ESCAPE Science Projects for EOSC-Future (WP6.3)

"Demonstrating EOSC Value through Cross-domain Research Science Projects"

20 July 2021

Authors:
Ian Bird – LAPP-CNRS (ESCAPE).
Cristina Doglioni –U. Lund (member of ATLAS/HL-LHC)
Elena Cuoco – Scuola Normale, Pisa (member of Virgo/EGO)

Distribution:
WP6, TCB

ESCAPE Science Projects

- **Science Cluster, acronym of the SP:**
  - ESCAPE Dark Matter (DM), Extreme Universe (EU)

- **Participating legal entities & RIs**:
  - CERN, CNRS(LAPP, CPPM, CDS), FAU, NWO(ASTRON, Nikhef), INFN, Scuola Normale Pisa, Uni Lund (via LAPP), Uni Amsterdam.
  - HL-LHC(ATLAS, CMS), SKA, KM3Net, CTA, Darkside, Virgo, LOFAR, LSST, Einstein Telescope, Theory, FAIR

Description

Introduction

The ESCAPE project is building a virtual research environment (VRE) for the Astronomy, Astro-Particle, Particle, and Nuclear Physics communities, as a prototype of the European Open Science Cloud (EOSC). It is intended that the developments of tools and services made in ESCAPE will eventually be contributed to the EOSC implementation. The mechanism for that migration will be the EOSC-Future project, within which the 5 science cluster projects (including ESCAPE) will be the science drivers.

Within ESCAPE it has been agreed that two large "Science Projects" (SP) will be deployed with a number of high-level objectives:

- To demonstrate new cutting-edge science capabilities, in particular those involving inter-RI collaboration and science outcomes;
- To validate on behalf of the science communities, that the software, tools, services, and infra-structure developed within ESCAPE are what is required by the science use cases;
- To provide feedback to the ESCAPE project, and ultimately to the EOSC community, that will help guide the future direction and development of the EOSC.

Thus, while the intention is to do real cutting-edge science, it must be *science within the context of EOSC*, and so equal weight must be given to both scientific activities and to infrastructure and tool development and validation. This may make the science cases harder to implement initially, but the longer term benefits of these synergies will be considerable.

The ESCAPE objectives are supported by the thematic consortia (made up from the national funding agencies) - ECFA[1], APPEC[2], ASTRONET[3], NuPECC[4], and the collaboration of those bodies within JENAA[5]. The European Strategy for Particle Physics update in 2020 made it clear that at the highest level, synergies between these research infrastructures should be encouraged and strengthened, and ESCAPE is a visible mechanism for that. JENAA supports the development of this new generation of synergistic VREs.

Goals of the SPs

The two SPs are indeed guided by Expressions of Interest at the level of these coordinated consortia (discussed above). The two science projects selected are:

[1] ECFA: https://www.ecfa.org/
[2] APPEC: https://www.appec.org/
[3] ASTRONET: https://www.astronet-eu.org/
[4] NuPECC: http://www.nupecc.org/
[5] JENAA: http://nupecc.org/jenaa/?display=about

Dark Matter (DM):
The Dark Matter project recognises that many of the RIs within ESCAPE have experiments that are searching for Dark Matter. There is a clear complementarity between these experiments under a variety of dark matter hypotheses.

The presence of DM in astrophysical observations, combined with the absence of clues for DM particles in experiments, indicates that if DM has interactions with ordinary matter they must be very feeble, and produce subtle experimental signals. Connecting results and potential discoveries from different experiments requires the engagement of all scientific communities involved - astrophysics, particle physics and nuclear physics - as already recommended within the update of the European Strategy of Particle Physics[6].

Besides the interpretation of results in terms of dark matter theories, synergies also exist between different communities and experiments in the tools needed to produce those results, in particular in terms of data management, data analysis and computing. This is one of the keystones of the Dark Matter Science Project within the European Science Cluster of Astronomy and Particle physics ESFRI research infrastructures (ESCAPE) project.

In particular, this Science Project aims at:

1. further understanding the nature of DM by performing new analyses within the experiments involved, and collecting all the digital objects related to those analyses (data, metadata and software) on a broad platform that will be ultimately hosted on the EOSC Portal and will allow these analyses to be reproducible within the various collaborations and by the entire community wherever possible;
2. exploiting synergies and complementarities across different communities, as the final output of each workflow will be individual experimental curves that can be interpreted in terms of dark matter particle properties and displayed on the same plots summarizing the complementarity;
3. enhancing the research participation in the EOSC, by providing a working example of open science that has started from a bottom-up effort by different experiments.

