commonalities which are there at several levels. Many ESCAPE partners are today members of the EOSC Association. ESCAPE aims to address the Open Science challenges shared by its partners and the community. These challenges are technical, operational, sociological and scientific. Open Science allows scientific information, data and outputs to be more widely accessible and harnessed. In order to achieve the inclusiveness of the ESCAPE scientific domains in the global implementation of EOSC, a work programme is structured to deliver an “EOSC cell” where five technical Work Packages are in charge of the development and deployment of five corresponding pillars of such a cell for open data research.

- WP2: Data Lake for Open Science (DIOS).
- WP3: Open-source scientific Software and Service Repository (OSSR).
- WP4: Connecting ESFRI projects to EOSC through the Virtual Observatory framework (CEVO).
- WP5: ESFRI Science Analysis Platform (ESAP).
- WP6: Engagement, Communication and Citizen Science (ECO).



1) Integration with the EOSC infrastructure [1 page max]

ESCAPE aims to produce versatile solutions, with great potential for discovery, to support the implementation of EOSC thanks to open data management, cross-border and multi-disciplinary open environment.

ESCAPE WP2 Data Lake for Open Science (DIOS) works on a prototype Data Lake, a federated data infrastructure. The Data Lake prototype and its overarching orchestration and identity management layer is providing the foundations of what could be seen as an early EOSC cell. These proto-cells are environments where data services co-exist with analysis tools to provide researchers with the required experiment specific framework. These frameworks would contain the full experiment computing environment, including data browsability and data access, software and code repos. The integration of Data Lake tools on standard notebook technologies and analysis platforms opens the door to connect the scientific big data coming from the experiments with the diverse scientific community:

- Researchers log into these experiment specific EOSC cells and find the environment to start working with minimal configuration overheads.
- WP2 DLaaS (Data Lake as a Service) toolkit bridges experiment related Big Data with local site CPU and storage resources through capabilities for data browsing, download, upload and content delivery (caches) orchestration.

ESCAPE WP3 Open-source scientific Software and Service Repository (OSSR), is linked with the EOSC infrastructure directly via openAIRE, it collects software and services; as of 11/11/2021, 90 research outcomes have been exposed. OSSR is in close contact with EOSC-Future Onboarding in T2.2 and a registration as provider on the EOSC portal level will be finalised by end of 2022, a coordination and feedback round with EOSC Enhance took place.

ESCAPE WP4 is bringing the astronomical Virtual Observatory (VO) services into EOSC. The VO is the disciplinary interoperability framework which enables astronomers to find, access, interoperate and reuse the data they need for their research. The framework is defined at the international level by the International Virtual Observatory Alliance (IVOA) and is widely used by data providers worldwide. The approach being taken in Task 4.1, “Integration of astronomy VO data and services into the EOSC” is to map the IVOA Registry of Resources into the EUDAT B2FIND as the way of connecting to EOSC. This has been achieved, resulting in some 24459 IVOA resources in B2FIND. This has been possible because the IVOA Registry is built on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and includes the Dublin Core plus disciplinary extensions in its metadata schema, facilitating its inclusion in EUDAT B2FIND via the ESCAPE activities. This astronomy EOSC use case is highlighted in the SRIA v1, and also ESCAPE D4.4.

ESCAPE WP5 ESFRI Science Analysis Platform (ESAP) is implementing a platform for the analysis of open access data. No ESAP-related services have yet been integrated with EOSC. However, ESAP is being designed with the twin goals of modularity and adaptability, so that its integration into both at-scale EOSC systems and the data processing landscape associated with individual research infrastructures will be straightforward. The flexible and expandable ESAP system, when exposed through the EOSC, will facilitate re-use of existing ESFRI data collections and innovative analysis capabilities at large data rates and scales. Current ESAP development includes integration with the IVOA and specific research infrastructures and related projects including LOFAR, WSRT/Apertif, JIV-ERIC, Zooniverse, CTA, FAIR and EGO. Zooniverse in particular is intrinsically cross-cluster, covering SSHOC, EOSC-Life, ENVRI-FAIR, PANOSC, and ESCAPE.

All ESCAPE Services that build together a thematic EOSC-cell, along with certified FAIR archives of any digital object, will be accessible from a single-entry point: the ESCAPE Virtual Research Environment (VRE) running within the EOSC portal.



2) FAIR principles implementation and repositories [1 page max]

One of ESCAPE main goal is to provide data with FAIR principles thanks to scientific data interoperability and new methodological approaches for quality certified data and tools sharing.

