The ESCAPE collaboration

Giovanni Lamanna[1*]

[1]LAPP – Laboratoire d'Annecy de Physique des Particules, CNRS-IN2P3 & USMB, Annecy, France

Abstract. The European Union funded H2020 ESCAPE project has brought together the ESFRI and other world class research infrastructures in High Energy and Nuclear Physics, Astroparticle Physics, and Astronomy. In the four years of the project many synergistic and collaborative aspects have been highlighted and explored, from pure technical collaboration on common solutions for data management, AAI, and workflows, through development of new tools, such as AI/ML codes, and in education and training, for example in the area of research software and citizen science. In addition, the project has shown that the scientific communities have a lot in common, can act as a single voice towards the funding agencies, the European Commission, and other key programmes such as the implementation of ESOC (European Open Science Cloud). Consequently, ESCAPE partners consider forming a long-term international open collaboration that would exist independently of specific project funding, that can maintain the synergistic aspects of the cluster scientific communities. This work highlights the ESCAPE achievements and outlines some of the goals of the new collaboration.

1 The Cluster action for open science

The H2020 cluster concept introduced by the European Commission (EC), was proposed specifically in 2018 to foster new cooperative schemes among the legal entities managing or leading ESFRI and other pan-European projects within large scientific domains. Such an action aimed at supporting: (i) open-science data-intensive research, in order to raise the productivity of researchers and to lead to new insights and innovation; (ii) scientific communities at large in implementing the FAIRness of scientific data stewardship (where FAIR stands for Findable, Accessible, Interoperable and Reusable); (iii) connection of world-class Research Infrastructures (RIs) to the EOSC European Union (EU) initiative (a unique infrastructure that allows universal access to research data in Europe). ESCAPE is one of the five Science-Cluster projects that resulted from this action and that operated for four years (2019-2023). Other domain-based Science Clusters are: ENVRI-FAIR (Environment and Earth Sciences), EOSC-LIFE (Biomedical Science), PANOSC (Neutron and light sources facilities) and SSHOC (Social Science and Humanities).

* Corresponding author: giovanni.lamanna@lapp.in2p3.fr

The “European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures” – ESCAPE, brings together seven ESFRI facilities (CTA, ELT, EST, FAIR-GSI, HL-LHC, KM3NeT, SKA), two pan-European organizations (CERN, ESO), an ERIC (JIV-ERIC) and a pan-European research infrastructure (EGO-Virgo) in astronomy and particle/nuclear physics research domains to support the EOSC implementation. ESCAPE represents a unique cross-fertilization opportunity for the concerned scientific community that was fostered by two complementary excellences: the data stewardship of large astronomical archives through the Virtual Observatory (VO) infrastructure and the exabyte-scale data management and large-scale distributed computing in particle physics (such as WLCG). Furthermore, astrophysics and particle/nuclear physics communities are generators and consumers of large volumes of complex data and at the same time early adopters of innovative ICT and data management solutions pushing the state-of-the-art. The partners in ESCAPE recognise the strong synergies and potential commonalities which are there at several levels: the research collaborations themselves are often synergistic and overlapping, with cross-over between all parts of the community, and within their national research institutes; there are often common funding agencies for astronomy and particle/nuclear physics in many countries; the data and computing facilities that all of these ESCAPE partners use are often host to both astronomy and particle or nuclear physics experiments. Thus, the natural synergies of the science domains are also reinforced by these factors and have anticipated naturally a vision towards the architectural implementation of EOSC for our community. Indeed, the ESCAPE work structure aims at deploying a domain-based “EOSC cell” (Fig. 1) composed of a federated data infrastructure (DIOS), a co-created open-source scientific software and service repository (OSSR), the integration of (VO) services for multi-messenger astronomical archived data, the science analysis platform instances customized to the needs of RIs and user communities (ESAP) and services to improve access to data through citizen science crowdsourcing experiments for most of the ESCAPE facilities (ECO).