The **scientific added value** for this SP is provided by new dark matter analyses coming from the complementary experiments involved (particle colliders, direct and indirection experiments) as well as from theory and observational constraints, all interpreted within the same theoretical framework and displayed in a summary plot that showcases their synergies. During the analysis design, we will identify innovative algorithms (e.g. machine learning, but also procedures to reconstruct images to distinguish signal and background) that can be individually highlighted and shared for use by other scientific communities and / or in society.

The **Open Science added value** for this SP is that all the digital objects within these new DM analyses will be implemented within the ESCAPE services infrastructure. We will make use of the ESCAPE Data Infrastructure for Open Science in the European Open Science Cloud to store, distribute and provide data and software access to the broad dark matter scientific community. This is a unique link between DM as a fundamental science question and the Open Science services needed to answer it that benefits the scientific community as a whole.

The DM analyses within this SP will rely on the ESCAPE services infrastructure within EOSC-Future to see their experimental data, simulations and software procedures developed within sustainable analysis pipelines and converging into a bigger picture to constrain or discover dark matter.

Related detailed docs:
https://docs.google.com/document/d/1IeJgOzCxZvKp0IG9382eXhQ3qf-KUYYwKvKFl35Oj1Y/edit#heading=h.i7g3bsi7rwc4

[6] ESPP: https://europeanstrategy.cern/home

Project descriptions:
https://docs.google.com/presentation/d/1VBRvDGsukncYQKdx_TD5np7MALjAYoKIp_rDnIXPdw/edit#slide=id.gaf347900df_1_0

Extreme Universe and Gravitational Waves (EU):
The Extreme Universe project intends to develop a sustainable platform within which to enable multi-messenger/multi-probe astronomy (MMA). There are many studies of transient astrophysical phenomena that benefit from the combined use of many instruments at different wavelengths and different probe types. Many of these are based on the trigger of one instrument generating follow-ups from others at different timescales from seconds to days. The intention within ESCAPE is to build such a platform for MMA science in such a way as to make it sustainable.

ESCAPE took inspiration from the pioneering first observation in 2017 of a gravitational wave source, GW170817, from the merger of two neutron stars, and importantly by the rapid follow-up by instruments from astrophysics, cosmology and nuclear physics of this historical event. Another recent successful example of linking together existing telescopes is the Event Horizon Telescope Collaboration (EHT) that leverages considerable global investment and has been awarded a number of prestigious titles for its ground-breaking results in making the first-ever image of a black hole in the galaxy M87. Such observations could lead to images of strong gravitational effects that are expected near a black hole. Extreme energetic astrophysical pulsing phenomena such as GRB, AGN, FRB are also high-energy phenomena not yet fully understood. A data sharing and open-science approach are key to adding knowledge and progressing towards an understanding.

A holistic approach to black holes and exploiting gravitational waves for fundamental physics are the main guidelines of the “extreme universe” (EU) SP. The ‘frontier’ for multi-messenger science is to understand extreme matter and particle processes in strongly curved spacetime. The ESCAPE SP (also to be accessible from the EOSC portal) would implement an integrated platform for Multi-Messenger Astronomy where data from different wavelengths/messengers can be easily gathered, analyzed and modelled holistically, and not remain fragmented as at present.

The **scientific added value** of this project is to set up a sustainable platform for multi-messenger astronomy that is available for publishing and following up on interesting “trigger” events on a range of timescales, as well as a data analysis platform for MMA.

The **Open Science added value** of this project includes the possibility to perform “realistic” studies from the web driven ESCAPE/EOSC infrastructure which will lower the bar for access to scientific data and tools towards the utilization of the open data as made available by the ESCAPE collaborators. The technologies applied in the Extreme Universe SP datasets are general, and as such the SP is a showcase of the potential of Open Science research using EOSC tools.

Related detailed docs:
https://docs.google.com/document/d/1R29rLNEHGfbhmEcQko2KtU_EK4QSyjpTHo_knHPXO8w/edit#heading=h.ls4nxlwbbldn

SPs in ESCAPE and EOSC-Future
ESCAPE also intends to investigate a coherent bridge between these two SPs. For example,

- through Gravitational Waves that may probe the particle nature of dark matter, e.g. via Black Hole environments,
- the properties of extreme nuclear matter via neutron star mergers, and
- new physics in the early universe via phase transitions.

Within EOSC-Future, ESCAPE aims to integrate the SPs in a proto-EOSC environment where particle physicists, astrophysicists and astronomers, experimentalists as well as theorists, with a wide range of interests, can exploit synergies and complementarities across different communities.

