Project Number: [824064]

Project Acronym: [ESCAPE]

Project title: [European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures]

Periodic Technical Report

Part B

Period covered by the report: from [01/08/2020] to [31/01/2022]

Periodic report: *[*2nd*]*

Table of Content

1 Explanation of the work carried out by the beneficiaries and Overview of the progress

The European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures (ESCAPE) brings together seven ESFRI facilities (CTA, ELT, EST, FAIR, HL-LHC, KM3NeT, SKA), two pan-European organizations (CERN, ESO), an ERIC (JIV-ERIC) and a French-Italian private Consortium (EGO-Virgo) in astronomy and particle physics research domains to support the EOSC implementation. ESCAPE will deliver a list of services (Fig. 1) to ensure that after the end of the project, there is a clear pathway and sufficient momentum to take forward the full involvement of the ESFRI projects in the European Open Science Cloud as presented in the schematics below.

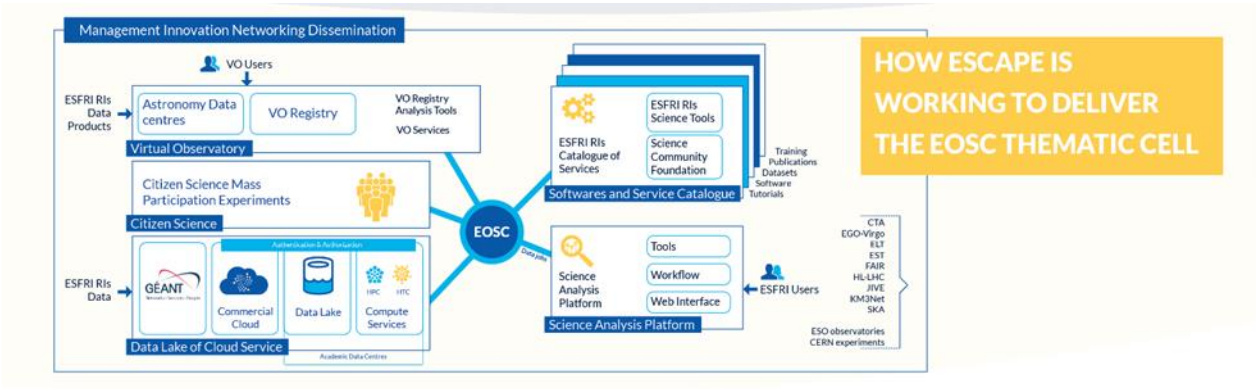

Fig. 1 ESCAPE Services to be integrated with European Open Science Cloud

Such an implementation has a three-fold impact: i) it aims at building a FAIR virtual research environment dedicated to the scientific community at large to interoperate data and workflows for fundamental science; ii) it raises by developing and deploying technological data management solutions mutually adopted by a large fraction of world-class RIs and potentially extendible to further disciplines; iii) it represents the declination of the “EOSC-portal” functionalities for our disciplines (which is guaranteed by the ESCAPE contributions in other related EC projects such as EOSC Enhance and EOSC Future).

As for the first one, also in this second reporting period, project management and beneficiaries made sure that the activities remain on track as per the description of work from Annex 1 (Part A) of the H2020-ESCAPE grant agreement. With active support from the work package leads, project management has been able to keep track of the project progress throughout the reporting period. Project activities were widely disseminated and communicated with multiple stakeholders during the reporting period. More specifically this second reporting period has been also distinguished by a very intense activity of promoting inter-cluster actions, dialog with major stakeholders such as the EOSC Association Board, the ESFRI board, the ERIC forum. That has also aimed at paving the way to sustain ESCAPE as long-term coordination platform for the benefit of EOSC as well as the European Research Area (ERA).

On the technical progress side, work plans were developed by the technical work packages in close consultations with participating ESFRI/RI projects. As elaborated in the following sections, most of the project objectives were addressed in this second reporting period.

Some of the significant progresses reached by ESCAPE during the second reporting period include:

- Active and intense networking with other Science Clusters as well as with EOSC and ESFRI boards. Deep investigation and elaboration of a plan for enhancing researchers' involvement in EOSC that included the ESCAPE participation in EOSC Future as well as the publication of a position statement about a long-term perspective.
- Technical coordination among ESFRI projects, topical meetings and co-development actions at the interface among works packages.
- From preliminary plans or first prototypes to operational components of the ESCAPE-EOSC cell for open science (figure 1). Namely through: (i) a functional Data Lake implemented with a large fraction of national storage nodes. Testing large scale data production workflows of ESFRI projects also interfacing public and commercial computing; (ii) interfacing astronomical VO services and archives with Data Lake data management services; (iii) deploy the open-sources ESCAPE catalogue of resources (data and software), providing standards and linked to services for data analysis; (iv) first analysis platform linking all the above components and providing user analysis environment.
- The startup of open Test Science Projects -TSP (within the context of the ESCAPE participation in EOSC Future) has boosted the ESCAPE work programme by customizing a (VRE) virtual research environment. The VRE is supposed to rely on the ESCAPE analysis platform, is hosting the TSP and leverages the interoperability of data demanded by any researcher. For education purpose and for serving also the long tail of science the VRE tackles from one side any level of outreach (including the one towards the citizens) and from another implements solution for rewarding scientists committing in open science.
- Preliminary design of the Open-source scientific Software and Service Repository was accomplished; the definition of technical solutions for its implementation have been developed and a first prototype was set up for internal use.
- ESCAPE participation at International Virtual Observatory Alliance (IVOA) interoperability meetings made sure that the priorities of the ESFRIs and RIs are being considered in the definition of common interoperability standards at the international level.
- Despite the COVID19 pandemic restriction the online training events and technical workshops have been organized according to our plans and have been a great success in terms of impact and popularity.

1.1 Objectives

As an introduction, we highlight here the main progress achieved during the second reporting period, according to most of the global objectives of the project derived from the explanation of work part (section 1.3.3) of the ESCAPE grant agreement 824064. More details are provided in the following sections dedicated to each work package.

- Establish the ESCAPE Management Support Team (E-MST), and will thus guarantee the smooth execution of all financial, administrative and reporting elements of the project. It will also permit the E-MST to exercise central control and oversight of the scientific progress and the inter-WPs technical coordination of the project, as measured by the successful receipt of deliverables and secured milestones.
- Establish a competence desk in support of the concerned ESFRI projects for FAIR data management issues.

The project manager follows up with day-to-day managerial activities of the project as per the grant agreement. During this second period, priority has been given in particular to: (i) the internal technical coordination by enhancing dedicated inter-work packages topical discussions and cooperative actions; (ii) surveys for cross-RIs data management service deployment; (iii) scientific use-case for interoperability of data and workflows. This has facilitated the spontaneous establishment of a competence desk/group of “FAIR data champions” of the concerned RIs.

- Networking and partnership with stakeholders, industries and other projects.
- Networking with other clusters.

The active participation of the ESCAPE coordinator in 2020 to the set-up of the EOSC Future work programme as well as in 2021 to establish a dedicated consortium, are among the most relevant achievements along such objectives. The coordination action with the other science clusters (EOSC-Life, SSHOC, ENVRI-FAIR and PANOSC) as well as with other pan-European e-infrastructures has been intense and successful. We acknowledge a large number of workshops, co-edition of position documents and meetings with EOSC Association Board, ESFRI Board and EC.

- Establish a supervisory committee, involving the management boards of the ESFRI facilities and other major research infrastructures encompassed within ESCAPE. Such a committee is convened about more general policy matters of common interest.

ESCAPE encompasses seven ESFRI facilities (CTA, ELT, EST, FAIR, HL-LHC, KM3NeT, SKA), two pan-European organizations (CERN, ESO), an ERIC (JIV-ERIC) and a French-Italian private Consortium (EGO-Virgo) in astronomy and particle/nuclear physics research domains to support the EOSC implementation. The members of the ESFRI RIs’ Directorates and Management Boards have been regularly consulted on multiple policy related aspects. More precisely the WP1 MIND addressed this objective in the reporting period by consulting the board on the EOSC Strategic Research and Innovation Agenda (SRIA) questionnaire, on

the second position statement document about the long-term perspective of the clusters, on the influence document about the evolution of the cluster work programme in the Horizon Europe framework, in proposing plan and a new MoU for a sustainable ESCAPE platform.

- Cooperation with thematic ERA-NET initiatives (such as ASTRONET, APPEC, NuPECC, etc.).

WP1 MIND addressed this objective in the reporting period. The chairs of the ASTRONET, APPEC, NuPECC and ECFA consortia together with a representative from ESA have been formally appointed as the ESCAPE External Expert Advisory Board (E-EAB) members. The E-EAB is comprised of the following five members.
- Andreas Haunghs, Chair of APPEC - Astroparticle Physics European Coordination committee
- Marek Lewitowicz, Chair of NuPPEC - Nuclear Physics European Collaboration Committee
- Karl Jacobs, Chair of ECFA - European Committee for Future Accelerators
- Colin Vincent, Chair of ASTRONET – Astronomy European Collaboration
- Christophe Arviset, ESA - European Space Agency.

The EEAB provides external advice and evaluation of the achievements of the project. The Board brings additional expertise to the project and comments on strategy to optimize project activities. The chairs of these consortia also support the dissemination and uptake of ESCAPE results in their respective consortia. The E-EAB members have been associated to the ESFRI supervisory committee members during the ESCAPE coordinator communications that have been addressing relevant issues concerning the rules of participation and work programme cluster actions devoted to enhance the researchers' participation in Open Science and in the EOSC implementation.

- A Data Lake prototype.
- Data Preservation, ensuring the long-term data preservation at the scale of tens of Exabytes is essential for the implementation of FAIR principles.

A pilot ESCAPE Data Lake was designed, deployed and assessed through a joint exercise by the partner ESFRI RIs labelled as Full-Dress Rehearsal (FDR20), from data recording from the telescopes/detectors/sensors to data browsing and access by users. This was followed by one more data-challenge (DAC21), successfully performed, where RIs designed several production-like activities covering processing workflows and data analysis pipelines as well. It is worth to acknowledge that during the DAC21 two astrophysics RIs drove the Data Lake Data Management in a common shared multi-VO storage infrastructure. Indeed, some end-to-end workflows have been running integrally in the Data Lake, that, although originally developed for particle physics, is now expanded to astroparticle physics, astronomy and cosmology communities thanks to ESCAPE.
One important progress of the second reporting period was to address expectations of a larger typology of potential users. This implied the evolution of the Data Lake to respond to a much narrower "just-a-bunch-of-files" individual user-perspective.

- Cooperative actions at national or regional and international level.
- Computing interface and scalability.
- Industrial and commercial involvement.

The ESCAPE Data Lake inherent flexibility enable heterogeneous data processing facilities in the Data Lake, activities progressed with commercial clouds, with the implementation of Swift/S3 storage endpoints in the Data Lake, with CINECA/HPC, and through a collaboration with the FENIX/HPC project.

- Skills Training and dissemination activities.
- Support a community-based approach in the global context of the EOSC catalogue of services.
- Enable open science interoperability and software re-use for the data analysis of the ESCAPE ESFRI projects based on *FAIR* principles.

ESCAPE beneficiaries contributed to the development, benchmarking and deployment of software. Gathering common practices and know-how towards the definition of best community approach for FAIRness of software was a highly relevant step forward for the benefit of the ESOC architecture.
During the reporting period, the partner contributed to and developed a series of software and services, enriching the ESCAPE catalogue content. All developments are openly available and an example project exists, showing the full possibilities of the system, from best practices in software development (e.g. licence and meta data), over continuous integration to test and upload the project to the catalogue, that is linked to the EOSC core services. It is based on the Zenodo repository, developed under the *escape2020* community (https://zenodo.org/communities/escape2020), with a dedicated landing page (http://purl.org/escape/ossr) that serves as entry point for all users. In continuation with the ASTERICS cluster school, ESCAPE has organized its first school (https://projectescape.eu/events/escape-summer-school-2021) devoted to software project development for astrophysics, astroparticle physics & particle physics. It provides theoretical and hands-on training on Data Science and Python development (coding environment and good code practices, version control and collaborative development, Python packaging, scientific libraries for data science analysis and machine learning). An unprecedented success story with 1000 registered participants, about 3000 views and more than 300 certificated delivered.

- Create an open innovation environment for establishing open standards, common regulations and shared software libraries for multi-messenger/multi-probe data**.**

Machine learning approaches to simulation and experiment data have been adapted and benchmarked; definition of data formats and different deep-learning approaches have been pursued as well as the exchange of experience and harmonisation of approaches for innovative workflows.

- Assess and implement the connection of the ESFRI and other astronomy Research Infrastructures to the EOSC through the Virtual Observatory framework.
- Refine and further pursue implementation of FAIR principles for astronomy data via the use and development of common standards for interoperability including the extension of the VO to new communities.

- Establish data stewardship practices for adding value to the scientific content of ESFRI data archives.

The integration of VO in the current EOSC infrastructure has progressed significantly during the reporting period. VO community of data providers and consumers has analysed largely connection points, requirements and challenges fur such a purpose. We enhanced ESCAPE participation to EOSC working groups on standards as well as contributions to events of the FAIRsFAIR project as well as RDA plenaries. Cross-WP activities have been key to implement the connection of VO with the other components of the ESCAPE cell as well as to prepare to connect VO to ESFRI projects' data archives.
Establishing data stewardship practices in astronomy deserved training. The "First Science with Interoperable Data School" was organised aiming at exposing early-career European astronomers to the variety of currently available VO tools and services.
Major progresses were achieved towards interoperability standards of tools definition for the multi-messengers' investigation with the future advent of the ESFRI projects currently in construction.

- Build a prototype science analysis platform tuned to the needs of the ESFRI projects.
- Assessment and adaption of the science platform to the specific needs of the various ESFRI projects and communities represented by the ESCAPE project.
- Development and porting of novel ESFRI tools and workflows to the science platform.

The main progresses accomplished along the plan for implementing the analysis platform concerned the "integration" activity. Namely, the one that enabled the prototype science analysis platform to interface to most of the other components of the ESCAPE cell: (i) "Data-Lake-as-a-Service" (DLaaS) project; (ii) enabling attached storage in analysis environments of any user, for making large data products available to the compute resource; (iii) VOSpace storage and query services; (iv) analysis software catalogue access and deployment; (v) user analysis environment services.

- Execution of the public engagement and communication strategy, based around project results, serving all WPs.
- Improve access to data and tools through citizen science crowdsourcing experiments for most of the facilities in the ESCAPE remit.

ESCAPE has been working to integrate into its analysis platform the Zooniverse citizen science platform, with the aim of addressing requirements from the largest possible categories of users. A recent extension of our tasks has been to incorporate engagement not just with the ESCAPE science community, but with the wider EOSC-Future communities, for instance by extending the remit of Second ESCAPE Citizen Science Workshop to research domains beyond ESCAPE. As a direct result, we now have crowdsourced data mining projects at an advanced stage of development in the domains of the SSHOC and ENVRI-FAIR research infrastructure clusters.

1.2 Explanation of the work carried out per WP

1.2.1 Work Package 1 MIND (Management, Innovation, Networking and Dissemination)

WP Lead	Giovanni LAMANNA (giovanni.lamanna@lapp.in2p3.fr)
WP Participants:	**CNRS**, CERN, NWO-I, FAU, OU

1.2.1.a **Introduction**

WP1 MIND focuses on internal project coordination between the ESCAPE work packages, project beneficiaries, participating ESFRIs, intergovernmental and pan-European organizations. This involves day-to-day management activities, continuous project reporting, financial management and project governance. In addition, WP1 MIND also acts as an interface between European commission and ESCAPE consortium members for contractual obligations. The interactions with EOSC community and EOSC stakeholder projects are coordinated by WP1 MIND.

1.2.1.b **Organisation**

To date ESCAPE management support team (E-MST) establishes project governance structure and handles overall day-to-day management of the ESCAPE project. It consists of the Project coordinator, Technical coordinator, Project Manager, Communications officer and Financial controller. The ESCAPE executive board (E-EB) is comprised of the ESCAPE coordinator, E-MST, work package leads and the chairperson of the ESCAPE General Assembly (E-GA). The executive board is the executing body of the project consortium.

The Project coordinator represents the consortium at the high-level strategic discussions with European Commission policy officers, ESFRI management board members as well as EOSC cluster coordinators and the EOSC community. The technical coordinator ensures the complementarity between ESCAPE work packages and ESCAPE ESFRIs with respect to the ESCAPE workplan. The Project manager along with the Financial controller and Communications officer support the consortium with continuous project reporting, periodic technical reporting, financial overview as well as communication and dissemination of project activities. Since the project kick-off, an annual progress report and an internal financial report was produced by the E-MST to provide consortium members a periodical project status update.

Among the terms of the INFRAEOSC-04-2018 call to which the ESCAPE consortium successfully applied, we recall the formal commitment from the Legal Entities and Directorates of the involved ESFRI projects and landmarks to engage with EOSC and its implementation, for the adoption of FAIR data stewardship methods and practices for the uptake of Open Science. Motivated by such a high-level commitment, the ESCAPE coordinator has

established the ESFRI supervisory committee, composed of the Directors of the 9 pan-European Research Infrastructures, original founding partners of the cluster initiative as well as among the beneficiaries of ESCAPE. The ESFRI supervisory committee provides policy related inputs to the E-MST. Their feedback is sought primarily while sharing expectations of major astronomy, astroparticle physics, nuclear and particle physics pan-European and world-class flagship RIs on the EOSC implementation with EC, EOSC association Board, ESFRI-EOSC task force, other Science Cluster project coordinators.
An ESCAPE External Expert Advisory Board (E-EAB) is part of the organization and is established in order to provide independent advice to the ESCAPE coordinator, and to conduct an independent assessment of the progress being made by the project. The success of EOSC is related to the capacity of responding to the needs and expectations of researchers, furthermore ESCAPE should build the bridge between the EOSC concept implementation and the EOSC exploitation by the European scientists and fellow citizens concerned by ESCAPE science. Therefore, for the composition of the E-EAB, the ESCAPE coordinator has proposed to rely on the Chairs of thematic consortia of national institutes and agencies of EU member states, namely APPEC, ASTRONET, ECFA, NuPPEC and a representative of ESA; all are concerned by the ESCAPE science. Through such a choice the project coordinator wishes to guarantee that the Board brings additional expertise to the project, comments on its progress and results, brings vision from relevant consortia of thematic national research institutes at European level, their encompassed communities as well as from the pan-European space research field to optimize the activities towards full achievement of the goals of ESCAPE with the most inclusive approach.

