ARPHA Preprints

PENSOFT

Grant Proposal

Author-formatted document posted on 05/12/2023

Published in a RIO article collection by decision of the collection editors.

DOI: https://doi.org/10.3897/arphapreprints.e116673

ESCAPE Dark Matter Science Project for EOSC Future project (WP6.3)

Pooja Bhattacharjee, Ian Bird, Francesca Calore, Caterina Doglioni, Christopher Eckner, Elena Gazzarrini, Lukas Heinrich, Tetiana Hryn'ova, Valerio Ippolito, Jared Little, Stephen Serjeant, Mikhail Smirnov

ARPHA Preprints *Author-formatted document posted on 05/12/2023*. DOI: https://doi.org/10.3897/arphapreprints.e116673

Dark Matter, the ESCAPE Science Project for EOSC Future Project

Authors: Caterina, D., HEINRICH, L., Calore, F., Bird, I., Ippolito,V., Bhattacharjee, P., Eckner, C., Little, J., Gallen,A., Ekman, A., Smirnov, M.

Abstract

The European Open Science Cloud (EOSC) forms an ecosystem of research data and related services that enable and enhance access and reuse of FAIR research outputs (i.e., data and other digital objects). The test science projects (TSPs) in EOSC Future will serve as examples of how joint projects can address major, if not global, challenges for Europe's societies and how research infrastructures can align to support Horizon Europe's missions within the EOSC. The five Science Clusters ENVRI-FAIR, ESCAPE, EOSC-Life, PANOSC and SSHOC have designed a set of TSPs to drive integration of research infrastructure data and services ("composability") between the science domains and, as they mature and become ready for production, demonstrate how services from EOSC-Core can be incorporated into the routine provisioning of research infrastructure services. As one key output from the science clusters towards EOSC, they will contribute thematic services to the EOSC Ecosystem which will be aggregated at the cluster level, reflect the needs, requirements and considerable digital assets and services of the associated research communities and mobilise the research communities for widespread use of EOSC resources. This paper present Science Project: Dark Matter.

Description:

i. Existing situation: (*e.g. what is the state of the art of the topic right now; this will need to be compared to the Impact sub-section below*)
The Dark Matter project recognises that many of the Research Infrastructures 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. 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 Test Science Project (TSP) within the European Science Cluster of Astronomy and Particle physics ESFRI research infrastructures (ESCAPE) project.

ii. Objectives:
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). Within ESCAPE it has been agreed that two large “Test Science Projects” (TSP) will be deployed with a number of high-level objectives:

ii.i. To demonstrate new cutting-edge science capabilities, in particular those involving inter-RI collaboration and science outcomes;

ii.ii. To validate on behalf of the science communities, that the software, tools, services, and infrastructure developed within ESCAPE are what is required by the science use cases;

ii.iii. 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, APPEC, ASTRONET, NuPECC, and the collaboration of those bodies within JENAA. 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.

iii. Compliance to criteria: (*these criteria come directly from the proposal last version*)

- Eligibility: (*SPs are open to participation by research communities (RPOs: universities and research institutes) and not-for-profit providers*)
  This SP it encompasses research groups from major dark matter experiments, but it's also open to scientists from legal entities which are not members of the EOSC Future Consortium – this is part of the nature of large particle physics experiments.
- Contribution to EOSC: (*e.g. cross-domain and composability features, contributing to EOSC; the feasibility of integrating the proposed services into the EOSC ecosystem; the type of services these proposals ask from EOSC (e.g. security, monitoring, AAI)*)
- 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.
- Quality: (*e.g. the quality of the proposed research, from the hypotheses to be tested all the way to the interpretation of the results; the quality of the proposed incoming services, based on elements like standardisation, protocols, best practices, efficient use of resources, semantics, data access (FAIR-compliance) the anticipated added value of the proposed services and their workflows, beyond those making the original request*)
  FAIR-compliant datasets will be used in this SP, and it will contribute to progress within the field. For example, CERN (one of the major RIs in this project) has recently approved its new Open Science Policy, and scientists in this SP are working within their experiments to make their datasets FAIR. This SP is also at the forefront of the discussion of sustainable software and reproducible analysis: in fields like particle physics, it is not enough to make data FAIR, it is also necessary to provide users with robust tools for analysing this data.
- Relevance: (*there will be a preference for service catalogues that address the EU focus areas for research and societal demands, including infectious diseases, the European Green Deal, circular economy, Data Spaces and EC Missions like climate change, health, agriculture; the user communities they bring and attract; gender issues; potential for Open Science and open innovation*):
  The **scientific added value** for this TSP 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 TSP 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 TSP 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.