Addressing FAIR principles is at the core of WP2 activities. The Data Lake model and tools should enable experiments to implement their Data Management plans. Providing the ability to humans and computers to **look for data and metadata** is at the core of RUCIO, the Data Lake's Data Management System. RUCIO connects the physical location of files at the sites (storage) with Logical File Names and associated metadata and/or file collections (datasets). Extended support for metadata is an ongoing activity, together with scientific communities and clusters. The possibilities to extend further browsability and searchability through metadata are huge for radio-astronomy, astronomy and cosmology. Data **accessibility** in the Data Lake is guaranteed via a common authentication and authorization method for users (ESCAPE IAM) and making use of well-defined standard protocols for data access (http and S3/SWIFT, and legacy protocols like xroot, gfal). The commonality of these protocols for data access (upload, download, list, search) ensures a common **interoperability** from the client side on different services and platforms used for: a) data management (including raw data recording), b) data analysis and workflows implementation and c) data processing in heterogeneous resources (batch systems, commercial clouds and HPCs). In activities and data challenges, we pursued these commonalities across experiments and sites providing resources, as a vision for further consolidation of the Data Lake, and the services and tools being prototyped. Perhaps one of the immediate consequences is being developed together with EOSC-Future project, which, besides the three principles mentioned before, will pursue the idea of analysis **reproducibility** by integrating workflows in a well-defined framework for data **reusability**, making use of existing technologies like REANA and RECAST, and well-defined repositories like Zenodo.

The main point of WP3, is to cross-fertilise across the partners the development of open scientific software and service under the FAIR paradigm. The objectives are to facilitate and support continuous development, deployment, exposure and preservation of partners' software/tools/services; foster interoperability, software re-use and cross-fertilisation between ESFRIs; offer an open innovation environment for open standards, common regulations and shared software. All the objectives follow the paradigm of enabling open science with software as "first class citizen" in the digital research products and a community-based, inclusive approach. The solution allows for a federation of community repositories. The harmonisation and standardisation are following in all WP3 tasks: 3.1: Management Activities, Policy and Support Action; 3.2: ESFRI Software and Services Collection; 3.3: Common Approaches: Software and Services; 3.4: Foundation of Competence for Software and Service Innovation; 3.5: Repository Implementation and Deployment. The building blocks of the OSSR are separable and should become part of the EOSC interoperability framework.

For WP4, the major task 4.2 identifies the ESFRI (and other RIs) needs for common standards to enable high level data products and archive services to be interoperable in the VO framework, so that they can connect to the EOSC through the VO. It supports the adoption of the VO framework and participation in the International Virtual Observatory Alliance (IVOA) for definition and updates of the necessary standards, in particular for inclusion of new communities. ESCAPE efforts in this area have promoted the use of common standards, to support FAIR across the cluster and also in the wider astronomy community. Results at an early phase in the project have been outlined in D4.2. Support is also provided to the astrophysics community in the scientific use of FAIR data via training events.

WP5 ESAP provides an extensible and consistent interface to access data from across a variety of ESFRI and other infrastructures. This includes providing access to standardized interoperable services provided by e.g. VO protocols, but also adapting individual ESFRI archives to a common format and exposing those through standardized interfaces (the ESAP GUI and REST APIs).



3) Technical, semantic, legal and organisational interoperability [1 page max]

ESCAPE goal is to enable interoperability between facilities, encourage cross-fertilization, and develop joint multiwavelength and multimessenger data processing and analysis capabilities

In WP2, Data Lake is providing the required capabilities to enable the implementation of Data Policies and Rules that can fit a variety of dimensions: from data protection making use of AAI methods (e.g. embargo data, protected data, read-only, etc.) to data geo-location rules (e.g. replicas only available in a site, region or continent or an arbitrary number of center locate world-wide). This ability is provided regardless of the underlying storage technology or the scientific apparatus or the scientific community, the goal is to have common interfaces, tools and protocols for resources providers and all scientific use cases. Initiatives outside ESCAPE have been initiated in order to expand our model to fellow science clusters and different disciplines. The implementation of data lifecycles is a good example and it has been a very successful demonstrator for many of the experiments during the data challenges we carried over in WP2, the usage of different Storage Quality of Service and access policies through the lifetime of a file (from data taking at the source to the final analysis by the scientists) is at the core of interoperability boosted with policies and rules.

In WP3, OSSR discusses and publishes common/best practices for software-related policies in the full software lifecycle; all the developments of the infrastructure follow common open standards (like CodeMeta metadata, open source library and API). One important issue is licensing of software and relevant practices have been collected and published to the OSSR landing page.

Discussions with other clusters, as well as national initiatives (Punch4NDFI in Germany, DMA Project of Helmholtz Association in Germany) are taking place.

In WP4, IVOA directly addresses all of the points specified above, perhaps with the exception of legal issues. IVOA contributes to the international Virtual Observatory initiative which supports global astronomy interoperability by the creation and maintenance of common open standards. IVOA Technical Interoperability is considered among the best practices as described in the EOSC Interoperability Framework document (page 11, doi: 10.2777/620649), and these standards should become part of the EOSC Technical Interoperability layer.