1.2.1.c **Deliverables and Milestones**

During RP2	
Deliverables	**Milestones**
	MS38: First periodic report **MS4:** Mid-Term Review + E-EAB evaluation + Acceptance of periodic report 1 **MS5:** E-EAB evaluation + Acceptance of periodic report 2 MS39 Means of verification - Review of deliverables of M21-M36

1.2.1.d **Task 1.1. Governance, coordination and project management**

During this second reporting period, the cluster management has known some evolutions:

i. A technical coordinator joined the E-MST. He was indeed recruited at the end of the first period and he has been contributing to ESCAPE during the full second period, strengthening the cooperation among ESFRI RIs data scientists by organizing dedicated inter-ESFRI projects meetings, supporting and supervising technical interfaces among the work packages, providing cooperative connections with as well

as participating in the technical coordination group of the H2020 EOSC Future consortium.

ii. More recently a turnover of the Chairperson of the E-GA, as foreseen by our consortium agreement at the twenty-fourth month term of ESCAPE project, took place. The E-MST has therefore welcomed a new member previously representing ESO within the E-GA and elected by the E-GA as its Chair.

iii. More recently (towards the end of the second reporting period) with the beginning of the clusters' activities within the EOSC Future consortium, the coordinators of the Test Science Projects (TSP) have been appointed and invited to join the E-EB. This helps to strength progressively the uptake of ESCAPE services by the scientific community concerned by the TSPs.

iv. A new project manager was recruited (April 2021) and a new financial controller (December 2021) joined the team.

v. A turnover of the Chairpersons of the three consortia of national institutes APPEC and ECFA reflected in a turnover of two out of five members of the E-EAB.

In the second reporting period of the project, E-MST organized ten ESCAPE executive board (E-EB) meetings through videoconferencing. ESCAPE general assembly has met once in 2021. Meanwhile, a few more occasions, namely virtual meetings, request of feedback on reports and surveys were conducted with the E-GA as well as with the Directorates of the concerned ESFRIs, in this second period. They concerned mainly the long-term perspectives of the cluster action and the bottom-up requests raised by scientists as well as partner projects to keep alive the ESCAPE cross-fertilization scheme among RIs further than the current H2020 grant duration.

The negative impact of the COVID-19 pandemic during the second period concerns mainly the impossibility to have meetings in person. As a consequence, some residual budget originally dedicated for travelling and events' organization, was partially reassigned or in process of being re-oriented. This was in part the subject of two approved amendment requests (AMD-824064-32) that privileged namely: - personnel costs aiming at extending the work contracts of some key members covering the extra six months of extension of the ESCAPE grant as established earlier as mitigation measure of the first COVID-19 lockdown; - an industry-laboratory cooperation action by supporting further one of our beneficiary SMEs for some innovative workflow developments; - alternative communication tools and initiatives such as broadcast, videos, webinars and virtual forums; - strengthening the Citizen Science plans towards a multidisciplinary inter-cluster action by acting within the EOSC Future WP10 cooperative context.

The E-EAB members attended the ESCAPE annual progress meetings in 2021 and other events. A follow up "question-time" to the project coordinator was also organized upon request of the E-EAB and a review report on the second period activities of the ESCAPE consortium was delivered in January 2022. The E-EAB review and recommendations are listed in appendix 1 of this report. E-EAB members are concerned by the ESCAPE achievements for the benefit of the consortia that they chair and have been active in supporting the ESCAPE work programme evolutions even further than the one established by the current Grant. Namely, they have welcomed the explorative initiative of the "Test Science Projects", addressing transversal open-science research projects concerning the ESCAPE scientific community. TSPs are part of the EOSC Future H2020 project work programme, one of the main commitments of ESCAPE consortium in cooperation with pan-European e-infrastructures

and one highly relevant achievement of the E-MST during this period. TSPs are seen by the E-EAB as *"… fundamental for the success of ESCAPE. TSP are critical to test the central tools under development (data lake and software catalogue, e.g.), while at the same time having a functional role to connect researchers to the EOSC*".

1.2.1.e **Task 1.2 Dissemination, innovation and networking programme**

After the ESCAPE kickoff meeting in February 2019, and a 1st Progress Meeting in February 2020, on the 28th and 29th September 2021 the E-MST organized a two-day annual progress meeting bringing together the H2020-ESCAPE consortium members, ESCAPE General Assembly members, ESCAPE external expert advisory board members as well as members from EOSC stakeholder projects. The event was organized aiming at satisfying as much as possible all level of interest with the maximum inclusiveness towards members, stakeholders and potential users. The first day of the event (https://indico.in2p3.fr/event/24690/) was dedicated to more technical contexts to encourage cross-work-package discussions within ESCAPE project members. The second day (https://indico.in2p3.fr/event/24500/), after a morning session open to all the ESCAPE consortium members and invitees from EOSC stakeholder projects, an afternoon closed session dedicated to E-GA, E-EAB and E-EB followed.

Such an annual general meeting was also the opportunity to report to all beneficiaries about the ESCAPE participation in EOSC Future. That is indeed one major result of the networking action of the E-MST and a major achievement towards the validation of ESCAPE as a long-term thematic platform for open-data science within EOSC.

Members of the E-EAB invite the ESCAPE coordinator to present the project and discuss how to support open science and transversal research activities (within EOSC) with the thematic national agencies. Namely with an invited talk at the 107th Plenary ECFA meeting, one more at the 2021 annual NuPECC Meeting also published (http://nupecc.org/npn/npn313.pdf) as well as at the recent APPEC General Assembly Meeting or the publication on the CERN courier.

This helps the E-MST to receive advice and shape in a more effective way the links between the ESCAPE work programme and the national institutes' expectations, increase inclusiveness of ESCAPE in collecting contributions and interest from RIs that are not yet currently beneficiaries of our project. E-EAB represents a large scientific community with roughly 50000 researchers and provides the feedback from the scientific perspectives on ESCAPE services to enhance researchers' engagement.
Such a networking has facilitated and encouraged new actions that strengthen the ESCAPE impact. Namely:

- The participation of ESCAPE in EOSC Future mainly by supporting two large "Test Science Projects" (TSP) deployed with a number of high-level objectives: a) To demonstrate new cutting-edge science capabilities, in particular those involving cross experiment collaboration and science outcomes; b) 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; c) 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. The two TSPs selected in ESCAPE are: - Dark Matter: to bring together on a common science platform, the data and analysis software of collider, direct and indirect detection experiments involved in ESCAPE; - Extreme Universe: to implement a sustainable platform for multi-messenger astronomy, to enable the use of probes from telescopes and detectors from across the spectrum, and including Gravitational Waves, neutrinos, and cosmic rays.

- Extend the inclusiveness of ESCAPE scopes towards different categories of users (e.g. theorists and students with dedicated education services). Enlarging the networking within the Nuclear Physics research domain by engaging with the new funded Horizon Europe project EURO-LABS. Finally supporting the vision for next generation accelerator-based facilities at CERN such as FCC within the dedicated H2020 FCC Innovation Study project, or the preparatory phase of new ESFRI projects such as Einstein Telescope.

The ESCAPE coordinator has been involved together with the coordinators of the remaining four thematic clusters into the dialogue with the ESFRI chair and the ESFRI-EOSC Task Force members in order to improve the dialogue with the ESFRI RIs, the researchers and organize the ESFRI RIs-EOSC workshops in 2020 and 2022. Networking among the five thematic Science Clusters has been the most relevant action of the E-MST during the review period. (For two years the Science Cluster coordinators have being met constantly every two weeks.) Indeed, the ESCAPE coordinator has been proactive in this context and at the centre of major initiatives:

- After a first Science Clusters Common Statement about their full and long-term commitment to the EOSC implementation in the month of April 2020, the Science Clusters have published their **2nd position statement document** in 2021.
- Such a document was presented during a dedicated **workshop** that was organized by the E-MST. After an open session https://indico.in2p3.fr/event/24327/ where the Science Clusters, ENVRI-FAIR, EOSC-Life, ESCAPE, PANOSC and SSHOC, paved the way towards their long-term commitment in implementing EOSC, a closed afternoon session was dedicated to delegations of the European Commission, the EOSC Association Board, ESFRI Board and representatives of some international research organisations (https://indico.in2p3.fr/event/24328/).

The resilient plans of European countries tackle different domains and priorities. In many of them some funding schemes are dedicated to open data science and digital innovation with the consequent uptake of initiatives very similar to the ESCAPE's one but leveraging national scientific communities. Such national initiatives like the German one PUNCH4NFDI is joining ESCAPE willing to strengthen through their own contribution the longer-term objective of the European cluster. The ESCAPE Coordinator works in this prospective and along the same line with other potential national initiatives and consortia.

1.2.1.f **Plans for future activities**

In the remaining 11 months period of the current H2020 grant, as well as continuing to follow the agreed programme of work, within ESCAPE we are committed to:

- Confirm and extend our action within EOSC Future at three main levels: (i) technology assets on matter such as AAI infrastructure, common metadata EOSC central service provision for scientific software; (ii) citizen science for education and society engagement through data; (iii) transversal science projects for enhancing researchers' participation and exploitation of EOSC.
- Consider to explore and implement further real cross-cutting science use cases in addition to those foreseen by the individual ESFRIs. Support new ESFRI and other RIs and reinforce the dialog with more Nuclear Physics facilities.
- Implement our sustainability plans by engaging with the large community and establishing a new consortium agreement for ESCAPE cluster as a thematic platform for open science data commons and society. A current draft consortium agreement has been validated by all legal entities of partner RIs. The new consortium will be ready to enter in operation in the first semester 2022; a new governance will be established at the end of the current H2020 grant.

One of the most urgent steps ahead of our consortium is the delivery of the first Virtual Research Environment (VRE) prototype. It will be the web-based entry point for open science hosting the current TSP and the gateway to the global ESCAPE as thematic EOSC-cell (figure1). The current work already started in the second reporting period has leveraged the ESOC Enhance project to harmonize the ESCAPE VRE requirements and the EOSC portal goals.
Current networking within the Science Clusters should bring to future engagement in new common projects within the next Horizon Europe framework.
Enhanced coordination with the EOSC Association is expected.
The next period will last less that 11 months and will be the conclusive one. An important community event and blue print will not mark the end of the project but the startup of a second phase of ESCAPE.

1.2.2 Work package 2 DIOS (Data Infrastructure for Open Science)

WP Lead	Xavier ESPINAL (xavier.espinal@cern.ch)
WP Participants:	**CERN**, DESY, GSI, IFAE, CNRS, INFN, RUG, NWO, FAIR, SKAO, SURFSARA.

1.2.2.a **Introduction**

ESCAPE WP2 main objective for the second phase was to assess and evolve the pilot Data Lake that was designed and deployed during the first phase of the project. TThe early pilot Data Lake assessment culminated in a joint exercise labelled as Full Dress Rehearsal (FDR20). By November 2020, the several layers of the pilot Infrastructure were exercised during a 24h production-like window where experiments executed relevant workloads covering a wide range of activities, e.g. from data recording from the experiment detectors/sources to data browsing and access via notebooks for user analysis purposes.

The FDR20 served to confirm that the infrastructure was solid and the concept was addressing the actual needs of the experiments present in ESCAPE. Nevertheless, the FDR20 exercise allowed us to pinpoint areas for improvement and where to put efforts for the following year.

The efforts during this second period culminated in a full scale Data and Analysis Challenge (DAC21) performed in November 2021 enlarging the Data Lake infrastructure consolidation, usability and sustainability.

During the 10 days of the DAC21 exercise the experiments in ESCAPE ran production-like Data Management, Processing and Analysis workloads. This included data acquisition activities from data sources, policy driven data replication and data lifecycles implementation. Data processing was a fundamental target of the DAC21, therefore a big emphasis was put to push different use-cases of processing activities including interplay possibilities using large scale resources (batch systems and clouds) and user-analysis oriented platforms (online notebooks and analysis platforms).

At the DAC21 several experiments deployed and used for the first time private installations of the Data Lake Data Management and File transfer tools demonstrating its seamless integration in a wide common Data Lake Storage infrastructure. Hence demonstrating good potential for sustainability and reinforcing the synergies with the ramping-up activity in the EOSC Future.

There are several aspects worth to mention detail that were fundamental part of this second phase and exercised during several of the DAC21 experiment workloads:

Boosted integration of the **Token-based authentication** integration on the several layers of the Data Lake infrastructure: Rucio, FTS, storages and potential integration with other AAI providers. Some end-to-end workflows ran integrally using token-based authentication during DAC21, including the data access and processing from analysis platforms.

Data life-cycle accommodation with the ability to define replication rules and policies: data redundancy, location and an eventual mapping to storage quality of service provided at the sites to adjust data popularity needs and value to the right storage resource-types and cost.

Webdav/HTTP capabilities promoted to be the de-facto standard in the Data Lake. The widespread knowledge of HTTP protocols also provides a flexible way to interact and integrate with other storage resources.

Rucio Evolution. In WP2 we have been gathering and providing feedback from the new scientific communities using Rucio. This interest and feedback is being taken into consideration by the Rucio core team and evolving the service to cater for these new scopes and needs. There is a special interest from astroparticle, astronomy and cosmology communities to enlarge and expand Rucio's metadata capabilities. A WP2 team member has now become the metadata technical lead within the Rucio project.

Enlarged **Data Lake monitoring** capabilities. The monitoring platform provides a good overview of the ongoing activities in the infrastructure, such as in-flight transfers and scheduled data movements, but also actual usage of resources in terms of storage by service provider, and by experiment.

Active **Deployment and Operations team** (DepOps). Early in the project we identified the need to share experience with the tools and the operations. The DepOps team is formed by experiments and site representatives and coordinated via a well-established meeting, demonstrated to be instrumental in fostering knowledge transfer and expertise sharing. DepOps teamwork was crucial to drive the preparations for the DAC21. The responsible chairs of these meetings were rotated among different people in the experiments and/or sites.

One of the missions during this phase was to prove that the Data Lake "big-data-portal" could be integrated also into a much narrower "just-a-bunch-of-files" user-perspective. This aimed to develop a Rucio extension allowing to connect the Data Lake infrastructure with the several analysis "machines" a user might be using. In this spirit, we developed a Rucio extension pluggable to vanilla Jupyter notebooks, allowing us to browse and download data in the ESCAPE Data Lake from the web browser/notebook. New improvements and ideas resulted in further development and integration of the Data Lake capabilities with the user environments, and finalised a product labelled as "**Data Lake as a Service**" (DLaaS), providing increased browse/download/upload data capabilities, token integration, data movement within the notebook, integration of local storage, leveraging content delivery and caches, etc., allowing end extending the integration possibilities with Analysis Platforms, computing infrastructures, and resources local to the services provider.

Integration of heterogeneous resources has been demonstrated. The ESCAPE Data Lake inherent flexibility enables heterogeneous data processing facilities in the Data Lake, activities with commercial clouds, the implementation of Swift/S3 storage endpoints in the Data Lake, and activities with CINECA/HPC, and we started a collaboration with the FENIX/HPC project.

The different activities briefly summarized above drove the momentum of the WP2 community during the last 18 months and served to achieve a steady state of the ESCAPE Data Lake

infrastructure able to perform and deliver over a broad range of activities, finalised in the DAC21 exercise where experiments could put at work production-like activities covering its various data management needs and data processing workflows.

1.2.2.b **Organisation**

The organisational structure of work package 2 is following the five tasks below :

- Task 1: Data Lake infrastructure and federation services
- Task 2: Data Lake Orchestration Service
- Task 3: Integration with Compute Services
- Task 4: Networking
- Task 5: Authentication and Authorization Mechanisms

1.2.2.c **Deliverables and milestones**

During RP2	
Deliverables	**Milestones**
D2.2: Assessment and analysis of performance of the first pilot data lake	**MS9:** Second WP2 workshop to analyse the performance of the pilot, prepare D2.2 **MS10:** Expanded prototype - more data centres including 3rd party centres, demonstrate integrated data management tools, verify RI data accessibility from compute platforms including commercial clouds **MS11:** Extension of the Data Lake to efficiently serve data to external compute resources providers

1.2.2.d **Task 2.1: Data Lake infrastructure and federation services**

The deployment of the Data Lake infrastructure, related services and tools have undergone significant upgrades during the 2nd reporting period. The deployment moved from OpenStack virtual machines (VMs) and Puppet-managed to Kubernetes (k8s) allowing to implement redundancy for the key Rucio components.

This ability to deploy, customize, and manage services at convenience has proven to be a successful approach and allowed ESCAPE experiments to start deploying and operating their own Rucio instances.

The ESCAPE Rucio instance follows the last minor and major Rucio releases, profiting from new functionalities and features, which many times come from contributions of the ESCAPE partners. The improved project sustainability and successful exportability adoption of the common infrastructure are in line with the goal to allow them to operate such services and tools according to their scientific needs beyond the ESCAPE project term.

The Data Lake has progressively increased storage capacity as in the number of Rucio Storage Elements (RSEs) federating 28 storage endpints provided by 10 partner sites, totalising a shared capacity of ~1PB and currently hosting 4.4M files from the participating ESFRIs. The storage endpoints are heterogeneous in terms of size and technology (EOS, DPM, dCache, StoRM, XRootD), overall supporting HTTP, xroot, and GridFTP access protocols. Moreover, efforts on the integration of heterogeneous and opportunistic resources, such as HPC elements, and AWS and Openstack-based storages have been carried out.

Specific testing-focused time windows have been identified to assess the robustness of the Pilot and Prototype phases of the various Data Lake components, tools, and services: the 2020 Full Dress Rehearsal (FDR20) and the 2021 Data and Analysis Challenge (DAC21) exercises. The goal was to demonstrate the fulfilment of communities' needs and the addressing of required functionalities and experiment-specific use cases.

A cross-WP collaboration has seen WP2 contributing to ESAP WP5 with a ready-to-be-used product named DataLake-as-a-Service (DLaaS). The focus was on further integration of the Data Lake, data access, and the related data management capabilities with the activities ongoing in the area of Science Platforms. The CERN team focused on developing a state-of-the-art “data and analysis portal” as an interface to the Data Lake offering a wide range of possibilities, from very simple I/O access to more complex workflows such as enabling content delivery technologies and integration local storage facilities at the facility hosting the notebook. The DLaaS project allows end-users to interact with the Data Lake in an easily-understandable and user-friendly way, and it is based on the JupyterLab and JupyterHub software packages. The Rucio JupyterLab software package is used to integrate the service with the ESCAPE Rucio instance. The DLaaS was built on top of JupyterHub running on CERN OpenStack dedicated resources, and deployed on the same k8s cluster hosting the other Data Lake services and tools. Examples of the features of the DLaaS include token-based OpenID Connect authentication to ESCAPE IAM, data browser, data download and upload, local storage backend access to enlarge scratch notebook space, multiple environment options, and a content delivery low latency-hiding data access layer based on XRootD-XCache.

1.2.2.e **Task 2.2: Data Lake Orchestration Service**

One of the main outcomes of this second period is that the way to address Storage Quality of Service has been understood and coded in Rucio. This provided the mapping of QoS classes with individual storage endpoints (Rucio Storage Elements, RSEs), and enabled the experiment's work-flows to target specific QoS, without requiring knowledge of which RSE. As a consequence the QoS strategies of the various ESFRI communities could be implemented during the DAC21 and data life-cycles exercised. In particular there was a close collaboration with LOFAR in developing their QoS strategy. Helped identify possible cost-saving through

provisioning different storage for data ingress and data egress, as they have different QoS requirements.

During a series of meetings, we worked collectively to build up the QoS use-cases of various ESFRI communities through which the QoS concept could be demonstrated during DAC21. Undertook dedicated meetings with FAIR to discuss participation in DAC21. This covered various aspects, including to what extent it would be possible to include data from the CBM test detector (miniCBM) during DAC21.