Implementation

i. Plan of work: (Tasks, Milestones, Responsible entities/persons, etc.)

Months 1-6

Organise recruitment, define datasets, resources and algorithms

Months 6-12

Focus for postdocs working on SP is new dark matter analyses
Data lake is operational and usable
First version of the VRE (JupyterHub) with implementations

First REANA implementation usable
Discussion on presentation of results integrated (and first draft available) within the Snowmass project, whitepapers to be delivered as outcomes of the project

Months 12-24

Focus for postdocs working on SP is shared between finalizing dark matter analyses and implementation
Focus for VRE postdocs is integration with EOSC and Core Services

Months 24-30

Consolidation and documentation
Combination of results
Dissemination

Details on 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 this and the other ESCAPE SP 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 TSPs 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 TSPs**
- The ESCAPE TSPs 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.
  - AAI (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 TSP’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.
  - Data Lake (WP2): federated storage services should be made available to the TSPs, 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 TSP), the Authorization mechanism must ensure this. The full set of tools that implement the DL should be available and used
  - 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.
  - ESCAPE FTS (File Transfer Service) - key for moving data around; integrated with the AAI service and the storage endpoints.
  - 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.
  - (ESFRIs, WP4): Publication of data sets into the DL - required from the TSP partners of the ESFRIs and WP4.
  - ESCAPE (WP3) software catalogue should publish all of the needed analysis components, and make them available for the various groups involved in the TSP work.
  - (WP5) An analysis environment, with a Jupyter notebook deployment, and access to scalable compute resources behind.
  - RECAST, REANA from CERN
  - Virtual Research Environment for each of the TSP 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.

ii. Use of resources (*data, services, etc.*): Open-access data-sets provided by RI partners in ESCAPE, as well as from other relevant RIs and data repositories (e.g. those referred to above). Open code web services delivered by all partners, which have to do with the data access, their management and their analysis.

iii. What are the demands of the SP from EOSC Future platform (*e.g. Vertical and Horizontal composability services, HPC power, storage volume, etc.*): The DM SP requires the following EOSC services to be integrated and available:

- Federation of the ESCAPE AAI (IAM) service with the EOSC AAI, such that authorised users with ESCAPE credentials have access to all EOSC-Future services and resources without additional authentication steps;
- Use of the EOSC helpdesk service to support the work of the ESCAPE SP team and interactions with EOSC infrastructure support;
- EOSC core monitoring to provide information related to ESCAPE use of EOSC services;
- EOSC core accounting to provide information related to ESCAPE use of EOSC resources;
- Availability of cloud storage and cloud compute, provisioned through EOSC-Future and associated EOSC resource projects. The storage resources will be attached to the ESCAPE data lake federated storage service, and the compute resource will be used for processing of the SP workflows.
  - Storage: several TB as proof-of-concept for integration; eventual need 100 TB; provisioned as cloud storage (e.g. S3 or Swift)
  - Compute: cluster of 500 cores initially, capable of hosting a Kubernetes cluster; eventual need scaling to 2000 cores;
  - Transparent access to HPC resources for AI/ML training tasks.

iv. What the SP brings to EOSC Future platform: 1) engagement of ESCAPE and general High Energy Physics community; 2) complex data analysis use case, of major interest for science, as well as for public; 3) infrastructure and workflow for performing complex data analysis on the EOSC (Virtual Research Environment), including off-the-shelf ML algorithms that are usable by others.

v. Partners: **CNRS, INFN**, **FAU**, **CERN**

vi. Effort (PMs by partner) *(straight from the proposal budget; please include the legal entities which are not yet members of the Consortium*): **CNRS** - 59 PM; **INFN** - 10 PM; **FAU** – 8.5 PM, **CERN** – 34 PM. In-kind contribution by University of Manchester, TUM.

Impact

i. Strategic: (*where EOSC platform will be after the end of the SPs (e.g. what will be the state of the art of the topic after the end of the SP*)
This Science Project will bring:

i.i. further understanding of 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;

i.ii. leveraging 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;

i.iii. enhanced research participation in the EOSC, by providing a working example of open science that has started from a bottom-up effort by different experiments.

ii. Scientific/User communities: This SP brings together a number of scientific communities that have been sharing results at conferences so far, but have not yet shared workflows, in a way that helps them fully understand each other's work. Whether the quest for dark matter leads to a discovery or to further constraints, FAIR datasets as well as sustainable and interoperable software are the key for these different communities to work in synergy and make progress.

iii. Societal/Economic: From the societal point of view, the topic of dark matter, which constitutes 85% of the matter in the universe, holds a major role in the understanding of our universe. A discovery of the nature of dark matter would constitute a scientific revolution.

iv. EU Policies: No EU policies related to this topic.

v. Other: Cultural. Mentioned already in the societal impact above.