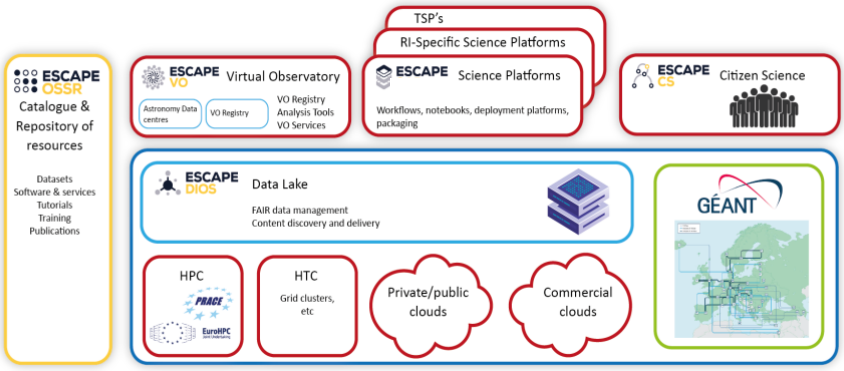

Fig. 1. ESCAPE work programme aimed at building a domain-based implementation of EOSC

1.1 The FAIRness of Scientific Software

The activities conducted by ESCAPE within OSSR aim at pursuing four main intents:

(i) Support a community-based approach for continuous development, deployment, exposure and preservation of domain-specific open-source scientific software and services in the

global context of the EOSC catalogue of services - the OSSR itself. The OSSR architecture [1] is based on collecting software in the form of a code repository or container image in a Zenodo community [2], which is connected to a specialized landing page [3] and searchable via a python client library. The landing page is the entry point of users to the OSSR products, as well as to other services within the ESCAPE EOSC cell. It also contains links to documentation and training materials (Figure 2).

The development platform provides a place to gather the common developments, ideas, guidelines and templates for the community, as well as a platform for new developments if required by an institution/group without access to another solution. It showcases the full software lifecycle up to the publication in the repository.

(ii) Enable open science interoperability and software re-use for the data analysis of the ESCAPE ESFRI projects based on FAIR principles. The OSSR is designed to be flexible through the customized adaptation of the metadata. Perspective entries to the repository have to be stored as software or datasets in Zenodo, but could range from installable software packages to containerized images of software to full analysis environments including data, or smaller databases. Extended metadata for installation or integration of the content in a larger research environment lies within the responsibility of the content provider, however, metadata required for interoperability and findability of specific entries can be integrated in the OSSR schema definition and the eOSSR library functions on request. Thus, the OSSR functionality can be adapted to the requirements depending on the contributors' use cases.

(iii) Create an open innovation environment for establishing open standards, common regulations and shared software libraries for multi-messenger/multi-probe data. Partners in the ESCAPE project and the wider science community are encouraged to onboard their scientific software, public datasets (limited in size), container images, or repositories with full analyses environments to the OSSR. They are requested to complete an onboarding process which involves the curated presentation of their project and upload of the contribution to the Zenodo community, triggering a short review process. An exemplary effort to foster cooperation is given in the ConCORDIA project to produce CORSIKA turnkey containers for various use cases in astroparticle physics and linked research fields shared for research conducted with primary and secondary detectors underwater or in ice and for muons in the low atmosphere. It involves setting up the CORSIKA containers and running test productions in various computing centres and assessing and certifying the quality and physics relevance domain of the simulation. An Innovation Competence Group was established for cross-fertilisation activities on the implementation of innovative workflows in astro & particle physics. The work therein is mainly in terms of innovation in data management, software and data analysis, in particular regarding the use of artificial intelligence and especially in the form of deep-learning techniques for querying large data archives, pre- processing data, object classification and parameter inference.

(iv) Educate stewards for FAIR software by knowledge transfer, collection of best practices and software schools. Producing FAIR software in practice requires know-how. During the ESCAPE Data Science schools, scientists in the field of astronomy, astro-particle and particle physics are taught the necessary ingredients for their software to become a part of open science by experienced code custodians. The 2021 special online edition welcomed more than 1000 registered participants. Following the FAIR paradigm and as an example of good practices in code development, all the school material is openly available online, including scientific programme, agenda and links to all contributions (software repository, notebooks, presentations and recordings) [4].