Preparatory work for the Data and Analysis Challenge (DAC21) included two additional QoS endpoints in the Data Lake:

Tape storage: tape resources were provided by INFN/CNAF, PIC, SurfSARA and DESY. Access to these sites was facilitated by adding new RSEs into the ESCAPE testbed. These new RSEs were manually tested. To alleviate concerns from partner institutes, the automated testing machinery was adjusted to avoid writing test data.

Erasure Coding: Enabled Erasure Coding storage in the CERN's Data Lake storage endpoint, allowing ESFRI communities to explore the impact of erasure coding on their workflows.

Close collaboration with the CERN team on improving the monitoring, in particular the Grafana dashboard, based on earlier feedback from LSST during FDR. The joint work has been organised within a true collaboration spirit and continuous feedback. As a result, the QoS monitoring abilities was largely improved and fundamental for the DAC21 exercise.

Alsdo during this period the integration of external Cloud storage (based on S3 protocol) was demonstrated with a commercial cloud provider (Amazon) and with a private cloud provider witrh surfSARA ESCAPE.

1.2.2.f **Task 2.3: Integration with Compute Services**

From the point of view of interaction with the compute resources infrastructure, the main focus during the reporting period has been to work on the integration between the Science Analysis Platform developed in WP5/ESAP and the Data Lake. One fundamental part of this integration has been the DLaaS. From the integration point of view, ESAP has been extended with Rucio query functionality, meaning that data in the Data Lake can be put in the ESAP shopping basket. The contents of the shopping basket can be read from a python library, which also makes it possible to define specific connectors for data collections. One of those is the Rucio connector.

Another important part of the general integration of the Data Lake has been the further development of OpenID Connect authentication and authorisation (further details in Task 2.5 section), making the link between the Analysis platform, the compute resources and the Data Lake more streamlined and accessible for a broader community.

All the functionalities discussed above have been demonstrated in the DAC21, for instance in one of the LOFAR use cases that consisted of querying data from multiple collections (including the Data Lake) in ESAP, running a DLaaS instance for further processing, obtaining

the data listed in the shopping basket, processing it, and put the result back in the Data Lake, all using a single user login through the ESCAPE IAM infrastructure.

1.2.2.g **Task 2.4: Networking**

Over the reporting period, the work within the Networking task has focused on three main areas 1) developing the throughput monitoring suite and improving monitoring dashboards 2) expanding support and uptake for additional data transfer protocols and understanding of third-party copy capabilities and limitations and 3) scaling out ESCAPE partner Data Lake prototypes to include longer-distance transfers.

Monitoring throughput: We have continued to develop and maintain the network components of the testing stack. For the assessment of the pilot Data Lake this included deployment of PerfSONAR hardware at all sites and the development of an ESCAPE Data Lake PerfSONAR testing matrix.

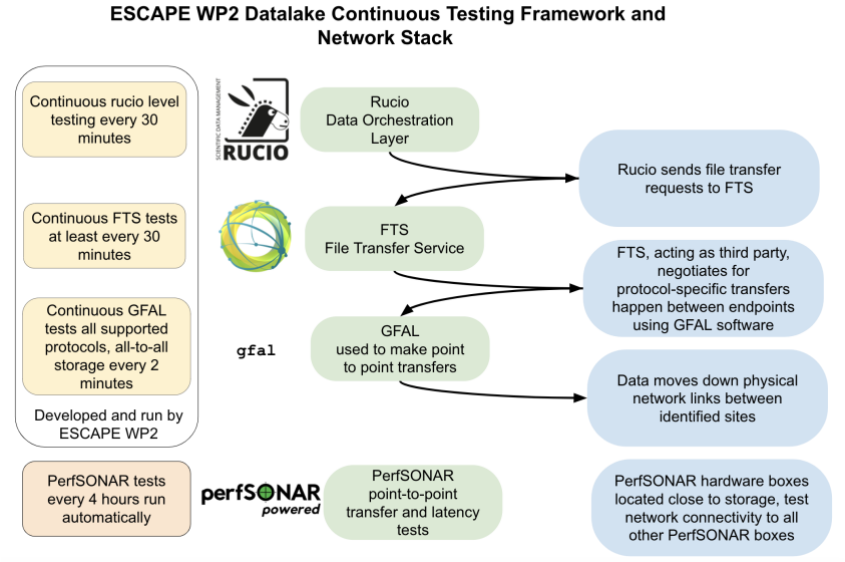

Fig. 2 Representation of the nested data transfer stack and the continuous tests running at each level during the FDR exercise. Note that at each level delegation of the task is passed down to the level below - Rucio instructs FTS which uses GFAL to make transfers. Testing at each layer in the stack is essential to understand failures. At the base level, perfSONAR tests the link health

Maintenance and creation of new dashboards: created a new Rucio Stats dashboard that reflects the statistics of the replicas in an interconnected way. That interconnection provides the possibility to get more details while filtering. The user can use as many filters as needed (RSE, experiment, QoS…) and the replicas metrics are shown in different panels depending on that filtering.

Main activities carried out as DepOps meetings - weekly health check, issue management:

- Multi-protocol support, DOMA Third Party Copy working group
- Long-haul transfers: although long-haul transfers within the WLCG are commonplace, confidence is still lacking in the ability of the Data Lake orchestration stack to handle

intercontinental transfers in non-HEP applications; work in ESCAPE WP2 has greatly improved. The ESCAPE team at SKAO has deployed its own Rucio instance, with support from the ESCAPE Data Lake team and from ESCAPE storage sites. The Rucio service for this runs on a kubernetes cluster. Rucio endpoints in the UK (harnessing Grid sites in Manchester and Lancaster), in Germany (at DESY) provide the European "receiving" sites to complement "remote" storage elements in the two SKAO host countries at the IDIA (https://www.idia.ac.za/) site in Cape Town, South Africa and at AARNET (https://www.aarnet.edu.au/) in Perth, Australia. The SKA Rucio instance was first deployed in March 2021 and since then the team have expanded the work to maintain both a production and a development instance, and implemented automated, continuous transfer tests to enable easy monitoring. In the months leading to DAC21, these tests were augmented to include throughput measuring tests using a range of file size (50MB, 500MB and 5GB) for transfers from each of the two "remote" locations into Europe. The team has found that the ESCAPE Data Lake sw stack continues to perform well and to show resilience. The HEP specificity of some elements does present some barriers to onboarding new, non-HEP storage sites, so the recent progress with token-based flows will significantly improve participation opportunities for more sites at representative SKA Regional Centre locations (e.g. Canada, India) in the years immediately following the ESCAPE project.

ESCAPE team members at MAGIC / PIC also deployed a custom Rucio instance and have been using it to test observatory-to-data centre replication flows across 2000+km from La Palma to Barcelona. This has enabled the CTAO team in ESCAPE WP2 to also test a range of use cases and strengthens the interest in the use of Rucio to support very long distance transfers, since CTAO's sites will be at La Palma and also at Paranal in Chile, with offices in Europe.

Furthermore the Rubin Observatory (VRU, LSST) project is assessing the suitability of Rucio to support long-distance data management between partner sites in the USA and in Europe, so the work done within WP2 is likely to lead to the development of a strong community of world-class astronomy projects using ESCAPE.

1.2.2.h **Task 2.5: Authentication and Authorization Mechanisms**

During the reporting period, the work on authentication and authorization mechanisms focused on the following tasks:

- Providing a reliable central authentication and authorization service deployment (the IAM ESCAPE instance) supporting X.509/VOMS and token-based authentication and authorization.
- Evolving the INDIGO IAM codebase to fix bugs, introduce support for a highly available (HA) deployment and honour requirements emerging in the ESCAPE context.
- Defining the strategy to support token-based authentication and fine-grained authorization on ESCAPE Data Lake, by proposing an incremental approach assessed by a continuous and automated verification process based on an integration test suite.
- Participating in the definition and implementation of DAC21 Activity 3, focused on supporting DAC21 AAI use cases.

- Developing and deploying the ESCAPE authN/Z test suite to verify ESCAPE Data Lake compliance with agreed upon authentication and authorization shared policies.
- Supporting cross WP AAI integration activities, organizing training events and providing expert advice in the project communications channels.
- Participating in the DepOps groups looking after the Data Lake sites health and helping sort out AAI-related issues.
- Investigating and realizing a prototype replicated highly available (HA) deployment of the INDIGO IAM and assessing its reliability.
- Participating in the definition of the token-based authentication and authorization flows in support of an AAI interoperability exercise aimed at demonstrating data-transfers between ESCAPE Data Lake and a FENIX project HPC centre.

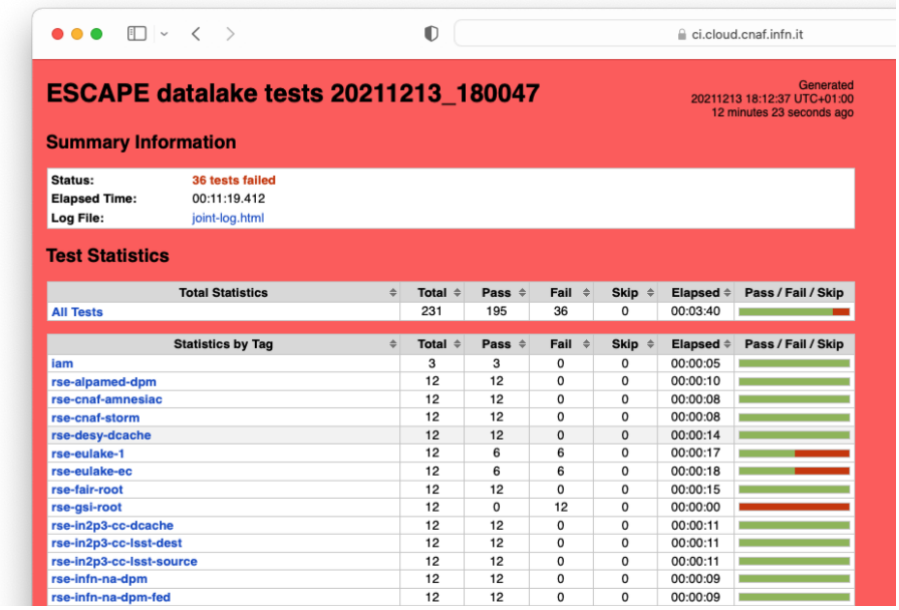

Fig. 3 Some of the results of the token-based authN/Z automated test suite results on the datalake (live monitoring)

1.2.2.i **Joint Cross-Work Package Activities**

After certifying the maturity of the Data Lake orchestration technologies through the FDR20 exercise effort has been made into strengthening cross work-package activities. The joint work with the ESCAPE Analysis Platform (WP5) being the most relevant to boost the interaction between data and computing oriented towards user analysis.

With this spirit in mind a joint workshop DIOS and ESAP was organised to identify common areas of interest and to make joint efforts. The Data Lake as a Service, providing further integration of the Data Management capabilities, is a good example of a driver to integrate the different layers (AAI, Storage QoS, Data browsing/uploading/downloading, etc.) a scientist would like to have at hand when logging-in to the work environment. This integration has been reflected in the recent DAC21 exercise where many of the experiment's use-cases included

end user data processing making use of the several developed and deployed Analysis Platform flavours with different options tailored to the needs.

Progress has been made towards understanding the integration with external software repositories, i.e. experiment specific software, code, calibration data, etc. The final integration between the different ESCAPE repositories in OSSR (WP3) and the scientific workflows that run in the FDR20 and the DAC21 has not yet been fully demonstrated. The plan is to converge on integration software access methods in order to envision a single entry point for the users where: to find and access the data, exploit the cpu and the analysis tools, and access to required software. This is being shaped under the concept of a VRE (Virtual Research environment).

The understanding between the ESCAPE Virtual Observatory (EVO) and the ESCAPE Data Lake communities progressed considerably. One has to keep in mind that the approach to Data Management and the needs of these communities are quite different in terms of structure, usage and needs. Nevertheless we started observing and discussing certain commonalities. During the ESCAPE Extended Discussion Day prior to the ESCAPE annual Progress Meeting, there was a specific session on describing the ESCAPE Data Lake infrastructure, the concept and the usability. As a result there was constructive discussion at the meeting and afterwards between the ESCAPE Data Lake experts, the EVO and the ESA scientists addressing fundamental conceptual points about Data Management, File Catalogues and Storage. A common technical workshop is envisaged for next year, to start identifying common points to explore options for future collaborations. As a result, there is also the intention to evaluate the feasibility to integrate one of the EVO data repositories (HIPS) in the ESCAPE Data Lake.

1.2.2.j **Synergies with external Projects**

In these past 18 months ESCAPE WP2 set some focus to cover additional aspects that were identified as potentially relevant extensions of the project original goals. The encouragement to pursue these extra activities was gathered during the first Project Review and via fundamental feedback in several workshops and meetings. The Data Lake orchestration layer and the tools rapidly proved to give a solid response during the initial phase of the project, the experiments managed to port workflows in the early prototype as demonstrated during the FDR20 exercise. This resulted in the adoption of the core Data Management services by two of the ESFRIs already in this second phase. The identified relevant extensions were geared towards widening the Data Lake flexibility to demonstrate its usability and benefits for scientific communities with more punctual and less demanding storage and computing demands, and also to enhance its usability for end-user analysis, Open Data and also as enabling technology able to foster Teaching and Outreach initiatives.

Photon and Neutron Sciences:

PaN user communities are very diverse and wide-ranging in their requirements, the commonality is to have an increasing amount of data and a need for distributed computing. The ESCAPE Data Lake model and vision, with its Data Management and File Transfer

service, was shared with these communities. There is an opportunity to use the Data Lake as a simple service, as the complexity can be much reduced depending on the use case.

Together with PANOSC and EXpands we organized a joint workshop to understand the scientific data management needs of the PaN community on 12th January 2021: "PaN ESCAPE Data Management Workshop". Next steps should be geared towards prototyping a real use case to: store, distribute and access data from PaN source.

FENIX HPC project: integrating heterogeneous computing resources:

Exploitation of heterogeneous resources, cpu cycles and storage, have been addressed and successfully integrated with the Data Lake as mentioned before in the document for the CMS/CINECA and AWS/Google cloud resources integration. But besides these punctual Proof of Concept exercises we started a well established collaboration with the FENIX project [link] composed by six European supercomputing centers: BSC (Spain), CEA (France), CINECA (Italy), CSC (Finland), CSCS (Switzerland) and JSC (Germany) which agreed to align their services to facilitate the creation of a common Infrastructure. The synergies between FENIX and ESCAPE WP2 are worth exploring and with this aim we started a joint collaboration. The goal is to provide a recognisable infrastructure with a framework of common tools to enable sciences to make use of HPC resources with Data Management capabilities to be able to access data, and steer/move data the HPC facilities. To achieve this we need the AAI infrastructures to be synchronized and to connect the internal HPC storage with the Data Lake storages where the experiments place the data. In the last months this advanced quite a lot and FTS have Swift/S3 capabilities ready while the ESCAPE IAM and the FENIX AAI systems are close to trust each other, permitting transparent access and data flow for registered users in either of the systems. This activity will be certainly pursued after the ESCAPE project in the framework of EOSC-Future.

CS3MESH4EOSC: sync&share platforms integration:

Concerning the CS3MESH4EOSC project, a collaboration started to integrate data transfer services based on RUCIO/FTS and the ESCAPE IAM into _ScienceMesh_, the new federated infrastructure being built within CS3MESH4EOSC based on storage sync&share platforms. To this end, FTS-driven HTTP Third-Party transfers were demonstrated between nodes of ScienceMesh, and in the coming months it is foreseen to provide a full integration between ESCAPE and ScienceMesh sites, such that managed data transfers can seamlessly take place across them.

EOSC-Future project: an ESCAPE legacy:

The successful ESCAPE model is already being integrated into an early stage development of the European Open Science Cloud "EOSC-Future" project. The synergies between the two European projects, both targeting scientific collaboration across diverse platforms, are being exploited in various ways.

The ESCAPE Data Lake infrastructure's functionalities, namely data injection, replication, processing and monitoring, are being extended to serve the Astrophysics scientific community as well as the High Energy Physics one, providing a reliable AAI framework and the federated storage services with the required Data Management needs. The EOSC Future Project is aimed at bringing together diverse analysis workflows from both communities, demonstrating how a common platform could support the goals of the Dark Matter Science Project and the Extreme Universe Science Project in the respect of FAIR data policies. Such platform is being implemented as a Virtual Research Environment (VRE), an ecosystem with different complexity levels enhanced by a seamless Web UI interface, where both scientists and new onboarding users will be able to run and reuse analysis workflows. This will be achieved fetching the data from the Rucio Data Lake, using the updated software version preserved on Zenodo, and retrieving the computational environment by leveraging the power of REANA, a reproducible research analysis platform initially developed at CERN, and of RECAST, the tool which enables the automation of signal passing through an analysis at the development time.

The ESCAPE DataLake-as-a-Service model will be utilized as a springboard to spawn flexible re-analyses with the possibility to incorporate distributed hybrid workflows thanks to the synergies between file system handling specific to Reana and FTS. The VRE will then provide the common hub to run workflows on external resources, taking advantage of the flexibility of containerized applications.

1.2.2.k **Outreach and communication**

- Invited presentation to PUNCH4NFDI: "*Data Lakes in the PUNCH sciences"*
- PaN ESCAPE Data Management Workshop
- ESCAPE ESAP AAI Workshop *"Integrating services with the ESCAPE AAI"*
- ESCAPE ESAP AAI Workshop *"OIDC support in RUCIO"*
- International Conference on Computing in High-Energy and Nuclear Physics (vCHEP 2021) *"The ESCAPE Data Lake: The machinery behind testing, monitoring and supporting a unified federated storage infrastructure of the exabyte-scale"*
- World-wide LHC Computing Grid (WLCG) - Grid Deployment Board (GDB), "ESCAPE Full Dress Rehearsal Exercise results"
- International Symposium on Grids & Clouds 2021 (ISGC 2021), *"ESCAPE, next generation management of exabytes of cross discipline scientific data"*
- International Conference on Computing in High-Energy and Nuclear Physics (vCHEP 2021), *"ESCAPE Data Lake: Next-generation management of cross-discipline Exabyte-scale scientific data"*
- CERN Joint Computing Seminar - ITTF Session, *"Data Lake as a Service for Open Science"*
- 4th Rucio Community Workshop, *"ESCAPE Data Lake as a Service"*
- EGI Conference 2021, *"The ESCAPE Data Lake as the bridgehead for the EOSC"*

1.2.2.l **Plans for future activities**

The primary target for the next months in WP2 is to analyse and discuss the broad range of activities and results from the DAC21 exercise. This will be done at the 3rd WP2 Workshop (project milestone). This gathering with the WP2 community will allow us to analyse the successes but most importantly identify potential improvements for the future to ensure a proper transition after the project ending. There are several items that make us think the ESCAPE activities and legacy will be not stopping after the project, on the contrary:

- ESCAPE and EOSC-Future co-existence, synergies and transition: the ESCAPE Data Lake services, infrastructure and tools will be the basis for the EOSC-Future Science Projects.
- Consolidation: several experiments already deployed Data Lake orchestration tools (RUCIO and FTS) for potential addressing their upcoming Data Management needs.
- ESCAPE WP2 is leading the activity for commissioning a new AAI scenario in HEP (x509 migration to token-based AAI)
- ESCAPE WP2 is leading the activity to address the file metadata needs of non-HEP sciences, this is reflected in the involvement with the RUCIO development team and the joint interest to pursue a system that can deliver the required needs for a broad range of sciences
- The ability demonstrated by a Data Lake architecture is of special relevance for the usage in heterogeneous resources, the ongoing collaboration with the FENIX/HPC project will persist after ESCAPE and within EOSC-Future.
- Large room for collaboration with sciences with well established data infrastructures, historically separated from HEP but with bigger commonalities than initially supposed. Synergies should be further explored and pursued in the next months with the astronomy community: Virtual Observatory community and ESA in particular. A common technical workshop will be organised before the end of the project.