This approach has a unique ambition of developing a multi-probe and cross-domain Open Science Cloud environment. One of the objectives for ESCAPE within EOSC-Future is to develop a permanent platform where the different communities can identify cross-fertilization opportunities for mutual benefit, with an even broader perspective of the complementary set of experimental searches, astrophysical/cosmological observations and theoretical benchmarks.

As well as supporting the continuation and consolidation of the key ESCAPE deliverables (data and software catalogues, analysis platforms) the inclusion of existing cooperative frameworks such as the Virtual Observatory, the CERN OpenData portal and ESO platforms, would further strengthen such a platform for the long-term.

Technical objectives

The sustainable, long-term management, curation, comparison and scientific exploitation of data of the next generation ESFRI facilities are key objectives of the ESCAPE approach. The aim is to maximize the exposure to multi-messenger and multi-probe data from astronomy and accelerator-based particle and nuclear physics for the open science challenges of a new generation of researchers.

The intention of these projects is to make use of the building blocks provided by the ESCAPE project work packages initially, and to build upon that with developments and infrastructure integration anticipated in EOSC-Future.

Since the two SPs do not neatly sit within a single RI, the question of how resources will be made available to support them is important (and will be important in the future as EOSC encourages and enables cross-RI and cross-discipline collaboration). To that end, via EOSC-Future the SPs should try and have access to the resources provided by the INFRAEOSC-07 projects (EGI-ACE, OPENAIRE-Nexus), that are tasked with that provision. In addition there is the procurement funding within EOSC-Future (WP8) that can also be a part of the resource provision.

Building blocks: ESCAPE Services Required for the SPs

The 2 SPs will make use of a variety of services developed in ESCAPE, to demonstrate that these enable the required scientific functionality. Within EOSC-Future we will deploy these services within the EOSC context to ensure that the ESCAPE “EOSC cell” is fully compatible with the long term EOSC environment.

The following lists the services developed by ESCAPE that are essential, although these can be treated as independent and we can incrementally include these into the overall science environment.

1. AAI (ESCAPE-WP2): A fully developed AAI solution following the AARC blueprint is fundamental. In EOSC-Future we must ensure that the ESCAPE solution is fully interoperable with EOSC. Scientists in the SP’s should be using a single user identity for all aspects of work. Authorization services must enable key access controls to various data sets and resources. We need to plan how the integration of ESCAPE AAI with EOSC-Future will be managed.
2. Data Lake (ESCAPE-WP2): federated storage services should be made available to the SPs, allowing all of the data sets required to be openly accessible to all participants. Because some of the data sets may be subject to embargo (with permission to use for the SP), the Authorization mechanism must ensure this. The full set of tools that implement the DL should be available and used:

a. Rucio - data location catalogue and policy engine: This may need a specific set up, the important point is that users must be able to select data across multiple experiments and infrastructures that may have their own Rucio catalogue.
b. ESCAPE FTS (File Transfer Service) - key for moving data around; integrated with the AAI service and the storage endpoints.
c. Caching and streaming services to deliver data to processing and analysis. These are services operational in ESCAPE; once AAI services are integrated they should be operational in the EOSC environment.

3. (ESFRIs, ESCAPE-WP4): Publication of data sets into the DL - required from the SP partners of the ESFRIs and WP4.
4. ESCAPE (ESCAPE-WP3) software catalogue should publish all of the needed analysis components, and make them available for the various groups involved in the SP work.
5. (ESCAPE-WP5) An analysis environment, with a Jupyter notebook deployment, and access to scalable compute resources behind.
   - RECAST, REANA from CERN; tools that enable reproducible analyses, essential in making the scientific outcomes FAIR
6. Virtual Research Environment for each of the SP as the outcome of the integration of the above together with publication services (WP3) for the scientific results and outputs of the work.

ESCAPE intends to make all of these services and tools available to other RI's or service providers in EOSC via EOSC-Future.

Work on the SPs would also be a good opportunity to further engage as ESCAPE (and EOSC-Future) with the HEP Software Foundation[7].

Building Blocks: Resource requirements

In addition to the above services, the science projects will also require some dedicated resources. This scenario already exposes the problem of doing cross-RI science, since none of them may feel directly responsible to provide the resources. In all cases it is important that the resources are integrated and accessible via a transparent (to the end-user scientist) mechanism which requires no special requests. We foresee the following requirements:

- Storage: sufficient federated storage with the DL for all of the data sets.
  - Estimate: needed by SP and group?
  - Some should be provided through EOSC-Future provisioning, and integrated with our Data Lake

- Compute resources: we should foresee provisioning from the following sources:
  - Partner owned resources, at ESCAPE computing centres.
  - Cloud resources -
    - Initially funded through ESCAPE set-aside funds;
    - in a second step, funded via the EOSC-Future procurement mechanisms, e.g. via the INFRAE0SC-07 projects
  - HPC resources, where useful, for AI/ML and similar appropriate activities. Here we will use the EOSC-Future/FENIX(EuroHPC) collaboration brokered via the EC as the mechanism. This is under active discussion and will be followed up.
- Networking: while this is unlikely to be an issue, we will continue to coordinate with GEANT who are heavily involved in EOSC-Future to ensure this is not a problem.