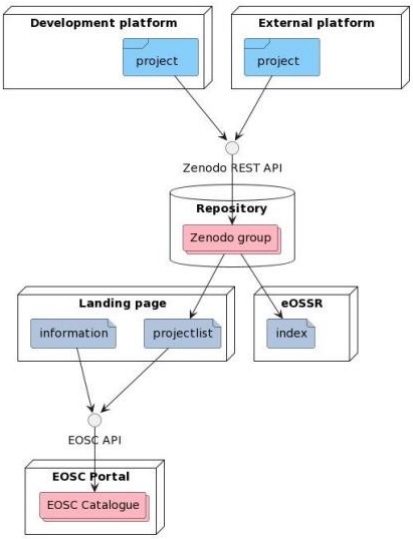

Fig. 2. OSSR architecture

1.2 Data Management for Open Science

The ESCAPE Data Infrastructure for Open Science (DIOS) activities aim at developing a federated data infrastructure, Data Lake, able to cater for the multi-Exabyte needs of the overall current and future RIs in astrophysics and particle/nuclear physics. The DIOS vision is a single data entity, hiding the inherent complexity and enabling policies, rules and data life-cycles to be applied to the Data Lake infrastructure as a whole, e.g., it is a form of storage consolidation offering a structure where data is concentrated in selected sites, globally orchestrated and accessed from the Data Lake as an entity, not as a specific site. Its architecture is based on the integration of common set of tools:
(i) to orchestrate the computing resources provided by computing centres and providers;
(ii) to facilitate data management, data transfer, data access, an information system and a common identity management infrastructure;
(iii) to respond to FAIR data management principles, interpreting the similar specification of the ESCAPE ESFRI partners as well as addressing the researchers' needs, therefore serving global and varied scientific communities in a scalable and performing manner.

The storage services hosted at different facilities are the core of the infrastructure, exposing different storage systems to the Data Lake. These systems are seamlessly integrated to work as a whole via well-known protocols: gridFTP, http/ webdav, and xroot; offering download, upload, and data streaming functionalities, third-party transfer capabilities and data deletion on a unified namespace. The open-source data management system Rucio [5] is the enabling technology that implements data orchestration in the Data Lake enabling: data policies, replication rules, file layout transitions, data life-cycles, distributed redundancy, etc. The data transfer technology stack is implemented through GFAL [6], FTS [7] and Rucio. GFAL is a multi-protocol data management library providing an abstraction layer of storage system complexity. FTS provides reliable data transfer at large scale between the storage systems enabling a Third-Party Copy ability to transfer data directly between arbitrary storage end-

points. Rucio uses GFAL for upload/download operations, FTS employs GFAL in order to perform the actual transfer. GFAL also works on the storage file system level with the supported protocols. Rucio provides a common namespace for the users to interact with data.

The authentication and authorization schema used to manage access control relies on the capabilities are implemented through the INDIGO IAM (Identity and Access Management) [8] service and accommodates the needs of RIs to deal with embargoed/restricted-access data. IAM allows ESCAPE to use both legacy X.509 certificates as well as token-based authorization as needed for the future.

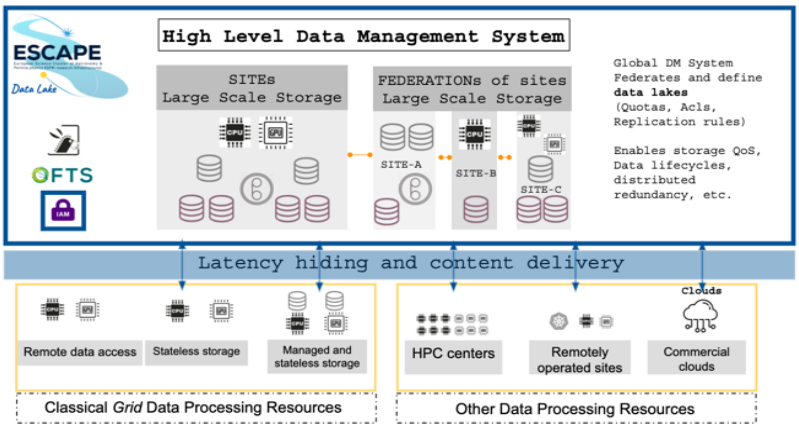

Fig. 3. The ESCAPE Data Lake Model. An overarching Data Management and Orchestration layer to deliver data to a different type of computing infrastructures and resources.