1.2.3 Work package 3 OSSR (Open-source scientific Software and Service Repository)

WP Lead:	Kay GRAF (kay.graf@FAU.DE)
WP Participants:	AIP, CERN, CNRS, CTAO, EGO, **FAU,** GSI, HITS, IFAE, INFN, JIV-ERIC, MPG, NWO-I, OROBIX, SKAO, UCM, UNITOV

1.2.3.a **Introduction**

The aim of WP3 is to expose the open science software and services of the ESCAPE ESFRI projects in a repository under the EOSC catalogue of services, ensuring compatibility with FAIR principles.

The objectives are to:

- Facilitate and support continuous **development, deployment, exposure and preservation** of partners' software, tools and services;
- Foster **interoperability, software re-use and cross-fertilisation** between ESFRIs (e.g. for simulation software);
- Offer an **open innovation environment for open standards** (e.g. for workflows and data-formats), **common regulations** and **shared (novel) software** for multi-messenger & multi-probe data.
- **Enabling open science** - with **software as "first class citizen";**
- Implementing a **community-based, inclusive** approach;
- Following **FAIR principles** for open science resources software and derivatives;
- **Federating** available resources of the partners and research infrastructures.

The ESCAPE catalogue implemented in WP3 is central to the ESCAPE-EOSC thematic cell. It does not only highlight and serve the users - i.e. the community scientists - as the centre of EOSC, but also establishes links to other communities, as the software and services collected will form a knowledge base that can be re-used and stimulate cross-fertilisation across domains.

The ESCAPE catalogue will form a single entry point to the thematic cell for test data sets, software and services, tutorials, trainings and publication; federating and pointing to the individual ESFRI and RI services. This approach is a result of the collection and classification of ESCAPE user, partner and ESFRI requirements and coordination with the EOSC community.

During the second reporting period the OSSR has been prototyped in close co-operation with all the project partners, as well as other communities. The work in this period as well as the plans for the last reporting period are summarised below.

1.2.3.b **Organisation**

The goals of WP3 - OSSR are defined by the deliverables and milestones of the work package.

With the first reporting period, the start-up phase of the project was concluded. The work package project plan was set-up including the consensus and input of all 19 work-package partner institutes. The organisational structure is following the five tasks defined in the DoA, the roles of the partners and the collaborative environment have been defined and established. Each task has a dedicated task leader representing one major partner institute of WP3.

The aims of the tasks, following the objectives, have been defined. All the tasks have been fully implemented, all of the foreseen critical objectives have been achieved:

Task 3.1: Management Activities, Policy and Support Actions (MAPS): Transversal activities on management, policy and support for OSSR. Lead by FAU.

Task 3.2: ESFRI Software and Services Collection (ESSC): Collection of software, evaluation and exploitation of common practices in software development, maintenance and distribution. Lead by FAU.

Task 3.3: Common Approaches: Software and Services (CASS): Sharing existing solutions and fostering joint development of new tools for specific analysis, simulation and processing tasks. Lead by INFN.

Task 3.4: Foundation of Competence for Software and Service Innovation (COSSI): Creation of a team of scientists from different ESFRIs who will investigate new, innovative approaches to data analysis, exploring machine-learning and deep-learning techniques with a special focus on the multi-messenger approach. Lead by EGO.

Task 3.5: Repository Implementation and Deployment (RIAD): Implementation of a trusted digital repository to integrate open-source scientific software and services developed by ESFRI in tasks 3.2-3.4 and integration of this repository in the EOSC catalogue of services. Lead by LAPP.

For an efficient organisation of day-to-day work in an open knowledge-sharing environment, focus groups have been successfully established. The work of the focus groups is community-driven on most important aspects of implementing the projects' objectives, currently on the collection of the community software, the technical implementation of the OSSR repository, innovative workflows, and common approaches to joint development of scientific software.

The resources and contributions of the partners are monitored by partner feedback collection on a half-yearly basis. The currently available resources are deemed adequate to reach the aims and objectives of the work package. The main contact persons of all 19 partners are identified and actively involved in the project progress. All partners are well linked to the working group, following the activities and meetings as well as contributing their work. Due to the generally good economic situation in the EU and later by the COVID-19 pandemic, several hiring processes have been delayed at the start and some drop-out personnel could not be replaced immediately, however, without significant impact on the project.

Some challenges that arose in the beginning of the project have been solved. Still, the high number of partners and diversity in software development and deployment techniques in the different communities linked in ESCAPE is an issue that need constant attendance within the focus groups.

To foster the open development environment and knowledge sharing, appropriate IT services (mailing lists, meeting-, cloud-, chat- and wiki servers) have been set up to share all relevant information (meeting agendas, presentations, minutes of meetings, etc.) with all partners. In addition a project platform was established to steer the use cases across all ESCAPE work packages as well as follow the onboarding procedure for the OSSR. The entry point for information on OSSR is its wiki page.

1.2.3.c **Deliverables and Milestones**

During RP2	
Deliverables	**Milestones**
D3.6: Mid-term technology WP3 project progress	**MS16:** Software and service repository demonstrator
D3.4: Establishing of innovation competence group via a workshop	**MS15:** Establishment of innovation competence group
D3.5: Thematic training event - 1st school	**MS17:** Progress of common software and service proposition
D3.7: License and provenance model for the repository	

1.2.3.d **Task 3.1: Management Activities, Policy and Support Actions - CNRS-LAPP, FAU, MPG-MPIK**

Task 3.1 implements the organisation of the work package as described above. It is led by FAU providing the WG coordinator and project scientist. Constraints and interfaces have been defined between the work packages of ESCAPE and within WP3, the main responsible persons acting as interfaces between OSSR and the ESFRI's have been identified as well. The following links have been established:

- within ESCAPE:
  - A helpdesk for all partners within ESCAPE (by project scientist) has been setup;
  - The technical coordination group is followed by the WG coordinator;
  - Frequent conversations with the other WP coordinators in the E-EB are held.
- within EOSC environment:
  - WP3 was represented in the EOSC Architecture Task Force and is in the EOSC Infrastructures for Quality Research Software Task Force by the WG coordinator. Here, it was possible to establish the software as an important ingredient to open science products on par with data.
  - The coordination with other EOSC projects (EOSC Enhance, EOSC Future) is maintained.
    - Especially also the technical implementation of the test science projects as cross-project activities between ESCAPE and EOSC-Future (Extreme Universe and Dark Matter) are supported by Task 3.1;

    - Links to OpenAIRE, Software Heritage and Zenodo have been established.
- to national projects:
    - Punch4NFDI (German national scientific data management project) - OSSR, data-lake and analysis platform discussions on PunchLunches were supported.
    - An agreement with the HGF (German Helmholtz Institute funding) DMA project to use the OSSR infrastructure as software repository has been reached, the first onboarding has started in 2021.

Also, the communication and outreach activities of WP3 are organised by Task 3.1, for details see the report of WP6. The most notable organisation where that of a webinar on the open science benefits enabled by OSSR (use cases from different ESFRIs[1]), the IWAPP workshop and the ESCAPE Data Science Summer school[2] that primarily organised by CNRS-LAPP, FAU and MPG-MPIK.

1.2.3.e **Task 3.2: ESFRI Software and Services Collection - CERN, CNRS-LAPP, CTAO, FAU, GSI, INFN, JIVE, MPG-MPIK, SKAO**

In task 3.2, the development, benchmarking and deployment of software within and across partners has started as well as gathering of common practices and know-how towards the definition of best practices to be shared with the community. A first round of software to become part of the repository is collected, the workflow for software onboarding from ESCAPE partners set up, and the partners have started to prepare the software for the latter purpose.

The software onboarding workflow to OSSR contains a survey, a presentation of the software and an (non-obligatory) technical report, all of the contributions will be collected on the landing page. The workflow is followed via issue on the ESCAPE project platform, set up and supported by FAU. This workflow is extended and cross-checked with the DMA software by GSI. 16 onboarding processes started out of which 8 of finished up-to-date. The checklist for ESCAPE partners is available at https://escape2020.pages.in2p3.fr/wp3/ossr-pages/

The described workflow will be extended to external partners and streamlined, also the terms of use as coordinated with EOSC-Future will be implemented. The onboarding is supported by a group of experts, primarily from CNRS-LAPP, CTAO (that is also coordinating the CTA contributions), FAU, GSI, INFN, JIVE and MPG-MPIK.

During the reporting periods, the partner contributed and developed the following software and services:

- CNRS-LAPP developped *hipeRTA*, a high-performance analysis software for the event reconstruction of CTA. The library is being integrated in the CTAO ACADA framework and is being deployed and tested on the first CTA prototype telescope being built on-site, processing the full rate of LST-1 data in real-time.

[1] *https://projectescape.eu/events/webinar-escape-ossr-enhancing-science-through-sharing-software-benefits-use-cases*
[2] *https://projectescape.eu/events/escape-summer-school-2021*

- GSI onboarded the software package *FairRoot*, *FairMQ* and *DDS*. A series of meetings has been organized and held, dedicated to joining efforts of MT-DMA[3] Subtopic 2 ("The digital scientific method") and ESCAPE OSSR. It was agreed to onboard DMA projects as a domain in OSSR. GSI also took over the management process of the DMA onboarding, which is currently planned to be performed within the activities of Focus Group 1 of WP3. The onboarding of the first DMA project R3BRoot has started.
- IFAE finished onboarding of *agnpy* - a python package for modelling the radiative processes in the jets of active galaxies to be used e.g. in combination with gammapy (MPIK's contribution) for on CTA within the Data Challenge 2021 (DAC21) and of gLike, a ROOT-based software package for maximum-likelihood analyses of heterogeneous data samples. gLike was used as a showcase for the ESCAPE General Assembly meeting in September 2021.
- JIVE progressed towards improving the CASA VLBI tools. Requirements, regression tests and documentation were developed and included in the CASA repository and regression-test environment. The Jupyter-casa kernel was updated to be built around the 6.4.0 release and the OSSR onboarding process for CASA/the Jupyter casa kernel has started.
- MPIK works on developing open source high-level science tools and data formats for CTA and astroparticle physics in general. Several releases of the Gammapy package have been issued which available on Zenodo and was selected as the of official Science Tool for CTA. To foster the community foundation aspect, independent "coding sprint" events were organized at a frequency of two per year since 2019 (remotely in 2020/2021), gathering also external participants.
- SKAO's further developed the reproducible workflow based around the SKAO's first Science Data Challenge (SDC1), updating the repository to run on a newer BinderHub instance. The advantage of this approach is that it renders modifications at the platform level unnecessary; BinderHub automatically generates the notebook environment, with requisite software dependencies, from the repository itself. This repository has also been integrated with the EOSSR, by releasing a stable version of the code on Zenodo.

1.2.3.f **Task 3.3: Common Approaches: Software and Services - CNRS-CPPM, FAU, GSI, IFAE, INFN, MPG-MPIK, UCM**

The initiative to enhance and simplify production of cosmic ray simulations, envisaged in the previous reporting period, has been onboarded with the name of ConCORDIA (Containers for CORSIKA on DIRAC). It includes multiple coordinated actions for using well-qualified simulation tools to be published through the EOSC. First, a library of ready-made containers based on the CORSIKA cosmic ray shower simulation software is provided; each container comes with a small test production which serves as a validation sample and to provide quantitative performance estimators, from both points of view of physics and computing. As a second action, choosing an existing container or creating a new one is made easier by means of a Web Graphical User Interface (GUI) and/or a dedicated scripting interface. Providing a

[3] *Matter and Technologies (MT) is a German national research program within the Helmholtz Association, that combines development of both accelerators and detectors into one program including the work package Data Management and Analysis (DMA)*

distributed computing environment to manage container creation and simulation jobs is the third action: the ConCORDIA application has been developed using the widely used DIRAC middleware. Finally, the fourth action is to integrate DIRAC, and hence ConCORDIA, in the ESAP analysis platform developed in WP5, consistently with other software and services produced in ESCAPE. Short-term and long-term storage of simulation outputs in the ESCAPE Data Lake via RUCIO is envisaged and will be the subject of further study and development. Most contributions are coming from INFN, CNRS-CPPM, IFAE and CTAO.

The classical CORSIKA simulation package (Corsika 7, FORTRAN version) is receiving a new development contribution from Task 3.3 in the shape of original code for fluorescence radiation simulation. It completes the description of optical emission by Extensive Air Showers, a useful feature for many cosmic and gamma ray observatories. The code has been adapted to latest versions (7.6x and 7.7x) of CORSIKA and successfully tested. Physical aspects of this simulation are well under control, while currently the focus is on optimizing the computation performance. This contribution is planned to be included in one of the next official CORSIKA distributions. This activity is mainly undertaken by UCM.

A detailed assessment of computing requirements for various relevant setups of the existing functions of the CORSIKA simulation package has been conducted, mostly by INFN. The CPU load and storage needs have been profiled for different choices of the hadronic interaction models, final particle energy cuts and compression options.

The initiative to establish a common data format between CTA and KM3NeT to foster interoperability among different astroparticle observatories has produced a considerable interest in the two major Collaborations. A coordinated effort with CEVO (WP4) has started, with the aim of bringing together also the long-term experience of IVOA and the ongoing efforts by the communities of GADF and gammapy. A crucial point is the ability of the format to correctly represent the instrument response functions (IRFs) of gamma ray observatories and neutrino telescopes, which are different from optical and radio telescopes. This work is mostly furthered by CTAO, FAU and INFN.

The development and consolidation of common analysis codes have been furthered, primarily by GSI and IFAE which have been also been uploaded to the repository prototype.

1.2.3.g **Task 3.4: Foundation of Competence for Software and Service Innovation (COSSI) - AIP, CNRS-LAPP, EGO, HITS, INFN, NWO-I-CWI, OROBIX, UNITOV**

In task 3.4, machine learning approaches to simulation and experiment data have been adapted and benchmarked; definition of data formats and different deep-learning approaches have been pursued as well as the exchange of experience and harmonisation of approaches for innovative workflows. Data processing workflows between different partners (AIP, CNRS, EGO, FAU, HITS, INFN, NWO, OROBIX, UNITOV) have been gathered and the establishment of use cases for multi-messenger analysis work flows connecting several ESFRIs has started. The gathering of competence in innovative work flows was intensified and concluded in the Innovative Workflows in Astro- & Particle Physics (IWAPP) workshop in 03/2021, where the innovation working group was formed.

The establishment of use cases for multi-messenger analysis workflows connecting several ESFRIs are pursued further and will be progressed via the test science projects.

A first collaboration between the OROBIX and Virgo-EGO partners has been set up to link a deep learning pipeline developed for Gravitational Wave analysis to the data versioning tool HANGAR. A machine-learning based multi-messenger workflow to search for transients in real-time simulated data from CTA and Virgo-EGO is in preparation between CTAO and EGO, FAU is coordinating a possible KM3NeT contribution. Currently the focus is on extending and/or complement Wavefier (https://zenodo.org/record/3356656) to perform joint analysis of data from different messengers and experiments, considering the different data formats, instrument sensitivities etc. One of the main ingredients needed for the project is a set of real and/or simulated, high-level data from different messengers (gravitational waves - GWs, photons, neutrinos) collected by different detectors. Within this context, EGO collected the information on the public databases already available for gamma-ray and neutrino data, as well as on the public simulation and data analysis tools that could potentially be included in the MM-Wavefier framework. Furthermore, a pipeline to simulate populations of astrophysical transient sources and the associated GW and gamma-ray signals as observed by different GW detectors (e.g., Advanced LIGO and Advanced Virgo) and EM facilities (e.g., the *Fermi* satellite and CTA) was developed. A first realistic, synthetic catalogue of data has been generated and used in a proof of concept study (Cuoco, Patricelli et al. 2021). A request was prepared for CTA and KM3NeT collaborations to get access to information that are not public and to produce, within these collaborations, sets of simulations of multi-messenger data associated with transient astrophysical sources. Also, it is investigated if data analysis pipelines developed by these collaborations could be included in the Wavefier-MM framework.

Other transversal collaboration will be pursed with WP4, in particular with the ESO team, where machine learning is used to access data in a large archive database; with WP2 to integrate access to the data lake for machine learning pipelines; and with WP5 to implement a real time classifier for gravitational wave signals using container solutions.

AIP develops a classification engine for solar and stellar spectra, and developed data conditioning and spectral inversion techniques for a pathfinder for the future EST instruments leading to three refereed articles[4]. After this preparatory work is published as “sTools” IDL software library[5], AIP will implement the classification engine the database for high-resolution solar spectra.

CNRS-LAPP develops innovative solutions for the reconstruction of CTA data using deep learning in the framework of the GammaLearn project and has been presented during the WP3 onboarding process in Task 3.2. Its addition to the OSSR is being finalized. A first application has been successfully applied to the commissioning data of the CTA first Large-Sized Telescope. Two publications resulted from this work. A work in collaboration with OROBIX has been carried to improve the transfer learning from simulated data to observational data

In cooperation with WP4, HITS introduced a novel explorative access method for the astronomical data stored in the ESO archive, based on machine learning. A first prototype that utilizes dimensionality reduction models was developed and presented IVOA Interop Northern

[4] *Dineva et al. 2020, Astron. Nachr. 341, 64; Kuckein et al. 2021, Astron. Astrophys. 653 A165); and Verma et al. 2021, Astrophys. J. 907, 54*
[5] *Kuckein et al. 2017, IAU Symp. 327, 20*

Spring 2021, ADASS 2021 as well as Astroinformatics 2021. This prototype provides functionalities to interact with the data such as searching for similarities, importing and exporting new data, making selections and creating catalogues.

The work on machine learning techniques (pLISA), started in KM3NeT and ASTERICS, has undergone considerable evolution by INFN. It has embraced OrcaNet for its foundation and includes more regression options. Both the updates on pLISA and OrcaNeT will be onboarded onto OSSR by the end of the project.

NWO-I-CWI participated in the on-boarding of a dataset and code for Corona Mass Ejection propagation, in collaboration with UNITOV. CWI has specifically contributed an interactive dashboard for the exploration of the data (https://zenodo.org/record/5538608) which is is currently extended to include various predictive models for CME propagation which will be valuable for the wider community.

UCM has continued the development of CTLearn and presented its results in different conferences (ADASS XXXI, Spanish Astronomical Society and International Cosmic Ray Conference) and the IWAPP workshop.

UNITOV contributes - in close coordination with CWI - on the tasks “Prediction of solar wind conditions (speed and magnetic field) at L1” and “Estimated arrival time of CMEs”. The aim is to improve the predictions for the arrival times of CMEs by combining a phenomenological model with additional data sources and machine learning approaches. This effort has led to the realization of a database of CME information and a dedicated visualization dashboard. The database and the dashboard have been onboarded to OSSR (https://doi.org/10.5281/zenodo.5516979). A further result is a research paper based on such a dataset that has been submitted to the Space Weather journal.