[7] HSF: https://hepsoftwarefoundation.org/

Implementation

Plan of Work

The ESCAPE plan of work for the SPs is given in the attached spreadsheet. The objectives and milestones are aligned with the EOSC-Future milestones at months 6, 18 and 30.

Workplan and milestones (evolving): https://docs.google.com/spreadsheets/d/1jij_tHH-z5Kwkw0ELu6xsVFkUj1iyP0GeXHcRcQ6tzc/edit?userstoinvite=elena.cuoco@gmail.com&ts=60ed814d&actionButton=1#gid=0

(snapshot below)

Scope	Aspect	M1-6	M6	M7-18	M18	M19-30	M30	
Project (Proposal milestones)	**Hiring**	Hiring of post-docs						
	SP Deployment		Preparation phase of SPs: cluster SPs started integration and adaptation of cluster/Research Infrastructure tools and services into the broader EOSC framework, including integration with AAI, data sources moved into EOSC (FAIR) data stores, and catalogued, software development and exchange platforms available, workflow deployment mechanisms adapted.		Initial deployment of SPs done: Most of the SPs have fully operational workflows using integrated EOSC services and tools, make use of EOSC Interoperability Framework, provide feedback from the science communities to the service deployment and operation, and can use resources made available through the clusters/Research Infrastructures.		Full scale operation of SPs to the point where many have full scientific analyses ready or close to publication as full demonstrations of open cross-disciplinary science. Demonstrations of the full lifecycle of data processing, storage, analysis and publishing supported by resources available and transparently integrated through EOSC. Workflows deployed across cluster/Research Infrastructure resources and where appropriate on commercial cloud and/or European HPC resources.	
	Clusters' Input to EOSC Horizontal Services		List of candidate services from clusters which can be generalised to be EOSC-wide offerings, with several already being developed to be EOSC-ified.		EOSC software catalogue/repository populated with cluster-provided services and tools.		General availability of Research Infrastructure-originated horizontal services (as appropriate) visible through EOSC Portal and catalogues.	
ESCAPE (services deployed for the TSPs)	**AAI**		AAI federated with EOSC-F; tools use hybrid of legacy & token based		DL tools, OSSR, VO, using federated AAI		Full workflows with EOSC AAI	
	Catalogue		OSSR 1st deployment, links to sw, datasets		OSSR accessible from EOSC portal		Full SP project lifecycle visible in OSSR	
	Data Lake		Pilot Data Lake in place, 5-10 sites ready				Full scale DL for both SPs	
					Demonstrate open vs embargoed vs limited access data in DL			
	Resources		Define resources needed for TSPs		Provision scalable resources from 07 projects, clouds		Workflows using all needed resources, HTC, HPC, AI, etc. with transparent provisioning	
			Put in place computing cloud to support TSPs, initially partner contributions		Access to HPC resources for some use cases, e.g. ML			
	VRE				Scalable notebook service in place		Full ESCAPE VRE in place	
	Citizen science						Portal access to CS projects and datasets	
	Open science						Public data releases	
	Horizontal Services		Initial list of candidate services, and SP tools		Visible via OSSR through EOSC portal		(EOSC-Future makes horizontal services available broadly)	
Dark Matter	**VRE**		Base use cases using data sets on DL		Analysis environment in place, supporting full DM workflows		Paper publications and open sw/data release	
					Using notebook service, REANA, Recast			
			Science groups build workflows to support individual DM projects		The software (e.g. reconstruction software, analysis pipelines) will be stored in the ESCAPE Software Catalogue.		Where applicable, the analyses involved in the TSP will be also linked to the ESCAPE Virtual Observatory and Citizen Science Platform, and to outreach material to be distributed at various levels (e.g. high school Masterclasses).	
Extreme Universe	**VRE**		Base use cases using data sets on DL		Analysis environment in place, supporting full EU workflows		Visualisation tools and software libraries for applications of (IVOA standard) space-time indexing of astronomical data and manipulation of complex sky regions (e.g. GW credible areas) (from ESCAPE WP4)	
					connected to Virtual Observatory registry to enable multi-wavelength data discovery		The possibility to perform "realistic" studies from the web driven ESCAPE infrastructure	
							Paper publications and open sw/data release	