Data Lake can hence be used both: *a la cloud* push/pull experience ironing out all the complexity for end-users, but also as an extremely capable fine-grained system that can be exploited by data management experts to implement access control, fine-tuned data replication, data pre-placement, storage quality of service, and leverage content delivery services, etc. The de-centralised Data Lake model can be flexible and leveraged to efficiently serve data for a variety of wide-ranging use cases. It is a quite different task to address a data processing campaign where the same job types might run on hundreds of Petabytes, compared to the case of hundreds of user-analysis jobs running on a few files of few Gigabytes each. Both extremes can be managed by the Data Lake. Furthermore, the way data gets delivered to the user community can be varied, from end-user laptop and notebooks or to the user-neighbourhood infrastructures (e.g., large storage at a university or caching layers). Therefore, the processing facility can range from: (i) a very simple resource, a laptop of a single user downloading a file via http, or a notebook service on a cloud accessing data via the Rucio-CLI to (ii) large structures, such as heavy-duty batch systems, a large cloud provisioning or a punctual but massive CPU/GPU allocation on a High-Performance Computing (HPC) centre. This reflects the fact that computing resources are getting very heterogeneous, the standard one-size-fits-all grid-site model is vanishing. Also hinting that data locality relying on a fully-fledged storage system is not always needed nor effective to exploit such a variety of computing resources.

DIOS activities have been dedicated to a continuous assessment and evolution of the Data Lake model during the last years. The Data Lake model (Figure 3) has been successfully proven as a capable working system. The ESCAPE ESFRIs/RIs have the possibility to

implement, in a production-like scenario, entire data life cycles ranging from raw data recording to data distribution for user access. Some of the tools and services used in the ESCAPE Data Lake have been already picked up by some ESFRIs/RIs for testing, and are currently being evaluated as a possible solution to address the upcoming challenges in their computing models [9]. Two examples can be quoted [10]: (i) A real data workflow use case at the MAGIC telescope (CTAO precursor) was represented by the remote storage in La Palma island being Data Lake aware and acting as a buffer injecting fresh data to the ESCAPE Data Lake. Once the files are consolidated and with the right replication level, the files are deleted from the local buffer allowing new data to be stored. (ii) LOFAR (SKAO precursor) data have been injected to the Data Lake from three radio source observations in external locations. Users in an external location were able to download the data, process and store results back to the Data Lake. Furthermore, by exploiting other ESCAPE services (ESAP as well as VO archived data) users have combined results stored with other public Hubble data, analysed them and finally uploaded the results of the combined analysis back to the Data Lake. (iii) The extreme long haul data management tests by SKAO certified end-to-end data lifecycles were performed injecting regular data products at each of the two "telescope" storage endpoints (IDIA in Cape Town and AARNET in Perth), incorporating the following aspects, to mimic telescope data product placement into a staging area: Upload the data to a non-deterministic source storage endpoint, register the data in place in Rucio, Subscription-based movement, Long haul transfer, Life-time based QoS transitions (including start and end dates), short lifetime (1hr) at source; then transfer data from IDIA and AARNET to Manchester and Lancaster (both UK sites). These tests were fully successful.

1.3 The Virtual Observatory

The VO (Virtual Observatory) is a framework of open standards for making astronomy data FAIR. It is an established and operational framework that has proven to be a great success for many aspects of the interoperability and FAIRness of astronomy data. The standards are defined by the community driven International Virtual Observatory Alliance (IVOA). Astronomy data providers, in particular ground- and space-based observatories publish their data using the IVOA standards, and compliant scientific tools and services enable the discovery, access and (re-)use of the data by the whole astronomy research community. The ESCAPE activities of the VO aim at: integrating the astronomical VO services into the EOSC; implementing FAIR principles for ESFRI data via the development and use of common interoperability standards; applying new techniques to data in astronomy archives to add value to the scientific content. One illustrative example of the original ESCAPE results is the development of sky-spatial and temporal coverage systems for indexing of astronomy data, designed to support fast data access and management of complex sky regions, for example for gravitational wave follow-up campaigns connected to EGO/Virgo. The IVOA MOC 2.0 standard which provides a "Multi-Order Coverage" map [11] based on the HEALPix tessellation, was led by ESCAPE and become an IVOA standard. This provides a general capability for being able to search and cross-match astronomy data sets based on their sky coverage and their temporal coverage.

The new capabilities were applied by EGO-Virgo and resulted in a journal paper "*Multi Order Coverage data structure to plan multi-messenger observations"* [12]. The paper includes an interactive python notebook and on-line video tutorial materials, both of which were used in the ESCAPE training event the '2nd Science with interoperable data school' [13]. Other notable results are the new VO services implemented by the ESCAPE partners in

the areas of radio astronomy, high energy astronomy and solar physics, and, furthermore, the maturation of the standards and tools for data provenance, in particular for high energy use cases of CTA and KM3Net. Achievements have also been made for the application of Deep Learning to the content of the ESO archive, with a number of demonstrations being presented to the astronomy community, and also scientific results published [14]. ESCAPE has made important progress on the visualisation of multi-scale astronomy data, with the development of a new WebGL enabled version of the Aladin Lite [15] application which serves as the visualisation component of a number of Astronomy archives. Implementation of the visualization tools built in ESCAPE are expected in the ESO and ESA archives. All software libraries and applications can be on-boarded to OSSR.