1.2.3.h **Task 3.5: Repository Implementation and Deployment - CNRS-LAPP, FAU**

Based on partner feedback a preliminary design of the repository and the definition of technical solutions for its implementation have been developed in task 3.5 - primarily lead by LAPP - and a first prototype set up.

The OSSR, based on the Zenodo repository hosted at CERN, was developed under the *escape2020* community (https://zenodo.org/communities/escape2020). Also, the landing page (http://purl.org/escape/ossr) that serves as entry point for the OSSR has been set up. On this page, users can browse and search the OSSR per categories and keywords, as well as software currently following the onboarding process. Developers can find the OSSR guidelines, tutorials and tools to contribute to the OSSR or to its development.

This prototype has been expanded to the needs found from the other tasks. Several software packages and services have already been hosted in the development platform and the repository. The concept includes a development platform to prepare the software and services that is not obligatory as also other platforms used by contributors can be linked to the repository.

The eOSSR Python library (code: https://gitlab.in2p3.fr/escape2020/wp3/eossr, documentation: https://escape2020.pages.in2p3.fr/wp3/eossr/) gathers all the developments made for the OSSR. In particular, it includes

- an API to programmatically access the OSSR, retrieve records and publish content;
- functions to map and crosswalk metadata between the CodeMeta schema adopted for the OSSR and Zenodo internal schema;
- functions to help developers automatically contribute to the OSSR, in particular using their continuous integration (see also code snippets).

CTAO assisted in integrating the eOSSR Python package into the ESAP (WP5). Both as a simple archive search function but also as part of the interactive analysis part of the platform.

After studying the available schemas to provide software metadata, the CodeMeta schema has been adopted for the OSSR. This choice and guidelines to contribute software to the OSSR has been explained in deliverable D3.7 (License and provenance model for the software and service repository).

All developments are openly available and an example project exists, showing the full possibilities of the system, from best practices in software development (e.g. licence and meta data), over continuous integration to test and upload the project to the OSSR - cf. Fig. 4. The OSSR is linked to the EOSC core services and portal via Zenodo and openAIRE, however the EOSC profile is currently under investigation to link the OSSR directly to the portal.

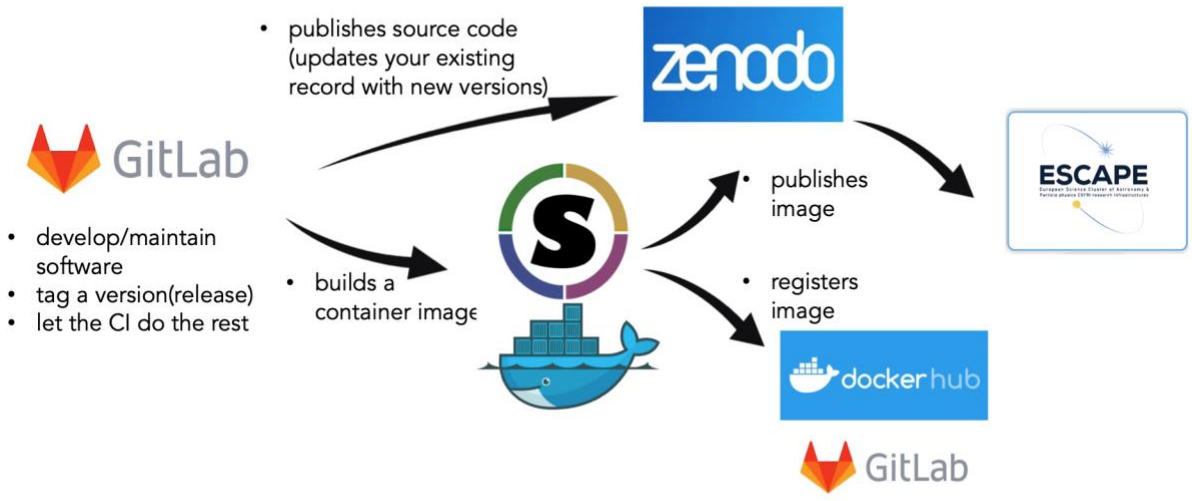

Fig. 4 OSSR Software/Service Onboarding Workflow

1.2.3.i **Joint Cross-Work Package Activities**

The cross-work package activities have naturally intensified in the second reporting period, together with the transition from a definition phase into an implementation phase.

Together with WP2, the usage of software in the datalake is pursued. The Datalake-as-a-Service container will be onboarded as outcome of WP2 to OSSR.

With WP4, the software metadata in the IVOA and OSSR are discussed and mapped. The onboarding or linking strategy of IVOA software to or with the OSSR is discussed, the IVOA will be included as key list of OSSR. Use cases with IVOA software have been defined and will be implemented in the next reporting period. The implementation and mapping of a VHE data level format (CTA, KM3NeT and other astroparticle experiment) with IVOA standards are scrutinised.

There is intense exchange with WP5 about the finding and running of software from OSSR in ESAP. As use cases Jupyter notebooks and mybinder environments have been agreed on. It is discussed how data processing/analysis workflows can be added to the OSSR, and picked up to be executed by ESAP and how to handle complex workflows. The same is done for containerized software.

1.2.3.j **Plans for future activities**

After the current reporting period, the work in WP3 will continue to be carried out along the well-laid lines of the deliverables and milestones of the project. The aim is to finalise the work prepared in the current reporting period and to establish the final OSSR with as many contributions from the partners to a FAIR open-science catalogue as possible.

For T3.1 a second ESCAPE school will be organised, as well as an all-hands meeting to scrutinize the OSSR status and cross-fertilization for open science at the end of the project. One important point to solve is the sustainment of OSSR.

T3.2 will finalise the onboarding of the partner software and services as well as enlarge the community engagement and T3.3 will continue with the data format specification as well as the ESAP backend to execute simulation software. T3.4 will continue to further the innovative approaches especially in the multi-messenger test science projects, the TSP will also be verifying the developments within OSSR and ESCAPE. T3.5 will continue the implementation tasks as well as the EOSC catalogue integration.

1.2.3.k **Review Discussions Summary**

The feedback received from the E-AB and the project review in RP1 have been implemented by listing the necessary resources for OSSR in D3.6 and intensifying the cross-WP work. The

OSSR landing page shows now the first onboarded projects and the user engagement has started with the TSPs.

In RP2 the E-AB has stressed the importance of the TSPs and encouraged to list all the software necessary to support them - this will be done with the clear aim of onboarding all the necassary software. Also the E-AB encouraged to engage with other software intiavites, and especially intensify engagment with the HSF, which is planned for this year.

1.2.4 Work package 4 CEVO (Connecting ESFRI projects to EOSC through VO framework)

WP Lead:	Mark Allen (mark.allen@ASTRO.UNISTRA.FR)
WP Participants:	**CNRS**, NWO-I, INAF, ESO, JIV-ERIC, KIS, ORB, SKAO, UEDIN, UHEI, CTAO GMBH, EGO, INTA

1.2.4.a **Introduction**

The aim of WP4 is to connect the ESFRI and other astronomy and astroparticle research infrastructures to the EOSC through the Virtual Observatory framework.

Objectives:

- Assess and implement the connection of the ESFRI and other astronomy Research Infrastructures to the EOSC through the Virtual Observatory framework.
- Refine and further pursue implementation of FAIR principles for astronomy data via the use and development of common standards for interoperability including the extension of the VO to new communities.
- Establish data stewardship practices for adding value to the scientific content of ESFRI data archives.

1.2.4.b **Organisation**

WP4 - CEVO is organized into three tasks related directly to the objectives, plus a management task. The milestones and deliverables are defined to guide the progress and record the results of the WP4 activities.

WP4 brings together partners with technical and scientific expertise in the VO (Virtual Observatory) framework, with partners who are connected to the ESFRI projects and other research infrastructures. VO expertise is provided by: CNRS-ObAS, INAF, INTA, UEDIN, UHEI and ObsParis. ObsParis provides a special link between VO and the CTA. The ESFRI and other research infrastructures are two projects from the ESFRI Roadmap: the European Solar Telescope (EST), and the cubic-kilometre-sized Neutrino Telescope (KM3NeT); and the ESFRI landmark projects: Cherenkov Telescope Array (CTA), Extremely Large Telescope (ELT) and the Square Kilometre Array (SKA). The pan-European International Organization European Southern Observatory (ESO) brings other world-class established astronomical observatories (e.g. ALMA, the La Silla Paranal observatories). Additionally, the research infrastructures European Gravitational-Wave Observatory (EGO-Virgo) and the Joint Institute for VLBI ERIC (JIVE) are also participating directly in the work package.

The project plan is structured into four tasks:

- Task 4.1: Integration of astronomy VO data and services into the EOSC (Led by INAF)
- Task 4.2: Implementation of FAIR principles for ESFRI data through the Virtual Observatory (Led by CNRS-ObAS)

- Task 4.3: Adding value to trusted content in astronomy archives (Led by CNRS-ObAS & ESO)
- Task 4.4: WP4 Management (Lead by CNRS-ObAS)

The detailed WP4 Project Plan (D4.1) provides descriptions of all of the tasks, sub-tasks, and the roles of the partners cross-referenced with the milestones and deliverables, and continues to be used for tracking progress of all of the tasks. The WP4 wiki pages[6] provide the top level organisational information about the tasks, as well as a list of all the events and meetings.

The work in Task 4.2 is organized in the Project Plan in a manner that groups together activities related to scientific domain areas: “Solar Physics”, “Radio and Millimeter Astronomy”, “High Energy Astrophysics”, UV/Optical/IR Astronomy, “Neutrino Astrophysics” and “Gravitational Wave Astrophysics”, and links these to the relevant ESFRI and RIs. The ESFRI/RIs in the different scientific domain areas are at different maturity levels concerning their use of interoperability standards and tools for making their data FAIR. This has advanced during the project, with Milestones being associated with IVOA interoperability meetings where we track the progress and priorities for the ESFRI/RIs use of the VO framework and its integration into EOSC. In RP2 we have continued the approach of organizing specific meetings to bring the relevant VO and ESFRI/RI expertise together based on the priorities and needs of the ESFRI/RIs. These meetings are all listed on the WP4 wiki pages.

The specific meetings are complemented by major annual meetings of the whole work package which are organized as Technology Forums. In RP2 the WP4 Technology Forum 2[7] was held as an on-line event 13-15 April 2021 with 59 participants including all partners (as well as cross-WP participation). This event, collected the results at the mid-point of the project, and served to plan the work for the second half of the project.

The organization of the interfaces of WP4 with other work packages was strengthened in RP2 by many joint activities, including the Technology Forum (WP3, WP5), but also in a regular monthly call with WP3, and specific meetings with WP2, WP3 and WP5. WP4 also actively participates in the ESCAPE Technical Coordination meetings, and the ESCAPE Progress meeting (28 September 2021) provided a focal point for cross-WP interaction that has led to progress in integrating WP4 results into other WPs (e.g. access to VO resources in science analysis platforms).

In RP2 there has been a strong emphasis on training events to bring the benefits of the WP4 developments to the research community and also to the community of data providers. The “Science with Interoperable Data” school (D4.3) provided hands-on experience with science tools and responded to the science cases of the participants. The Hands-on workshop for data providers (MS24) consolidated the efforts for enabling ESCAPE ESFRI and RI use and development of the VO framework and this event extended beyond the ESCAPE partners being open to the wider community of astrophysics data providers and those interested in the use of EOSC for astronomy. The WP4 training events have been organized in a way to enable the re-use of the materials, and this has been acted upon by a number of partners who have provided training in European and international events that have extended the reach of the ESCAPE project.

[6] *https://wiki.escape2020.de/index.php/WP4_-_CEVO*
[7] *https://indico.in2p3.fr/event/23481/*

The results of the WP4 are recorded in Milestone reports and Deliverable reports, and the conference proceedings and refereed journal publications are all listed on the WP4 wiki pages.

1.2.4.c **Deliverables and Milestones**

During RP2	
Deliverables	**Milestones**
D4.3: 1st Science with interoperable data school **D4.4:** Intermediate analysis report of VO data and service integration into EOSC **D4.5:** Prototype demonstrator for value-added archive services	**MS23:** Progress and priorities at IVOA (4) **MS24:** Hands on Workshop for data providers **MS25:** Progress and priorities at IVOA (5) **MS26:** Progress and priorities at IVOA (6)

1.2.4.d **Task 4.1: Integration of astronomy VO data and services into EOSC**

Task 4.1 has made important progress in this reporting period with the deliverable report D4.4 ***Intermediate analysis report of VO data and service integration into EOSC*** finalized in November 2020. This report provides an intermediate analysis report of the status of the integration, with success stories, ongoing efforts and description of the current and foreseen challenges. The progress on the activities is reported in D4.4 with respect to the insights they provide on the VO-to-EOSC integration development. The analysis builds on top of those activities and summarises the connection points, requirements and challenges that the VO community of data providers and consumers faces when trying to connect to the current EOSC infrastructure. Cross-WP activities relevant to the integration to EOSC are described, as well as connections to other EOSC related projects and participation in meetings and symposia in the EOSC landscape. The report also provides a future-look ahead to the remaining time span of the project and the remaining integration goals.

The main intermediate result of Task 4.1 is the mapping of VO Resource metadata against the EUDAT B2FIND metadata which was done as the first step of including the Virtual Observatory Registry in EOSC. This mapping of VO resources to the DataCite-based[8] EUDAT catalogue service has been validated as working, and it can be accessed through the EUDAT B2FIND service itself[9].

These results have been made visible in the ESCAPE project, at the Technology Forum 2, and also in the ESCAPE Progress meeting. Importantly we have used D4.4 as the basis for presenting the work to the wider data sharing communities in astronomy and in nearby disciplines. One example of this is the presentation at the SPIE 2020 Astronomical Telescopes and Instrumentation Conference in December 2020 by M. Molinaro (INAF), and subsequent

[8] *DataCite Metadata Schema; https://schema.datacite.org/*
[9] *http://b2find.eudat.eu/group?q=ivoa&sort=title+asc*

publication “The Virtual Observatory Ecosystem Facing the European Open Science Cloud” Molinaro et al. 2020[10].

In Task 4.1 we have contributed to specific EOSC events to represent the ESCAPE project and to interact with EOSC contact points. The events in this period have included the EOSC Symposia 2021[11], as well as RDA plenaries with co-located EOSC events. Other contributions have been made to events of the FAIRsFAIR project such as the synchronisation force workshop series (2021), and also the FAIRsFAIR Workshop on Metadata catalogues integration for Interdisciplinary Research (11 September 2020) - Presentation on ‘Metadata standards of the astronomical Virtual Observatory framework’[12]. The main partners contributing to this task in the reporting period are INAF, CNRS-ObAS, UHEI and UEDIN.

Regarding project-wise connections and cloud-based solutions, 2021 has seen the collaboration between CEVO and the H2020 NEANIAS WP4 (Space Services) members from INAF that led to the onboarding of two EOSC resources devoted to galactic astrophysics after porting them to a contributed cloud platform (GARR, Italian NREN, EOSC associated).

Connected to Task 4.1 goals, the participation of M. Molinaro in the EOSC Association Task Force on Semantic Interoperability (kick off meeting in Dec. ‘21) provides another way for ESCAPE (and CEVO) to work alongside EOSC in matching domain driven mature interoperability (from the VO) to the growing EOSC architecture and systems.

1.2.4.e **Task 4.2: Implementation of FAIR principles for ESFRI data through the Virtual Observatory**

The aim of this task is the definition and adoption of common open IVOA standards for interoperability based on ESFRI requirements. In the first reporting period of the project the initial requirements of the ESFRIs for VO standards and tools have been assessed and included in the work plan, and the priorities and progress have been reported in the deliverable D4.2 “Intermediate Analysis Report on Use of IVOA Standards for FAIR ESFRI and Community Data”. In RP2 the priorities have been evolved based on the initial progress and results, and also based on science cases and community feedback.

There has been progress on all of the activities of the task relating to the development of standards, tools for implementation, and training activities and community engagement. All of the WP4 partners have been active in this task. In the following we highlight the activities that have been carried out and the results that have been achieved in the reporting period.

Update and definition of standards based on the requirements and priorities, and representation of ESFRI interests in the global VO framework.

The WP4 work on the definition and implementation of standards is done by using and building on the IVOA framework of standards. The architecture of these standards (their organization, functions and inter-relationships) has been updated by the IVOA in 2021 with the release of

[10] SPIE paper: https://ui.adsabs.harvard.edu/abs/2020SPIE11449E..1SM/abstract
[11] On-line 16-18 June 2021 : https://www.eoscsecretariat.eu/events/eosc-symposium-2021
[12] Presentation link: https://cloud.projectescape.eu/index.php/s/oPDaMySEoMpbHPH

the IVOA Architecture Version 2.0[13]. ESCAPE has contributed to this via the participation of M. Molinaro (INAF) as the current Vice-Chair of the IVOA Technical Coordination Group (TCG) and via input to the various IVOA Working Groups. The ESCAPE ESFRI requirements for standards have been prepared via specific meetings and via the WP4 Technology Forum. This constitutes a large proportion of the work of WP4 in all of the partners as the requirements and concepts for standards must include a significant amount of technical detail, and the scientific concepts must also be well developed. These requirements are then brought to the IVOA meetings as contributed presentations, demonstrations, and input for discussion sessions at the meetings. The results of the WP4 input at the IVOA meetings are collected and assessed in Milestone reports.

The Milestone reports show that there was strong participation by WP4 partners in the IVOA meetings (15 contributions in November 2020, 29 contributions in May 2021, 14 contributions in November 2021) with representation of all of the VO-expert partners, and representation of ESFRI and RI partners: ROB for EST, ESO, SKAO, ASTRON for SKA/LOFAR, JIVE, CTAO, ObsPARIS for CTA, INFN EGO-Virgo, ECAP-FAU (WP3) for KM3NeT, plus input prepared for ALMA, and KM3NeT.

The table below lists the main results that have been achieved toward the development of interoperability standards and their implementation in tools and services related to the various ESFRI/RIs involved in WP4. The table also indicates the ESCAPE partners that have contributed to the results.

ESFRI/RIs	Partners	Results toward interoperability standards and tools

ESO-ELT.	CNRS-ObAS, ESO, HITS INAF, INTA, UEDIN, UHEI.	- IVOA standards for data access and visualisation:
		  - DataLink v1.1[14], MOC2.0,
		- Support of **VO standards in ESO archive** services - used as exemplary case to demonstrate the use of interoperability standards in a large ESFRI archive.
		- Standards relevant to Optical/IR/survey astronomy
		- Tools for visualisation
		  - Aladin Lite v3 prototype

EGO/VIRGO.	CNRS-ObAS, EGO (INFN).	- Development of **MOC2.0[15]** standard *(currently in review).* - Major update of **mocpy** as a reference implementation of the standard. - Tools & libraries integrated into GW community software *(e.g. ligo.skymap).* - Paper accepted in Astronomy & Computing journal. - Including python notebook and tutorial.
SKA, JIVE, ALMA, (LOFAR).	ASTRON, CNRS-ObAS ESO, INAF, JIVE, SKAO, UHEI.	- Creation and support of the **IVOA Radio Astronomy Interest Group**. - Publication of an IVOA note: *Radio astronomy in the VO: services implementation review v1.1*[16]. - Proposal for an **extension of the IVOA ObsCore Data Model** for radio visibility data.