Engagement Plan

I. Target groups: Researchers/engineers working on dark matter; Academics, researchers and students; Citizens.

II. SP key concept:
One overarching objective of science is to further increase our understanding of the universe and its composition. The nature of dark matter (DM), corresponding to 85% of the matter currently present in the universe is still unknown. The presence and distribution of DM is detected through its gravitational interactions by observatories and experiments, while the interactions of DM with ordinary matter particles can be observed indirectly and directly in astrophysics experiments. These interactions also allow for DM to be produced in collisions of ordinary matter and observed in experiments at colliders and at particle accelerators. These experiments provide complementary information about dark matter - ordinary matter interactions. Data from this wealth of astrophysics and particle physics experiments, represented in terms of direct, indirect detection and collider experiments in this SP, combined with theoretical models and interpretations, will shed new light on dark matter.
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. 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 this SP. The DM analyses within this TSP will rely on the ESCAPE services infrastructure to have experimental data, simulations and software procedures developed within sustainable analysis pipelines.

III. Dissemination measures (see links for details)

- Talk & proceedings at the TOOLS20 conference (review of state-of-the-art tools for high energy physics and cosmology
- Engagement via Consortia (ECFA, APPEC, NUPECC, JENAS)
    - DM is a recognised challenge for the communities
    - Invited poster by DM postdoc at the JENAS conference 2022
- Talk and upcoming paper from the FAIR4HEP workshop, to advance data & artificial intelligence (AI) models through the development of FAIR frameworks
- Inputs to "Snowmass" whitepapers submitted to arXiv and journals [1][2][3][4] (Snowmass: input to prioritization of particle physics, deciding US's science priorities)
    1. Figures made with code that is / will be onboarded to EOSC

Near / longer term future:

- Individual peer-reviewed papers
- International physics and computing conferences to present scientific and workflow results
- Wishlist: Citizen Science experiments & Events for the general public, *learning from the ESCAPE experience*

Table 1. List of basic services ready to be integrated in EOSC Future

No	***Name of organization providing the service/scientific product***	***Provide a high-level description of what the service/science products involved required in the SP do in terms of functionalities they provide to the end users, and a link to relevant documentation. Provide information about functional capabilities from a end user point of view.***	***Define who are the scientific communities and other target groups to which the SP services / scientific products provide value (e.g. scientific collaborations, research projects, long tail of science etc.)***
01	ESCAPE AAI Service/ INFN	AAI service used by ESCAPE, ready to be a member of the EOSC AAI federation	Target groups: researchers, service providers.. Application domains: cross-domain scientific

			collaboration, research project development, engineering development, information and communication material, education and training material;
02	Data Lake Federated storage and data services/ CERN, SKAO, INFN, ESCAPE collaboration	Data and storage federation, capable of managing Exabyte-scale data sets, in a robust, reliable and highly available manner. Data location and replication is policy-driven, and automated. The user sees only a single storage entity, while the physical storage is globally distributed and replicated. High level metadata and catalogue services provide data access. Services for high speed data transfer and data caching, enable just in time data streaming and delivery to compute nodes.	Target groups: researchers, academicians, students, policy makers, citizen scientists, stakeholders. Application domains: cross-domain scientific collaboration, research project development, engineering development, information and communication material, education and training material;
03	Virtual Observatory	Infrastructure supporting astronomy experiments, providing metadata and access to all astronomy datasets, standards, protocols etc. Environments to find and explore astronomy datasets.	Target groups: researchers, academicians, students, policy makers, citizen scientists, stakeholders. Application domains: cross-domain scientific collaboration, research project development, engineering development, information and communication material, education and training material;
04	HEP Open Data Portal/CERN	Portal exposing open scientific data from High Energy Physics. Includes sample code and workflows to make use of the data.	Target groups: researchers, academicians, students, policy makers, citizen scientists, stakeholders. Application domains: cross-domain scientific collaboration, research project development, engineering development, information and communication material, education and training material;
05	OSSR repository of software and services	Catalogue and repository of software and services developed and used within ESCAPE. Will be visible through the EOSC portal.	Target groups: researchers, academicians, students, policy makers, citizen scientists, stakeholders. Application domains: cross-domain scientific collaboration, research project development, engineering development,

			information and communication material, education and training material;