1.4 Unifying Science Analysis Platforms

Since several years many interactive analysis tools have been deployed and many heterogeneous developments in the field of "science platforms" have been produced. Such a new trend in user analysis, the potential benefits and the consequent adoption of which are evaluated by any researcher individually, was lacking a homogeneous and easy to apply approach in all domains. This has directed ESCAPE partners to activities related to ESAP (ESFRI Science Analysis Platform) aimed at: (i) surveying requirements and expectations of ESFRIs looking at ways to deploy visual and interactive open data-provision and -analysis platforms against the classical individual user approach of terminal-text based job-submission; (ii) exploiting notebook functionalities bridging barriers between outreach and large scale analysis; (iii) moving beyond fragmentation by proposing a model of common standards and interconnectedness among science platform efforts; (iv) providing ESAP prototype architecture that, individual ESFRIs or their communities that need particular capabilities can build on a common, interoperable technological basis, customizing and extending it.

Several of the ESCAPE-affiliated ESFRIs are under construction and are making strategic choices within their computing models for delivering open data, software and services to end users. For example, CTAO is currently evaluating how to provide user-facing services in the context of its up-coming Science Data Challenge, while SKAO is beginning prototype developments for its network of Science Regional Centres, which will be the observatory's primary means of delivering data to end users. Both of these infrastructures have been heavily involved in ESCAPE in general and in the development of ESAP in particular. The value of the cross-ESFRI approach within ESCAPE is determined by the fact that: a) no individual ESFRI will find it advantageous to assume responsibility for a legacy software system, even if that system provides useful functionality, while shared solutions within a large community involvement would be more reliable; b) since every ESFRI has responsibility for the quality of the released data sets and workflows to end-users, it is also aimed at guaranteeing support, maintenance services and long-term structural support for delivery of its own ESAP as an operational service ; c) a long-term Science Cluster action can foster the continuous inter-ESFRI cooperation to centre each ESAP around common technical standards. ESAP is finally a science platform toolkit: an integrated set of software components which ESFRIs, ESCAPE project partners, can use to rapidly assemble and deploy platforms that are customized to the needs of their particular user communities. These various deployed instances of ESAP then provide the key interfaces between the services delivered by the ESCAPE project and the wider scientific community, by abstracting the details of heterogeneous underlying infrastructures away from users, thanks to a modular, plugin-driven architecture: an ESAP instance is integrated with its surrounding environment by enabling and configuring an

appropriate selection of plugins, and new capabilities are easily added by writing new plugins. While this plugin-based system makes ESAP infinitely re-configurable, its basic functionality includes tools for discovering and accessing data, services, and software from ESCAPE project repositories; access to a range of computing and analysis services; and orchestration of data, services, and software to help users create and access research environments that meet particular needs. In addition, ESAP remains endlessly adaptable to integrate with new external service offerings. Many opportunities for these arise: from the management of OpenStack-based virtual machines to integration with workflow- or function-as-a-service systems.

In the future ESCAPE could foster a federation between distributed ESAP instances. At the same time ESCAPE has addressed one more need that has been raised by hosting open-science test projects: the perspective of a community-based platform for multiple-source data interoperability. This is the case when combining multi-messenger astrophysics and astroparticle data, or when conceiving a unique easy-to-access environment for combining High-Energy Physics or Nuclear legacy data, from different accelerator runs and/or independently of the experiments that produced it for re-analysis and results reproducibility. In such a vision ESAP provides the basis for a Virtual Research Environment, an online collaborative system for scientists to conduct high-impact research.