[13] *https://ivoa.net/documents/IVOAArchitecture/20211101/index.html*
[14] *DataLink v1.1 (in review) https://ivoa.net/documents/DataLink/20211115/index.html*
[15] *MOC v2.0 (in final review) https://ivoa.net/documents/MOC/20211101/index.html*
[16] *IVOA Note publication : https://ivoa.net/documents/Notes/RadioVOImp/index.html*

		- **Example TAP services** developed by ESCAPE partners accessible in VO tools and in the ESCAPE platform.
CTA, KM3NeT. 	CTAO, CNRS-ObAS, CNRS-CPPM, Obs-Paris, UHEI.	- **Data Provenance standard** (ProvDM) approved by IVOA. - Many activities for adoption and implementation. (Workshop held) - Reference paper published on a: ***Management System for Provenance Information.***
EST 	CNRS-ObAS, INTA, KIS, ORB, UHEI.	- **VO metadata developed for Solar Physics**. Preparation and submission of formal proposal for additional semantic metadata to be included in IVOA UCDs. - Prototype TAP services for solar data implemented at ROB using EPN-TAP

The ESO archive is an example of a mature operational implementation of the IVOA standards. A detailed presentation by the ESO partner at the WP4 Technology Forum 2[17] showed the web and programmatic interfaces that have been implemented, citing a long list[18] of IVOA standards and tools. The presentation provided a detailed overview of the approach taken at ESO since 2017 for the use of IVOA standards. One of the identified priorities is for support of the DataLink standard which is used heavily throughout the ESO archive system. This standard has been updated in RP2 to Datalink v1.1 (with CNRS-ObAS and UHEI authors). The ESO archive also uses the Aladin Lite embeddable visualization widget (provided by CNRS-ObAS) for the sky display of the archive interface. Embeddable visualization has been identified as a common need across many of the ESCAPE partners (ALMA, EGO/Virgo, SKA, ASTRON) as well as external interests (e.g. it is used in ESASky[19]), which lead to the priority to update Aladin Lite to use WebGL technologies and significant effort has been applied to this in RP2 with the prototype[20] Aladin Lite v3 demonstrated[21] at the ADASS conference in November 2020.

Sky-spatial and temporal coverage systems for indexing of astronomy data has also been identified as a high level priority for fast data access and management of complex sky regions. In RP2 this work has been driven by requirements for gravitational wave follow-up campaigns connected to EGO/Virgo. An important result is the development of the IVOA MOC 2.0 standard which provides a "Multi-Order Coverage" map based on the HEALPix tessellation, with the major upgrade that the MOC 2.0 standard now also provides a hierarchical system for describing the coverage in time. 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 standard is led by CNRS-ObAS with contribution from EGO-Virgo (INFN) and also from external partners including the Vera C. Rubin Observatory, and NASA. The standard was submitted for IVOA review in November 2021. As required by IVOA standards, we have

[17] https://indico.in2p3.fr/event/23481/
[18] *ADQL - Aladin Lite - DataLink - HiPS - ObsCore - SAMP - SODA - SSA - STC-S (point, circle, polygon, multi-polygon) - TAP (DALI, VOSI, UWS, UCD, UTYPE, ...) - TOPCAT - VOTable - pyvo*
[19] *https://sky.esa.int*
[20] *Aladin Lite v3 prototype (best viewed in Chrome browser) : http://cdsportal.u-strasbg.fr/webgl/*
[21] *Aladin Lite v3 demonstration at ADASS : https://schedule.adass2020.es/adass2020/talk/BGV8W9/*

developed “reference implementations” of the standard to demonstrate its capabilities. The mocpy python library developed by CNRS-ObAS was released in RP2. The new capabilities were applied by EGO-Virgo and resulted in a journal paper “*Multi Order Coverage data structure to plan multi-messenger observations”*[22]*.* The paper includes an interactive python notebook and on-line video tutorial materials, both of which will be used in the upcoming WP4 school. This results also contributes to the multi-messenger aspects of the Extreme Universe Test Science Case.

The ESFRI/RIs in WP4 that are related to radio and millimetre wavelength astronomy are SKA, JIVE, and ALMA with LOFAR being closely connected to ASTRON. ESCAPE WP4 has facilitated a new level of integration of radio astronomy into the VO. A new IVOA Interest Group[23] has been created with a CNRS-ObAS member as the vice-chair, and we have led an IVOA note on “Radio astronomy in the VO: services implementation review”, published in November 2021, which includes input developed in ESCAPE in particular for ALMA, ASTRON, JIVE as well as a multiple other European radio telescopes (Nançay, NenuFAR, INAF radio telescopes). Based on requirements identified in WP4 we have led the proposal (November 2021) for an extension to the IVOA Observation Core Data Model (ObsCore) to take into account radio visibility data. The support for the use of VO standards in radio astronomy within the WP4 has led to the implementation of new and improved services for data from JIVE and ASTRON. A “VO service for the European VLBI Network” has been opened by JIVE, and was presented by M. Kettenis at the ESCAPE European Data provider Forum[24] (M24). The ASTRON ALTA (Apertif Long-Term Archive) has started publishing timing/transient data in the VO and provides a number of services based on the DACHS publishing suite (from the UHEI partner) and presented this to the IVOA community in May 2021 (MS 25). The Table Access Protocol (TAP) services from ASTRON have also been used as an example of access to VO data in the WP5 analysis platform prototype. Other work for the use of the IVOA HiPS standard has been demonstrated for LOFAR data enabling it to be visualised in the Aladin client, and initial discussion have been held with SKAO and WP2 partners to explore the use case of a HiPS node in the Data Lake storage system. The ESCAPE work on radio astronomy has also been highlighted to the wider astronomy community at the European Astronomical Society meeting in a presentation on “*Radio Astronomy in the Virtual Observatory: Bringing radio data to researchers in the spirit of Open Science*”, by CNRS-ObAS and ASTRON partners.

The Provenance Data Model is identified as a high priority for ESCAPE WP4 partners (in particular CTA and KM3NeT related partners). This subject has reached a level of maturity where the IVOA standard (supported by ESCAPE work in RP1) was approved[25] in April 2020, then in RP2: a dedicated ESCAPE CEVO workshop[26] was held on 7-8 September 2020, and ESCAPE CEVO partners led a community “Birds of a Feather - BoF” discussion session during the ADASS conference in November 2020. The workshop was focused on the use of Provenance DM information to explore the traceability of data products, find contact information and acknowledge people. Providing provenance information allows a user to assess the quality and reliability of the products, and is a step toward supporting reproducibility. The workshop gathered 28 participants including ESCAPE partners ASTRON, CNRS-ObAS,

[22] *Greco et al. 2022, accepted in Astronomy and Computing*
[23] *IVOA Radio Astronomy IG wiki pages: https://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaRadio*
[24] *https://indico.in2p3.fr/event/23987/*
[25] *https://www.ivoa.net/documents/ProvenanceDM/20200411/index.html*
[26] *https://indico.in2p3.fr/event/21913/*

CTAO-GMBH, ECAP-FAU (WP3), INAF, JIVE, Obs-Paris. This workshop contained presentations on technical solutions, demonstrations, hands-on sessions and discussions, and was used to update and collect the requirements of ESFRI projects in order to build the road map of future developments. The results and feedback of the workshop were brought to the international level at the November 2020 IVOA interoperability meeting (MS23) by M. Servillat (Obs Paris) in a presentation on "*Provenance activities in the European project ESCAPE*". Furthermore, the workshop and BoF discussion were also complemented by a specific meeting on Provenance standards with (July 9, 2021) (SKAO, Obs-Paris, CNRS-ObAS and external partner LUPM). Another external workshop was also held on "Provenance in Practice"[27] in the French community, with contributions from ESCAPE partners. The work has also been presented and published at the "International Provenance and Annotation Workshop" (19-22 July 2021[28]) in a paper "*Towards a Provenance Management System for Astronomical Observatories*"[29].

In the area of High Energy astrophysics, the WP4 work has concentrated on the need for data models for future CTA data, and this activity has also been joined by KM3NeT and also by WP3 partners. A series of meetings on this subject have been pursued by CTAO, CNRS-ObAS, Obs-Paris, ECAP-FAU, with external participation from the IVOA Data Model WG chair and the gammapy project principal investigator.

The ROB partner associated with the EST ESFRI has in RP2 focused on making the Solar Physics thematic extension to the EPN-TAP protocol for the access to tabular data. Progress has been made on the linking of the SOLARNET metadata requirements, keywords used in Solar Orbiter mission, and existing IVOA semantic metdata called 'Unified Content Descriptors' (UCD). Metadata used in solar event databases (HEK, HFC) were also linked to existing UCDs. This highlighted cases where no UCD exists for a physical quantity used in solar physics. Following this, the IVOA "Vocabulary Enhancement Proposal" (VEP) process was initiated by ROB to include two new semantic terms in the IVOA UCDs.

Progress has also been made on the implementation of IVOA standard TAP services at ROB: Three services using EPN-TAP are being implemented: One service provides the information about the USET sunspot drawings (a list of jpg files), and the second service the information about the USET sunspot group catalogue. As a first test, configuration files for these two services have been submitted to the gitlab of the ObsParis, to be reviewed by the VESPA team (Dec 2021). Another service concerns a catalogue of coronal holes data (ROB SPOCA-CH), for which the mapping between quantities computed and EPN-tap parameters and the UCD has been done, leading to a next step of investigating how to include provenance information as part of this service. The feasibility of including a TAP client within the SOLARNET VO has also been investigated and will be continued.

Other areas pursued by the WP4 partners for the use of interoperability standards includes: LineTAP - A Proposal for a Relational Model for Spectral Lines (UHEI). This concern standardisation of the way spectral line data may be queried with TAP services in "LineTAP: a

[27] *https://indico.obspm.fr/event/1267/*
[28] *https://link.springer.com/conference/ipaw*
[29] *M. Servillat et al. 2021 : https://link.springer.com/chapter/10.1007/978-3-030-80960-7_20*

proposal for an IVOA relational model for spectral lines" by M. Castro-Neves (UHEI) in collaboration with the VAMDC consortium[30], IRAP[31], and PADC[32].

Training activities to support of the scientific community for the use of FAIR data

As mentioned above, there has been a strong emphasis on training events to bring the benefits of the WP4 developments to the research community and also to the community of data providers. This has been pursued by specific scheduled events in the original work plan, and also other events where the ESCAPE WP4 training materials have be re-used.

Milestone 24 was a Hands-on workshop aimed at the data centres and projects that serve as data providers in the astrophysics community. The workshop[33] was held as the "ESCAPE European Data Provider Forum and Training Event" 23-25 November 2021 hosted on-line by the UHEI partner in coordination with CNRS-ObAS, INAF and UEDIN. The event offered an opportunity to identify common challenges and problems in the dissemination of astronomical data, in particular using Virtual Observatory standards, to exchange solutions, and to share perspectives. The event attracted 59 participants including ESCAPE partners ASTRON, CTAO, ESO, JIVE, Obs-Paris, SKAO, ROB (EST) as well as WP3, and WP5. As intended, there were many participants (~50%) external to ESCAPE including representation from ESA, IRAM, and from outside Europe: Australia, India, South Africa and USA. The initial goal was to favor in-person hands-on activities and the live implementation of code and application of tools. In the necessary on-line setting this was modified to having morning sessions of presentations, and then interactive afternoon 'hands-on' workshops and tutorials which enabled a number of one-to-one discussions on detailed topics. One of the highlights was a detailed presentation from ESO on their use of IVOA standards from high level requirements to implementation of selected standards, going through analysis of constraints, evolution of existing archive infrastructure, selection of databases, DBMSes integration and maintenance in the operational environment, using off-the-shelf components, costs (FTEs), obsolescence. The workshop topics covered tools for implementation and operations of services using the IVOA standards: TAP, DaCHS, VOLLT, MOC and HIPS. All of the materials for this training event are available on the event page.

A key training event in RP2 was the "**First Science with Interoperable Data School**" which was Deliverable D4.3 of WP4 organised by INTA. The goals of the school were twofold: on the one hand, to expose early-career European astronomers to the variety of currently available VO tools and services so that they can use them efficiently for their own research and, on the other hand, to gather their feedback on the VO tools and services and the school itself. It also served to inform the community of young researchers about the new context of data sharing where the Virtual Observatory framework is being integrated into the EOSC.

The hands-on tutorials and the more advanced presentations were aimed at addressing the needs of the participants based on the proposed science cases they provided during the registration. These hands-on tutorials provided the starting point for participants to use the

[30] *http://www.vamdc.org*
[31] *https://www.irap.omp.eu/en/homepage-en/*
[32] *https://padc.obspm.fr*
[33] *MS24 : https://indico.in2p3.fr/event/23987/*

tools and services for their own science projects with the benefit of the expertise of the tutors. To support the science cases a tutor was assigned to each participant and interactions before the school were used to set up the initial steps. The science cases were worked on in the week following the school and participants returned for a session where they presented their results. The real strength of this event is in the interactions between the young scientists and the tutors, representing the strong engagement of ESCAPE with the community to enable new kinds of science and sharing the vision of ESCAPE. A paper about the school was presented at the ADASS 2021 conference[34].

Participants were encouraged to act as VO-ambassadors in their research institutes by giving informal talks with colleagues, seminars and scientific workshops and conferences, and making use of the material employed during the School. One of the follow-on events was a re-use of the materials in a LOFAR school[35] - Hands on session VO & LOFAR use case, 26 March 2021.

1.2.4.f **Task 4.3: Adding value to the scientific content of ESFRI data archives**

The work on Task 4.3 has been very active in RP2 and has matured to the level where a number of demonstrations of Deep Learning applied to the content of the ESO archive have been presented to the astronomy community, and also scientific results were obtained which have been published[36]. The Deliverable report D4.5 on **Release of prototype machine learning enabled archive services providing value-added content to archives** was provided to the ESCAPE Project Manager on 31 January 2022.

D4.5 details the activities carried out in Task 4.3 to explore the application of techniques from the field of Machine Learning to astronomical archive search capabilities. They were carried out in a collaboration between ESCAPE partners CDS, ESO and HITS. It describes the details of the activity and the demonstrations that have been presented to the astronomy community.

The main conclusion of the report is that the exploratory work that was carried out has shown that there is considerable potential of unsupervised machine learning techniques being applied to the science archives of ESFRI and other infrastructures that serve a broad user base. ESCAPE was instrumental in showing the basic feasibility of the stated goals, thus taking a first, very significant step towards eventually reaching them. It would be of great benefit to pursue further developments based on the results obtained in ESCAPE and accelerate the application of new techniques into ESFRI archives to enable deeper levels of Open Science.

A challenge was the early departure (by 7 months) of the person who was hired at ESO to work on the task. While the possibility of an early departure was identified in the project risk register, it has had a significant impact in the delicate phase of the finalisation of some of the work, which can now proceed only minimally. ESCAPE was instrumental in showing the basic feasibility of the stated goals, thus taking a first, very significant step towards eventually reaching them.

[34] *https://ui.adsabs.harvard.edu/abs/2021arXiv211207370J/abstract*
[35] *https://www.astron.nl/lofarschool2021/*
[36] *MNRAS paper : N. Sedaghat et al. (2021), Open Access in arXiv*

1.2.4.g **Task 4.4: WP4 Management**

The WP4 management task has ensured the operation of the work package, in particular the execution of the WP4 Project Plan (D4.1) and the coordination of 25 WP4 events in RP2 (as listed on the WP4 wiki pages).

The main challenge that was faced in RP2 was the continual adaption and re-planning of the work based on the COVID-19 situation. This has resulted in almost all WP4 events in RP2 being done in virtual mode. In some cases, this has enabled higher levels of participation due to the flexibility to attend virtual events, but the lack of in-person interactions has had an impact which has reduced efficiency. Innovative solutions for WP4 meetings and events have been implemented (e.g. on-line interactive meeting spaces). The challenge of moving the school to an on-line format was quite a success. This experience opened new ways of spreading knowledge, offering the opportunity of reaching a broader community. Deliverables and milestones have been re-scheduled, and these have all been completed during RP2 as planned.

The WP4 management is supported by an engineer at CNRS-ObAS who also contributes to the training activities, and this role is expected to continue to the end of the project.

The early departure of contract staff is recognized in the risk register. This risk has been realized at ESO and CNRS-ObAS. A contractor hired at ESO for expertise in Deep Learning for Task 4.3, has departed 6 months before the expected end of the contract. In the ESO case the relevant deliverable (D4.5) will be achieved, and the impact of the departure is a disruption to further development of the prototype tools and reduced availability of the demonstrators. A contractor hired at CNRS-ObAS for technical development and support of visualization tools in Task 4.2, departed in April 2021 with the impact of delaying the release of the Aladin Lite visualization widget to RP3, which will be supported by contributed effort at CNRS-ObAS.

1.2.4.h **Joint Cross-WP Activities**

Interactions with the other work packages have been strengthened in RP2 as the different strands of work have matured and have become ready for integration. Cross-WP activities have been supported by working together in the intersecting topics, and also by including specific sessions in events for detailed discussions and planning, such as in the WP4 Technology Forum 2.

Task 4.3 is a joint activity between WP3 and WP4 where the Deep Learning expertise in the HITS partner is combined with ESO and CNRS-ObAS partners for the application of innovative Deep Learning methods to data in the ESO archive. This work has been promoted within WP3 and WP4, in particular by contributions to the WP3 “Innovative workflows in astro and particle physics” (IWAPP[37]) with a presentation of MEGAVIS - Real-time spectra analysis and visualization with autoencoders (A. D’Isanto, HITS). This WP3 event also included a session

[37] *https://indico.in2p3.fr/event/20424/*

on Virtual Observatory (H. Heinl, CNRS-ObAS). WP3 participant also made important contributions in WP4 workshops such as the Provenance workshop in September 2021.

The WP3-WP4 activities have also addressed the inclusion of IVOA related software in OSSR by making a mapping of the relevant metadata, and sharing experience on execution planning. This topic has also been brought to the international level in the IVOA (MS26).

The WP4-WP5 interactions concern the implementation of access to VO services in science analysis platforms. The UEDIN partner provides a link between WP4 and WP5 on this topic in particular on the use of VO standards and solutions in the prototype analysis platforms. Access to IVOA standardized Table Access Protocol (TAP) services has been successfully demonstrated in the prototypes, and progress has also been made for the use of the IVOA SAMP messaging protocol for interoperability between platforms and applications.

Specific meetings have been held between WP2 and WP4 on the potential use of the Data Lake as a back-end for certain types of Virtual Observatory service. This has included an assessment of whether the Rucio based storage system can be adapted for use with the IVOA HiPS system for the hierarchical access to all-sky survey data.

1.2.4.i **Plans for future activities**

The work for the next period is well defined by the detailed project plan for WP4 (D4.1). Modifications to the plan have been done in response to the continuing issues caused by COVID-19, and the progress of the work in RP1 and RP2.