4) Stewardship of data [1 page max]

ESCAPE provides data with FAIR principles to increase researchers' efficiency but also deep training, education and capacity building programmes for the new generation of scientists and engineers that fully exploit ESFRI and EOSC facilities, to ensure the requirements and service features are properly understood and uptake.

The ESCAPE community is diverse with respect to the status of FAIR software and services, from these bases, OSSR has developed a work plan and focus groups that identify and address issues and find common solutions for the stewardship of FAIR open software.

A major aspect of the OSSR workbook is the training of software stewards, two schools are part of the projects' deliverables - one has happened with overwhelming success in Summer 2021.

WP4 addresses skills and techniques for data stewardship for ESCAPE partners in an annual Technology Forum, and also the wider community of data stewards with an up-coming "European Data provider Forum — Hands-on workshop for data providers".

Awareness of Data stewardship has been raised in the wider astrophysics community, for example at the main European Astronomical Society conference in a session co-organised with ESCAPE partners, on "Legacy and stewardship of astronomical archives towards multi-instrument, multi-wavelength and multi-messenger science." (<https://eas.unige.ch/EAS2021/session.jsp?id=SS14>). The data stewardship work of WP4 has also been highlighted at the Library and Information Services in Astronomy (LISA) conference (<https://lisa9.org>) with a paper on "Data stewardship of next-generation data products and services related to ESFRI and other astrophysics infrastructures" (<https://hal.archives-ouvertes.fr/hal-03372190v1>).



5) Cross-cluster collaboration activities and achievements [1 page max]

Because projects are not isolated islands, ESCAPE aims to develop synergies in the domain of open data management and data systems of the ESFRI projects, exploring standards, prototyping and deploying advanced open data services, towards the development of a consistent European research infrastructure ecosystem.

From the WP3 OSSR point of view ESCAPE participates in “Realising the Open Science Cloud: Towards a FAIR research data landscape for the social sciences, humanities and beyond” with all leaders of thematic discovery marketplaces:

- ENVRI-FAIR: Data/service catalogue
- ESCAPE: software and service repository
- EOSC-Life: Registries for software and workflows
- PANOSC: Federated Catalog service
- SSHOC: SSH Open Marketplace.

ESCAPE WP6 Engagement and Communication (ECO), has a cross-cluster coordination of both (a) communication and (b) citizen science activities with PANOSC, SSHOC, EOSC-Life, ENVRI-FAIR, FAIRsFAIR. Citizen science is intrinsically cross-domain so is a natural starting point for cross-cluster collaboration. Also, ECO has implemented ESAP Zooniverse access to crowdsourced data mining products throughout the cross-cluster domains including SSHOC, EOSC-Life, ENVRI-FAIR, PANOSC, and ESCAPE, with over 100 active projects and around 2 million volunteers.

ECO has been in charge of the coordination of EOSC crowdsourced data mining plans for EOSC-Future. The extension of ESCAPE milestone workshop on citizen science to the domains of other clusters, has resulted in two new citizen science projects under development in SSHOC and ENVRI-FAIR areas.

ESCAPE has participated in joint face-to-face communications meeting in abeyance since the pandemic, but resulting virtual presentations at EOSC Symposium 2021, Open Science Fair 2021 (in collaboration with REINFORCE H2020 project) and other meetings.

At Open Science Fair, a debate was organised on how to increase the subject knowledge of non-subject-specialists in EOSC to be published (in preparation).

Originally part of the ESCAPE work programme and included in the new H2020 EOSC Future project, the ESCAPE (Test) Science Projects (SP) were proposed to validate ESCAPE services at the end of the project. SPs are proposed to demonstrate multi-domain science integration across ESCAPE, new cutting-edge open science capabilities, making use of the services implemented within ESCAPE. While the SPs would provide feedback on the capabilities delivered by the ESCAPE EOSC-cell, will benefit real science goals in exploring synergies between the ESFRIs and largely among three scientific communities Astrophysics/Astroparticle, accelerator-based Particle and Nuclear Physics (supported by consortia of EU member states research agencies and institutes within JENAA). This is also the result of a strong cross-cluster coordination and a multidisciplinary shared action. The Science Projects are strengthening the opportunities delivered by EOSC in producing excellent cross-domain science and federating cross-cluster initiatives for innovation and education.

The work-programme of ESCAPE is evolving with the support of the concerned community and with a strong inter-clusters dialogue. Together, the five science clusters delivered a second co-signed position statement document with a formal explanation of the urgent need of European Commission to support a longer-term role of the five science clusters. The science clusters together aim at providing content to the EOSC, enhancing researchers’ involvement in Open Science and suggesting potential cooperative pathways in the Horizon-Europe framework and along with the EOSC Association roadmap.