1.5 A Virtual Research Environment for open-science projects

Within the ESCAPE programme, a number of scientific test projects (TSPs) were planned by bringing together more than one ESFRI, initially conceived as validation benches for ESCAPE services and their technical expertise to interoperate data and combine analyses. It became clear that such conclusive tests on scientific use cases would have two critical values: (1) the TSPs support the implementation of a landing gateway and web-based collaborative system for scientists to host their complete data analysis workflow and access all ESCAPE services, a virtual research environment (VRE) [16]. (2) The TSPs aim to produce new scientific results and publications that demonstrate the innovative impact of open science and strengthen the involvement of the wider scientific community in this field. It then proved appropriate to take advantage of the TSP approach in order to federate collaborative actions between researchers, possibly transversal (e.g., experimentalists and theorists; astrophysicists, particle physicists and nuclear physicists). These objectives were achieved through the ESCAPE consortium's participation in another dedicated EU project, “EOSC Future”, in which a series of pilot analysis projects are currently supported within two TSPs.

The *Dark Matter* TSP aims to highlight the synergies between different communities searching for dark matter by making the necessary data and software tools fully available, in particular focusing on data management, data analysis and computing, using experimental data and software procedures from selected direct detection, indirect detection, and particle collider experiments (such as HL-LHC, KM3NeT, CTA, DarkSide) [17][18][19].

A related objective of this Science Project is to support the creation of a collection of versioned repositories for experimental results as well as for theory predictions and interpretations. We will implement this service by making available experimental data, results and their interpretations discoverable through the ESCAPE VRE. In particular, in the HEP domain, one pilot focuses on the search for dark matter particles or mediators produced at the LHC and detected by ATLAS. A first stage consists of implementing the flagship analyses with visible and invisible signatures in ATLAS within the RECAST [20]

reproducible analysis pipeline. RECAST brings together all the software used for analysis from semi-calibrated data, and enables re-execution using a declarative workflow and containers. The semi-calibrated data will be retrieved from the ESCAPE Data Lake, and the software pipelines will be distributed and executed via the software catalogue OSSR. Once this is done, RECAST will be interfaced with REANA [21], a portal hosted by CERN where external users can request the execution of analysis pipelines to test new signal hypotheses.

The *Extreme Universe* TSP aims to prototype a platform and to develop innovative workflows for end-to-end analysis pipelines of multi-messenger astronomy, enabling the use of data from multiple telescopes of the full electromagnetic spectrum as well as gravitational waves, neutrinos and cosmic rays. Pilots of such a TSP concern: analyses about a combined broadband spectral energy distribution fitting of blazars (with Fermi, HESS and other archived data); machine learning-based workflows to combine on-line monitoring of gravitational-wave (Virgo) alerts and multi-messenger counterpart (the "Wavefier" project) as well as detection follow-up observations (within the VO); searching for fast radio burst persistent radio source (with LOFAR observations); exploiting data and simulations to produce and update Instrument Response Functions for neutrino point source analysis (for KM3NeT).

ESCAPE has deployed a VRE prototype. The user access to it is performed through the ESCAPE AAI that is also within the EOSC federation, meaning that potentially ESCAPE users' credentials could enable access to multidisciplinary open data. The entry point of such a platform is a jupyterhub instance deployed on top of a scalable Kubernetes infrastructure, providing an interactive graphical interface for researchers to access, analyse and share data. The data access and browsability is enabled through API calls to the high-level data management and storage orchestration software (Rucio) that enable access to the Data Lake. Scientists have access to software and workflows directly into the OSSR and they can select computational environments from Docker images and the integration of a re-analysis platform (Reana) supporting various distributed computing backends. One more key functionality of the VRE is the capability of researchers to ingest and upload their results enabling a continuous cooperative approach extended to the full community. The final goal of the VRE project, bringing together data and software access, workflow reproducibility and enhanced user interface, is to facilitate scientific collaboration, ultimately accelerating research in various fields.

1.6 Citizen Science

One more explorative goal of ESCAPE is to connect science to society at large, by involving fellow citizens more directly in open science through mass participation experiments, i.e. « citizen science ». The ESCAPE vision is to improve access to data and tools through citizen science crowdsourcing experiments for most of the facilities in ESCAPE. Furthermore, through the synergistic work conducted within the EOSC Future project, ESCAPE aims at extending the remit of these activities to wider science areas and facilities. From a technical point of view, ESCAPE removes any obstacle that would prevent a citizen from being, in the same way as a researcher, an end user of the EOSC-cell's services. Thanks to the authentication and authorisation service, all that needs to be done is to select datasets, workflows and training material suited to a citizen's level of ability and rights of access. A central vision of the EOSC is to make scientific data FAIR. Implicit in this vision is that FAIR data must also be useful. While this is a task for ESFRIs, which are responsible for the quality of their own scientific data, in the case of society, the main task is to provide scientific

data and projects that make everyone aware that they are part of the challenge of scientific research and at the same time guarantee advances in public learning.