Task 4.1 will build on the initial work of connecting the VO framework to EOSC as described in “Intermediate analysis report of VO data and service integration into EOSC (D4.4) which was delivered at the beginning of RP2. The maturation and development of EOSC in terms of the SRIA, the creation of the EOSC Association and the EOSC Partnership, as well as projects like EOSC Future will be taken into account in the analysis and assessments of the mapping the VO into the architecture of EOSC. This will involve more interaction with EOSC bodies, a stronger connection to other projects including H2020 EOSC Future, and also the H2020 NEANIAS project to address common aspects of the service on boarding task for a set of data resources and services. In terms of communications we plan to develop and disseminate a Vision of Virtual Observatory framework integration in EOSC. The work will form the basis of the main deliverable for Task 4.1, the “**Final analysis report of VO data and service integration into EOSC**” (D4.7) due in July 2022.

The next scientific training event of WP4 Task 4.2 is planned for 22-24 February 2022[38]. The “**2nd Science with interoperable data school**”, was initially advertised as a hybrid event, but given the new wave of the COVID-19 pandemic the school has been converted to a fully remote event. The school is a deliverable of WP4 (D4.6) and is led by CNRS-ObAS and INTA with INAF, UHEI, INFN partners contributing tutors, plus an additional tutor from the University of Bristol. The school has received some 39 detailed applications (which involved submission of proposed science cases). The school will use the well-established format as employed for

[38] *Web page for the 2nd Science with interoperable data school: https://indico.in2p3.fr/event/25225/*

the February 2021 event (D4.3) but we emphasize that the content of the tutorials reflects the maturity of the ESCAPE project, with many of the results and new capabilities developed within WP4 and ESCAPE as a whole, being integrated as training material. One example is the new tutorial on "*Multi-order coverage data structure to plan multi-messenger observations*", which focuses on scientific capabilities developed in WP4 based on EGO/VIRGO priorities, and which utilizes the interoperability standards for sky coverage maps and the data access tools (Aladin, mocpy) that have been developed in Task 4.2. The tutorial is published (as a notebook) as part of the paper and addresses some of the challenges of the Extreme Universe test science case for multi-messenger data access. As such, some of the young researchers engaged in the EOSC Future supported Science Cases will participate in the school.

Another new tutorial prepared for this school focuses on Time Domain astronomy: "Transient characterization using the Virtual Observatory". The tutorial aims at characterizing science alerts using the existing information in astronomical archives and benefiting from the advantages that the Virtual Observatory offers in terms of discovery, access and analysis of astronomical data. In particular, the tutorial makes use of TOPCAT, VizieR, SPLAT- VO, SVO Discovery Tool and VOSA. This methodology can be seen as a complement to follow-up observations.

A final Technology Forum in April/May 2022 will be the focal point that will gather the results of Tasks 4.2 and 4.3 as input for the deliverable report "**Final Analysis Report on Use of IVOA Standards for FAIR ESFRI and Community Data**" (D4.8) to be prepared for September 2022. It is also expected that the results of Task 4.2, in particular the enabled services, tools and standards will be employed in the test science cases. The inherent science cases of WP4 will also provide a way of highlighting the advances that have been made, following the model used for the EGO-Virgo, and ESO Deep Learning scientific results that have been presented in RP2 (i.e. demonstrations and publications). The results of the project are also expected to be presented at European and International meetings including the European Astronomical Society meeting (July 2022) ADASS conference (September 2022) and also at the IVOA interoperability meetings in 2022.

Interfaces of WP4 with other work packages will continue to grow in RP3, in particular as the test science cases combine the use of the different capabilities of the services and tools prototyped by the different work packages.

1.2.5 Work package 5 ESAP (ESFRI Science Analysis Platform)

WP Lead:	John Swinbank (swinbank@astron.nl)
WP Participants:	**NWO-I**, CERN, CNRS, CSIC, CTAO, EGO, FAU, IFAE, INAF, FAIR, JIV-ERIC, KIS, RUG, SKAO, UCM, UEDIN.

1.2.5.a **Introduction**

Activities in Work Package 5 are broadly divided into two major areas. First, the work package is developing ESAP, the ESFRI Science Analysis Platform: a unified mechanism by which users can discover and interact with the data products, software tools, workflows, and services that are made available through ESCAPE. Second, members of WP5 are preparing their own services, data products, and tools for integration with ESAP and their subsequent use within ESCAPE. This report describes both of these activities in parallel.

Much of this report focuses on the work which has been carried out on ESAP. A high-level understanding of its design and goals may therefore serve as a useful introduction. ESAP is designed to be the key interface between the services delivered by the ESCAPE project and the wider scientific community. It is designed to be the key interface between the services delivered by the ESCAPE project and the wider scientific community. It is designed to be extensible and flexible to adapt to both the heterogeneous needs of existing ESFRIs and other ESCAPE project partners, and to adapt to the emergent requirements of future projects.

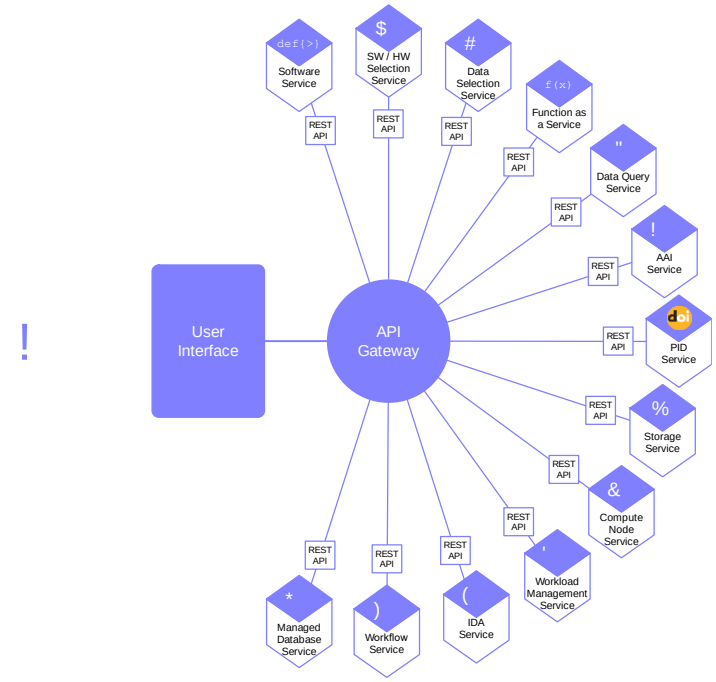

Fig. 5 The ESAP architecture is based around the API Gateway, which brokers requests between a customizable user interface and an array of RESTful services.

ESAP achieves this level of flexibility by abstracting the details of heterogeneous underlying infrastructures away from the user. This is based on a modular, plugin-driven architecture, illustrated in Fig. 5. The user connects to a service-independent web-based user interface,

which, in turn, passes requests to the API Gateway. The Gateway, in turn, uses REST[39] interfaces to communicate with a variety of external services. These REST interfaces are powered by ESAP's service connectors, which translate user requests into service-specific commands, and aggregate results for presentation. The Gateway is easily extended to address a wide variety of service types by implementing new service connectors. In this way, it can address whatever current or future capabilities are exposed through the EOSC.

The WP5 team are both working on the ESAP user interface and Gateway and developing service connectors for a variety of different service types that have been identified as important by ESFRIs and other ESCAPE project stakeholders. WP5 will also provide documentation and standardized interfaces to enable the wider community to implement their own service connectors for their particular infrastructure.

Implementing service connectors and other forms of integration for the deliverables from the other ESCAPE work packages is of paramount importance: ESAP draws together the ESCAPE service portfolio and presents it to the user as a coherent whole. This is illustrated in Fig. 6, which illustrates the various types of service which ESAP provides access to and shows how these connect to the other Work Packages.

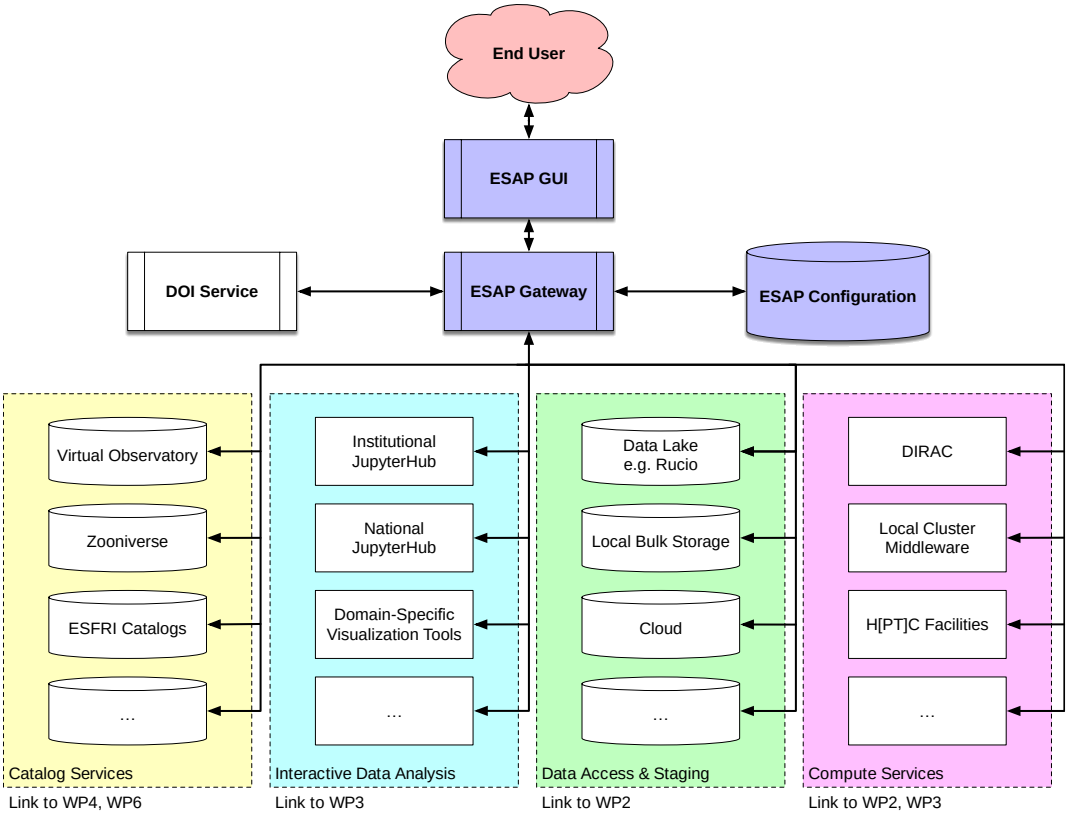

Fig. 6 ESAP in its environment, interacting with services drawn from across the ESCAPE portfolio and elsewhere. Links to other work packages are indicated.

[39] *REpresentational State Transfer*

It is important to note that — despite its name — ESAP is not a single instance of a science platform: there is no one single instance of ESAP which is supported by the ESCAPE project and its successors to provide access to all possible services. Instead, WP5 delivers ESAP as a science platform toolkit: ESFRIs, ESCAPE project partners, and other groups can use ESAP to rapidly assemble and deploy platforms which are customized to the needs of their particular user community, and which integrate their existing service portfolios. In this way, we are delivering a system which can cater to the specific and individual requirements of each ESFRI.

1.2.5.b **Organisation**

- Task 5.1. Data aggregation & staging
- Task 5.2 Software deployment & virtualization
- Task 5.3 Analysis interface, workflows, and reproducibility
- Task 5.4: Integration with HPC and HTC infrastructures
- Task 5.5: Work package management

1.2.5.c **Deliverables and Milestones**

During RP2	
Deliverables	**Milestones**
D5.3: Performance Assessment of Initial Science Platform Prototype	**MS29:** Initial science platform prototype with discovery and data staging **MS30:** Deployment of initial set of ESFRI software on prototype platform **MS31:** Second WP5 workshop to analyse prototype performance **MS32:** Integration of Science Platform with OSSR repository **MS33:** Integration of Science Platform with Data Lake expanded prototype

1.2.5.d **Task 5.1: Data aggregation & staging**

Task 5.1 aims to provide ESAP users with the capability to access and combine data from multiple collections and to stage that data for subsequent analysis.

Shopping Basket

The Shopping Basket is the central data management metaphor of the ESAP system. As the user queries archives and performs analysis operations, the results are placed in their Basket

and carried with them through the system: future analysis jobs or other tasks manipulate and augment the basket contents.

During RP2, the basket has been implemented and added to the ESAP system. This marks a major milestone in the development of ESAP as a cohesive system. Fig. 7 shows a screenshot of the current interface. This development was taken up by multiple project partners, with architectural leadership from ASTRON (NWO-I).

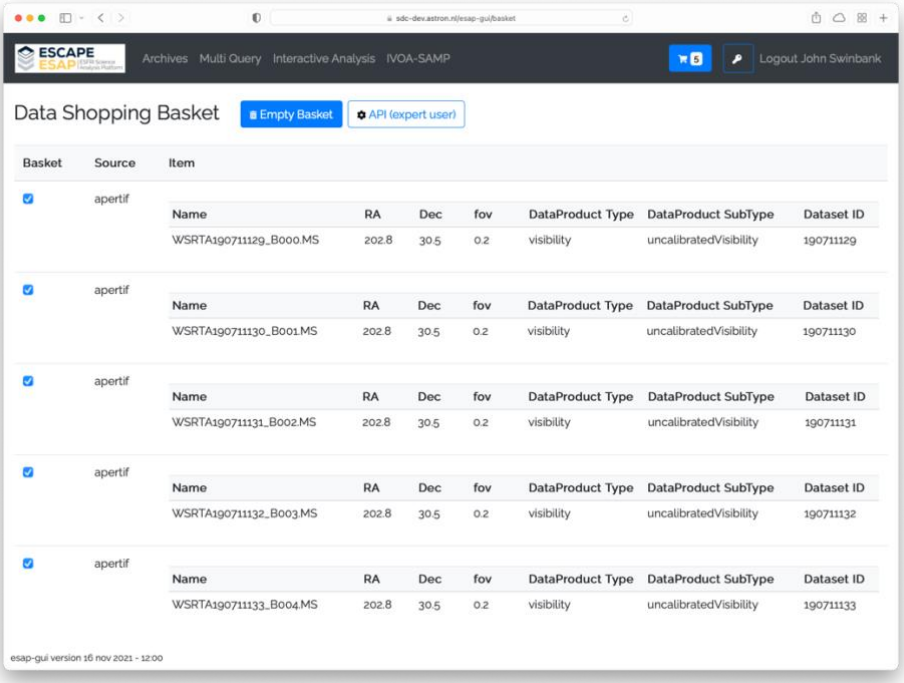

Fig. 7 The ESAP Shopping Basket in use.

Data-Lake-as-a-Service

While ESAP provides a wide range of capabilities to query archive and catalogue services, the fundamental system for bulk data storage and management developed within and used by the ESCAPE project is the Data Lake, developed by Work Package 2 (Data Infrastructure for Open Science; DIOS). It is therefore fundamental to the project as a whole that ESAP be effectively integrated with the DIOS infrastructure: users should be able to search the Data Lake from an ESAP instance, and to access and manipulate that data from within ESAP-supported analysis interfaces.

The major effort to integrate these services has been undertaken at CERN as the “Data-Lake-as-a-Service” project (DLaaS). The Rucio[40] JupyterLab software package that was developed by the CERN team during RP1 forms the basis of this effort. During RP2, this has been extended to provide a system running in the CERN OpenStack which includes:

[40] *https://rucio.cern.ch*

- Token-based OpenID Connect authentication to the ESCAPE Identity and Access Management system.
- A Data Lake data browser.
- File upload and download from the Data Lake.
- Local storage backend access to enlarge scratch notebook space.
- A latency-hiding content-delivery layer based on XCache[41].

The CERN team undertook a substantial development effort to deliver this functionality, including fundamental development and planning and engaging with a wide range of stakeholders across the project. The DLaaS was extensively tested and proved its value during the November 2021 Data and Analysis Challenge (DAC21), coordinated by ESCAPE Work Package 2.

Work is ongoing to integrate DLaaS with other core ESAP concepts. The team at ASTRON (NWO-I) has recently extended the Shopping Basket client to integrate with DLaaS, and we are considering ways to make similar capabilities available on systems outside CERN.

Network-attached storage

The team at SKAO has conducted an extensive study of enabling attached storage in analysis environments, with the aim of identifying efficient mechanisms for making large data products available to the compute resource. This has involved provisioning both NFS[42] and Ceph[43] - managed storage to the clusters on which the analysis environments are running and configuring these to present consistent data volumes across multiple user environments.

Virtual Observatory integration

The Work Package 5 team collaborates closely with their peers in WP4 to ensure integration of International Virtual Observatory Alliance (IVOA) standards throughout the ESAP data discovery and analysis systems. Within Task 5.1, this includes the following.

- *Simple Application Messaging Protocol*

The IVOA's SAMP[44] recommendation provides a convenient mechanism for compliant tools running on a single system to exchange data with each other: the user can, for example, query IVOA-compliant archives using one tool, then "broadcast" the results to another application for visualization.

ASTRON (NWO-I) and CSIC collaborated to integrate SAMP functionality into ESAP. This provides the user with the capability to broadcast data from applications running on their local system to ESAP, from where it can be aggregated and analysed with data collected from other sources. Although ESAP provides its own native VO query capability (described under VO Query Services, below), SAMP provides the user with the ability to choose their preferred tool from the diverse and feature-rich IVOA ecosystem, so they can select a tool which provides

[41] *http://slateci.io/XCache/*
[42] *Network File System*
[43] *https://ceph.io*
[44] *https://www.ivoa.net/documents/SAMP/*

specialist query services or which is most adaptable to their use case, rather than being restricted to the (necessarily) limited query interface supported by ESAP.

- *VO query services*

The ASTRON (NWO-I) and UEDIN teams have collaborated with colleagues in WP4 to refine our fully-featured service that is capable of querying IVOA-compliant services using a variety of standards, notably Table Access Protocol[45]. In particular, during RP2 work has been done to:

- Increase the number of VO services accessible through ESAP.
- Add a metadata "tree" explorer component.
- Make cosmetic enhancements and improve overall usability of the service.

- *VOSpace storage services*

The INAF team has collaborated with WP4 colleagues in the implementation of a VOSpace[46] interface for a storage service available at the INAF Italian Archives. This is not currently directly integrated with the ESAP service offering, but has been used as a testbed to develop and research potential technologies for future integration.

Other archives & data services

- *European VLBI Network archive*

The JIV-ERIC team developed a Jupyter plugin to query the EVN[47] archive by experiment code (necessary for PIs looking for their own data), by source name, or cone search using the VO ObsTAP service (developed under ESCAPE WP4) was finished. Having found data in the EVN archive, double clicking it, the plugin opens the latest known (i.e. best at the time) data reduction pipeline (data reduction scheme) in the form of a Jupyter notebook from a template with the data's file names already filled in; the calibration particulars still need to be specified by the user.

- *Zenodo*

The CTAO team integrated the eOSSR[48] Python module — developed by Work Package 3 — with ESAP. This makes it possible to search the Zenodo repository for ESCAPE community entries through the ESAP front end.

- *Cherenkov Telescope Array*

The UCM team has been making available samples of CTA data in support of integration of CTA data into the ESAP system. This work is ongoing; it is not yet exposed through ESAP.