ESCAPE has participated in some citizen science projects [22]: - the Zooniverse platform was used to develop a new citizen science project called *Galaxy Zoo: Clump Scout*, representing one of the first large-scale studies of clumps in local galaxies; - *Radio Galaxy Zoo: LOFAR*, is a project that invites volunteers to classify radio images extracted from the first data release of the LOFAR Two-metre Sky Survey (LoTSS) which covers 424 square degrees in the region of the HETDEX Spring Field. In this release, 325,694 individual radio sources were detected with a signal five times greater than a typical background noise fluctuation.

2 The ESCAPE open collaboration

The successful work programme, the achievements and the synergistic capacity of ESCAPE are widely recognised. The top-down approach of the ESFRI and other world-class RIs willing to continue cooperative actions by joining their efforts is confirmed. The bottom-up demands of the scientists involved not to interrupt but to continue the cross-fertilisation in science and innovation that ESCAPE has been able to build, are strongly considered. These reasons have motivated the establishment of a long-term and inclusive Science Cluster action after the end of the H2020 grant. A new "ESCAPE open collaboration" agreement came into force on January 1st 2023 and for the next three years, has been signed by the nine partners, CERN & HL-LHC, CTAO, EGO-Virgo, ESO & ELT, EST, FAIR-GSI, JIV-ERIC, KM3NeT, SKAO, and joined more recently by the Einstein Telescope (ET) ESFRI project. The collaboration is always open to any new related RI's willing to join, to recognise synergies and act as a coherent community. This will maintain the exceptional collaborative and human experience represented by the Science Cluster and strengthen the role and impact of astronomy and nuclear/particle physics in the field of open data-intensive research.

The ESCAPE work programme has been updated, although all current activities still represent its core; related working groups are either organizing actions for consolidation of services or fostering next developments. Some major orientations follow.

DIOS: The Data Lake will evolve continuously since co-developed by almost all partners, and several are adopting it in their own computing models. The sustainability of the DIOS activities is guaranteed by the RIs as well as by the national institutes' commitment. The other Science Clusters have expressed interest in the ESCAPE achievements, therefore within the new Horizon Europe (HE) EU framework a new action has been approved to explore and potentially extend the Data Lake infrastructure to other domains.

ESAP and VRE: The analysis platform further customization is a priority from some of the ESCAPE RIs and the VRE concept is conceived more like a community-based global service. While it is considered relevant by our scientific community it will be further developed and tested. The ESCAPE VRE, that is quite advanced technologically, catches the interest of other communities. The wish to extend the rationalisation arguments that motivated ESAP is shared by all Science Cluster communities, that will enter together, during the next months, in a mutual consolidation and convergence activity about their own VRE prototypes. It is worth to underline that the recent evaluation of the value of the ESCAPE VRE is quite encouraging. Within HEP the VRE is now seen as a solution to allow re-analysis of different

data sets of subsequent LHC runs, cumulating, combining or comparing data from all experiments (e.g., ATLAS and CMS) as well as an easy to take-over homogeneous and easy to use environment for data analysis for any user despite which experiment is involved. Consequently, since user rights are tuned to the user typology and automatically deployed through the ESCAPE AAI, the VRE and the legacy Science Projects results therein, would help researchers with the implementation of Particle Physics Masterclass exercises, Citizen science projects or outreach actions, within the same environment.

OSSR: The open-source scientific software and service catalogue is a pathway to onboarding ESCAPE products and ESCAPE communities results into EOSC. OSSR is a key service of the VRE and is supported by the ESCAPE collaboration towards three objectives: (i) continuation of up-to-dated training schemes for software programming and deployment of quality of code development and preservation standards; (ii) foster the cooperation within the ESCAPE community at large for sharing efforts for innovative software developments, though a "community foundation" approach, similar and in cooperation with HSF; (iii) develop, in collaboration with other Science Clusters, a European Virtual Software Institute for Research Software. The concept of the Virtual Software Institute is to tap into the research knowledge of University departments, software engineering schools for the benefit of (natural) science developments. The aim is to:
- Enable R&D resulting from collaborations of computer science and natural science;
- Establish a career path for scientists and engineers working in software and computing in natural science (in many fields the recognition of software work and finding / retaining experts is a major concern);
- Cross-fertilize knowledge between different science domains and make the acquired knowledge available across domain boundaries;
- Act as a lobbying organisation and raise awareness of software and computing in natural science.

The ESCAPE collaboration has succeeded in getting approval and financial support by the EU within the Horizon Europe work programme in support of all these objectives.
Coherently, the new work programme of the ESCAPE will also bring a relevant focus on the career development for young scientists and training in astronomy, astroparticle, astrophysics, cosmology, high-energy, and nuclear physics, specialising in scientific computing and research software. This action would be pursued also in collaboration with other Science Clusters and will build upon the activities of a Virtual Software Institute

The successful results of the explorative experience of the (Test) Science Projects motivate the ESCAPE community to target a second phase within Horizon Europe work programme to uptake new emerging and challenging "Open Science Objectives". Extend commitments from more RIs in current and new Open Science Projects. ESCAPE will also leverage the inter-cluster coordination for Cross-Cluster Open Science Projects and when relevant will act to reach out and support "the long tail" of science and multiple scientific communities. A dedicated action has been approved by the European Commission and an important funding is dedicated to support it. Namely, the five Science Clusters will organise and manage during the next years some calls for proposals of new open science projects and/or services. The ESCAPE partners are confident that the astro./particle/nuclear physics researchers will be able to take advantage of these opportunities and in this manner to bring a significant contribution to the success of the ESCAPE action.
The directorates of the ESCAPE partner RIs, as well the national research institutes that support them, share the idea of committing to the European Strategy for Data by exploring and building synergies on "Sector Data Spaces". In some of them interests of ESCAPE

Parties would emerge for transdisciplinary studies and for the provision of secure and FAIR-enabling European cloud services. For example:

1) Cross-sector sharing of data and Green Deal data for a unifying, forward-looking approach of any Big Science facility for energy efficiency, water management, support of circular economy and any environmental implications of RIs construction, etc.
2) Health data linked with high energy particle and nuclear physics facilities for preventing/treating diseases.
3) Industrial and Manufacturing data by exploiting FAIR digital objects from the R&D programmes for detectors, sensors, telescopes and other devices of the ESCAPE RIs.
4) Opening data and innovation projects to training actions by research for a Skills data space, to reduce the skills mismatches between the education and training systems and the labour market needs.

Thus, ESCAPE sees a strong motive for paving a challenging and successful path ahead, building and making use of the inherent synergies between the RIs and their communities. This in turn will optimise the significant long-term investments of the funding agencies by sharing facilities, tools, and experience across the ESCAPE community.

1. https://escape2020.pages.in2p3.fr/wp3/ossr-pages/page/tools/ossr_curation/
2. https://zenodo.org/communities/escape2020/
3. http://purl.org/escape/ossr
4. https://zenodo.org/record/5838436
5. M. Lassnig, *Extending Rucio with modern cloud storage support: Experiences from ATLAS, SKA and ESCAPE*, CHEP2023 (2023)
6. https://github.com/cern-fts/gfal2-util
7. https://fts.web.cern.ch/fts/
8. https://github.com/indigo-iam/iam
9. R. Bolton, *Global Data Management in Astronomy and the link to HEP: are we collaborators, consumers or competitors?* CHEP2023 (2023)
10. G. Lamanna, *Open Data for APPEC-ECFA-NuPECC* - JENAS, Madrid, (2022)
11. https://ivoa.net/documents/MOC/20220727/REC-moc-2.0-20220727.pdf
12. G. Greco, M. Punturo, M. Allen et al. Astronomy and Computing 39(1):100547 (2022)
13. https://indico.in2p3.fr/event/25225/
14. N. Sedaghat, M. Romaniello, J. E. Carrick. F.-X. Pineau, Monthly Notices of the Royal Astronomical Society, 501, 4, 6026 (2021)
15. https://aladin.cds.unistra.fr/AladinLite/
16. E. Gazzarrini et al., *The Virtual Research Environment: towards a comprehensive analysis platform*, CHEP2023 (2023)
17. J. Little et al., *The ESCAPE Dark Matter Science Project in the European Science Cloud*, CHEP2023 (2023)
18. "ESCAPE to the future" conference - Dark Matter TSP presentations (2022)
19. "ESCAPE to the future" conference - Extreme Universe TSP presentations (2022)
20. https://recast.docs.cern.ch/
21. https://reanahub.io/
22. "ESCAPE to the future" conference report about Citizen Science projects (2022)