[45] *https://www.ivoa.net/documents/TAP/*
[46] *https://www.ivoa.net/documents/VOSpace/*
[47] *European VLBI Network*
[48] *https://gitlab.in2p3.fr/escape2020/wp3/eossr*

- *Zooniverse*

The WP5 team has supported our colleagues, based primarily at the Open University and participating in ESCAPE Work Package 6 (Citizen Science), in integrating ESAP with the Zooniverse[49] citizen science platform. It is possible for ESAP users to query the Zooniverse back-end database, load the results into their ESAP Shopping Basket, and then send them to analysis services for further processing.

- *FAIR*

The team at GSI/FAIR has provided infrastructure has been provided so that experiment data from the miniCBM experiment can be injected directly to the WP2 Data Lake.

1.2.5.e **Task 5.2: Software deployment & virtualization**

Task 5.2 aims to make the codes, tools, scripts, and other packages developed in support of the various ESFRIs and other RIs readily available to ESAP users.

ESAP deployment and test systems

As discussed above, ESAP is intended to be used not as a single monolithic system, but rather for multiple instances to be customized and deployed in a range of different environments. In support of this, during RP2 the development team — notably the members at ASTRON (NWO-I), CTAO and SKAO, but with contributions from elsewhere — has substantially enhanced and documented ESAP deployment, including a fully containerized deployment process.

Although WP5 is not resourced to provide central supported instance of ESAP for general use, during RP2 we have provided access to a variety of independent ESAP installations, including:

• The team at ASTRON (NWO-I) maintains a core ESAP development instance, automatically deployed by the continuous integration system on hardware running at that institute. This runs the latest version of the codebase, a wide variety of service connectors, and is publicly available[50], though with clear disclaimers as to the level of service which may be expected from a development system. This serves as a testbed for much of the common development effort undertaken by WP5.

• The team at SKAO maintains an ESAP instance running on the STFC Cloud. As well as serving as a general resource, this has been used to test containerization and deployment of ESAP itself.

JupyterHub and BinderHub deployments

As described in Task 5.3: Analysis interface, workflows, and reproducibility, below, WP5 has focused on the development & deployment of Jupyter[51] notebooks for interactive data analysis

[49] *https://www.zooniverse.org*
[50] *https://sdc-dev.astron.nl/esap-gui*
[51] *https://jupyter.org*

to date. In support of this effort, we have deployed systems based on JupyterHub[52] (which provides a multi-user notebook server) and BinderHub[53] (a cloud service for launching Jupyter notebooks derived from code repositories). In particular:

- *CSIC*

The team at CSIC has deployed a JupyterHub service at the Spanish SKA Regional Centre prototype. This service has been used as a testbed for ESAP integration, including use of the ESCAPE Identify and Access Management service.

- *JIV-ERIC*

Building on the virtualized JupyterHub system provisioned in RP1, the team at JIV-ERIC has:

  - Upgraded their Jupyter-CASA kernel to CASA 6.4[54].
  - Deployed a non-virtualized JupyterHub system, including integration with ESCAPE Identity and Access Management.
  - Deployed a BinderHub environment, which is now integrated with the ASTRON (NWO-I) ESAP test system (that is, ESAP can be used to launch analysis jobs running on the JIV-ERIC system).

- *FAIR*
  - A prototype JupyterHub platform has been set up at GSI/FAIR.
  - Docker images are used to support a rich variety of functionality on this platform, and on the JupyterHub system at RUG, including integration with the WP2 Data Lake and with GSI/FAIR experiments' software frameworks.
  - In collaboration with RUG, they have focused on support for the R3BRoot analysis software (described under Analysis software packaging & deployment, below) and for the CMB[55] analysis environment, a key FAIR-GSI use case.

- *RUG*
  - The team at RUG have deployed a JupyterHub system, integrated with the ESCAPE Identity and Access Management service, which provides support for analysis activities at FAIR. This service is not yet integrated with the ESAP environment; that is planned for future work.
  - The IT department at RUG has made available 10 virtual machines, 20 TB storage, 50 GB RAM, and 50 CPUs from their cloud infrastructure to support this service.

[52] *https://jupyter.org/hub*
[53] *https://binderhub.readthedocs.io*
[54] *https://casa.nrao.edu*
[55] *https://www.cbm.gsi.de*

- *SKAO*

During RP1, the team at SKAO began development of a prototype JupyterHub-based system. This system was extended during RP2 by:

- Integrating it with the ESCAPE Identity and Access Management system.
- Providing load balancing across multiple worker nodes.
- Adding persistent work environments, so that users have a designated home space in which they can store their work and save their progress.
- Adding BinderHub, so that compute environments can be easily defined in code repositories.

Analysis software packaging & deployment

- The team at SKAO adapted their solution to the first SKA Science Data Challenge (SDC1; a source finding and classification exercise[56]) to run on the JupyterHub/BinderHub service (described in JupyterHub and BinderHub deployments, above). This software has been registered with the Open-source Scientific Software and services Repository (OSSR), developed by ESCAPE Work Package 3 for preservation and reusability[57].
- The teams at FAIR and RUG have packaged R3BRoot[58] and CmbRoot[59] and all its dependencies in a Docker[60] container image for distribution and deployment. This software is now available through the WP3 OSSR[61].
- The team at FAIR has prototype analysis workflows for the CBM experiment based on Jupyter notebooks and Docker containers. This notebook includes interfaces to the WP2 Data Lake.
- The teams at UCM and CTAO have been collaborating with Work Package 3 to develop CONCORDIA, a set of containers which make it possible to execute the CORSIKA air-shower simulation system[62] using the DIRAC workflow system (described in Task 5.4: Integration with HPC and HTC infrastructures, below).
- The team at CSIC provided the “HCG16 study”, which has been a key integration activity within ESCAPE over the last several months. This captures a workflow which describes the data reduction and analysis performed by Jones et al. (2021)[63] to investigate galaxy evolution in Hickson Compact Group 16. This workflow has been exposed through the ESAP platform and has been registered[64] with the Work Package 3 OSSR.

---

[56] *https://astronomers.skatelescope.org/ska-science-data-challenge-1/*
[57] *https://zenodo.org/record/5526844#.YanlUC8w1zU*
[58] *A framework used for simulations and data analysis of the R3B experiment at FAIR ; https://www.r3broot.gsi.de*
[59] *A software package for simulation, reconstruction, and analysis of the CBM experiment; https://redmine.cbm.gsi.de/projects/cbmroot*
[60] *https://www.docker.com*
[61] *https://www.zenodo.org/record/5549470#.Yat8_y8w2J8*
[62] *https://www.iap.kit.edu/corsika/*
[63] *https://ui.adsabs.harvard.edu/abs/2019A%26A...632A..78J/abstract*
[64] *https://www.zenodo.org/record/5534682#.YaoTci8w1zU*

- The team at CSIC has also produced a use case based on the analysis of time series data produced by various satellites including COROT and PLATO. Analysis notebooks for this use case are currently under development, making use of the CSIC JupyterHub service (see JupyterHub and BinderHub deployments, above).
- The team at EGO have packaged a variety of EGO/VIRGO data analysis pipelines for use with the ESCAPE Data Lake, developed by Work Package 2. These pipelines were used to participate in DAC21 during November 2021, which successfully demonstrated process data stored in the Data Lake with the WAVEFIER system[65].
- The team at EGO is now applying the same process to their Coherent Wave Burst pipeline[66].
- The team at IFAE has been developing GammaHub, a tool for interactive analysis of DL3[67] format files — the emerging standard in Gamma-ray astronomy. This has included development of data and computing models, interfacing with Apache Hive and Hadoop for data selection (see Integration with Hive and Hadoop, below), and developing interactive workflows using the Gammapy[68] package for data processing. GammaHub will be fully integrated with the ESAP system as the technology matures.
- The team at INAF as created a container which packages the LOFAR DDF Pipeline[69] in a self-contained and consistent environment, providing complete dependency management and portability. Although LOFAR is not an ESFRI, this work is explicitly targeting its role as a pathfinder to the SKA.

1.2.5.f **Task 5.3: Analysis interface, workflows, and reproducibility**

Task 5.3 integrates the data access services provided in Task 5.1 with the software from Task 5.2 to provide users with a coherent and coordinated approach to data analysis in the ESAP context.

Jupyter notebooks and OSSR integration

A major achievement during RP2 has been the integration of the core ESAP system with Jupyter notebook-based analysis environments and the WP3 OSSR. ESAP will help users identify BinderHub services which provide access to appropriate computing capabilities, and then search the OSSR for entries which can be launched onto those services. When combined with Shopping Basket interaction in the notebook (below), this provides the following analysis workflow:

1. Users search the various archives and populate their Shopping Basket with data of interest.

[65] *https://zenodo.org/record/3356656#.YaokPy8w1zU*
[66] *https://gwburst.gitlab.io*
[67] *https://gamma-astro-data-formats.readthedocs.io/en/latest/*
[68] *https://gammapy.org/*
[69] *https://github.com/mhardcastle/ddf-pipeline ; Shimwell et al (2019) ; Tasse et al (2021)*

2. Users search the OSSR and identify software which addresses their use cases.
3. Users use ESAP to identify a BinderHub service which provides them with access to an appropriate analysis platform.
4. ESAP launches the software taken from the OSSR into a notebook running within the BinderHub system.
5. From within that notebook, users can manipulate the data stored in their Shopping Basket.

Many WP5 participants have contributed to this effort, with leadership coming from CTAO, UEDIN and ASTRON (NWO-I); particularly important has been the contribution from FAU, which coordinates the WP3 effort and has acted as the key bridge to OSSR activities and integration.

Shopping Basket interaction

The Shopping Basket, described above, captures the "working set" of data under analysis by the user within the ESAP environment. It is therefore essential that the user be able to access and manipulate the Basket from within the analysis environment. The team has focused on supporting this interaction from the Jupyter notebook environment by developing a Python library which makes it possible for a notebook user to read and update their shopping basket from within the Jupyter system — or, indeed, from within any Python interpreter. Work is ongoing to refine this interface and to make integration of the Shopping Basket with the analysis environment as seamless as possible.

This work was supported by the WP5 team, but the bulk of development was contributed by the Open University to facilitate integration with the Zooniverse citizen science project as part of ESCAPE Work Package 6.

Managed Database

The team at EGO developed the concept for and began development of a Managed Database service which is integrated with ESAP. This complements the Shopping Basket, described above, by providing a fully-featured SQL database coupled to the ESAP core. Users can choose to retrieve tabular data from archives directly into their own database, rather than adding to the Shopping Basket. They can then use the full power of SQL to perform complex analytics on their data. This provides a powerful mechanism for enabling complex query workflows including tasks such as cross-matching between heterogeneous catalogues.

The Managed Database currently exists in an early prototype form, and is not yet deployed on any of the ESAP test platforms. Over the coming months, the prototype will be further developed to make it ready for production use, and the team will investigate integration with distributed SQL query engines like Trino[70] and openLookEng[71].

Reproducibility

The team at JIV-ERIC has started working on a mechanism whereby users can submit a version-controlled notebook from their private area back into the public EVN archive, with the intent that the published notebook will be accessible and, in the future, may be persistently

[70] *https://trino.io*
[71] *https://openlookeng.io*

identified and referred to in publications. In relation to this, JIV-ERIC was accepted as a member of the TUDelft DataCite consortium and gained access to DOI[72] publishing environments. Based on this, work is now underway to fully establish requirements for the EVN archive (and backend) to properly support DOIs for EVN data and / or other documents. Ultimately, this work should form the basis of the reproducibility and DOI-minting capabilities provided by ESAP.

1.2.5.g **Task 5.4: Integration with HPC and HTC infrastructures**

Task 5.4 aims to make it possible for ESAP users to deploy their workflows and analysis jobs at scale on a range of HPC and HTC infrastructures.

Integration with DIRAC

The CTAO team has taken the lead in DIRAC[73] integration, acting as focus group leaders. This serves as a test-case for the integration of workload management systems into ESAP for large-scale processing.

In addition to prototyping and design efforts, CTAO has organized several meetings to consult with external experts. From this work we have been able to secure the use of CTA-DIRAC infrastructure to begin testing our integration plans. A workflow has been defined. Work is still ongoing, with current efforts focusing on integrating user information provided by ESCAPE's Identity and Access Management service with a DIRAC system.

Integration with Hive and Hadoop

The team at IFAE has been developing interfaces for processing gamma-ray astronomical data using Apache Hive[74] and Hadoop[75]. This technology is deployed in experimental form to support the development of scientific workflows at IFAE; full integration with the ESAP system is expected to follow as those workflows mature.

Access to PANDA data

The team at FAIR has demonstrated batch-mode data processing of data from the PANDA[76] experiment stored on the WP2 Data Lake and using the GSI HPC cluster.

IVOA Execution Planner

The INAF team has collaborated with colleagues in ESCAPE WP4 to develop a technical note — IVOA Execution Planner[77] — which explores concepts for an HTTP web-service interface that provides a simple way to discover and access computing resources.

[72] *Digital Object Identifier*
[73] *http://diracgrid.org/*
[74] *https://hive.apache.org/*
[75] *https://hadoop.apache.org/*
[76] *https://panda.gsi.de*
[77] *https://github.com/ivoa/ExecutionPlannerNote*

This work aims to lay the foundations of an IVOA standard for interoperability of computing resources, in much the same way as existing IVOA recommendations target data interoperability. This work is explicitly being developed with the requirements of ESCAPE and ESAP in mind, while simultaneously building links with the wider IVOA community.

This is an exciting development, which may not reach maturity in time to be integrated with the ESAP software system over the course of this project, but which points the way to a potential future heterogeneous network of distributed science and analysis platforms.

Rosetta

The team at INAF has also developed Rosetta[78], a tool which serves both as an independent platform but is also in the process of being integrated with the ESAP core as an "execution engine".

Rosetta is built on top of a novel architecture which frames user tasks as microservices, providing full support for custom software environments and interaction methods. In this way, Rosetta can provide support for both remote desktop and command-line applications. This makes it complementary to the JupyterHub and BinderHub deployments discussed above, making a vastly expanded range of application environments available to the user.

Rosetta is currently deployed and used at INAF, as shown in Fig. 8. Integration with the ESAP core is in its planning stages at the time of writing.

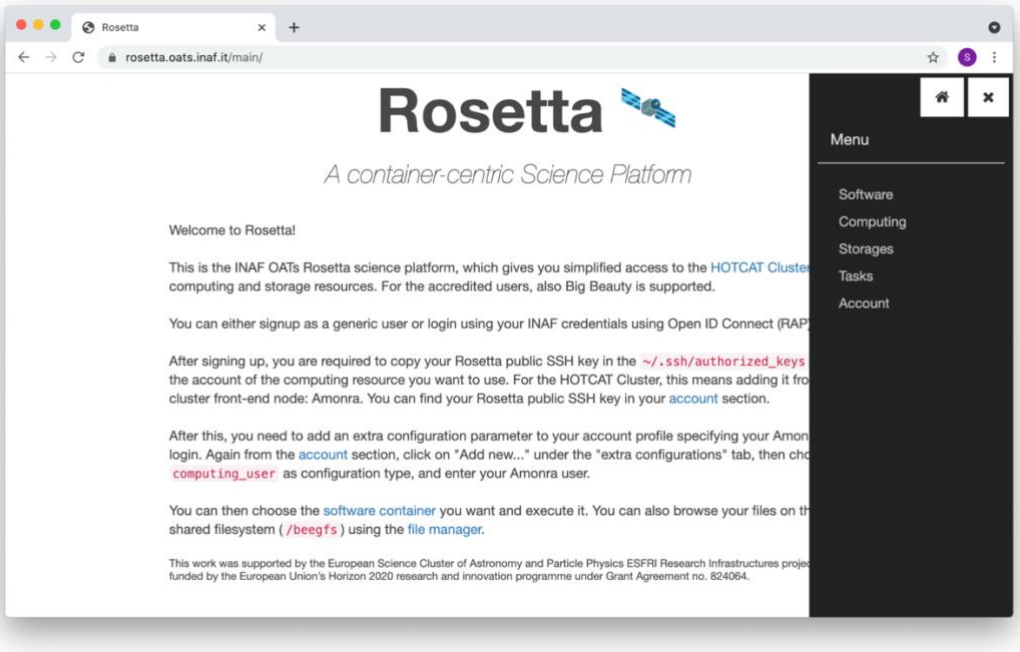

Fig. 8 The Rosetta platform deployed at INAF.

[78] *Source at https://www.ict.inaf.it/gitlab/exact/Rosetta ; draft publication at https://sarusso.github.io/Rosetta_draft.pdf*

1.2.5.h **Task 5.5: Work package management**

Task 5.5 is responsible for overall coordination of the WP5 effort. This includes not only managing the work package itself, but also coordinating activities with the other ESCAPE work packages, and managing relationships with the wider ecosystem of related and relevant projects.

Overall WP5 management

Overall management of WP5 is undertaken by ASTRON (NWO-I). During RP2, these activities included:

- Replacement of Michiel van Haarlem by John Swinbank as WP5 Coordinator.
- Facilitating communication across the work package by organizing regular WP5 meetings, technical discussions and presentations, and focus group meetings.
- Coordinating of cross-work package activities, in collaboration with the leaders of other work packages.
- Development and communication of the ESAP vision.
- Providing architectural guidance and review on code contributions to ESAP.
- Participation in ESCAPE Executive Board meetings and the ESCAPE General Assembly of September 2021.
- Soliciting, organizing, and addressing use cases from ESFRIs and project partners.
- Organized the Second WP5 Workshop in August 2021[79]. At this meeting, the status of ESAP development was presented and ESFRIs, project partners, and stakeholders were invited to provide feedback and further refine the direction of development.

Relationships with other work packages and ESFRI stakeholders

In addition, several other members of WP5 have made important contributions to communications with other work packages and stakeholders. In particular:

- FAU has created and curated a platform[80] for collecting use cases, and has facilitated the identification of common use cases between WP5 and project partners.
- CTAO has taken a lead in pushing for a refinement and formalization of the use cases.
- CERN has provided a key link in coordinating WP5 activities with WP2.
- FAU has provided a key link in coordinating WP5 activities with WP3.

[79] *http://indico.in2p3.fr/event/SecondWP5Workshop*
[80] *https://project.escape2020.de*

- UEDIN has provided a key link in coordinating WP5 activities with WP4.

Outreach and communication

The ASTRON (NWO-I team has presented WP5 activities and provided introductions to ESAP and ESCAPE at a variety of international workshops, conferences, and meetings. Highlights include:

- Invited presentation to PUNCH4NFDI[81] (April 2021)
- Discussion at the European Astronomical Society Annual Meeting 2021[82] (June 2021)
- Poster presentation at ADASS XXXI[83] (October 2021)
- Invited presentation to the European Data Provider Forum[84] (November 2021)
- Invited presentation at the Low-latency alerts & Data analysis for Multi-messenger Astrophysics meeting[85] (January 2022).

CSIC has also carried out various outreach activities for disseminating ESCAPE results to the Spanish community. Notably, these include:

- Preparation of a video for the European Researchers' Night 2019[86].
- A talk, entitled ESCAPE. Partículas y astros en un mar de datos, at the European Researchers' Night 2021[87].
- An interview with members of ESCAPE was published in CSIC Abierto[88].

Related efforts

While WP5 effort is centred around the development of ESAP itself, it is essential to make a careful assessment of similar efforts being carried out within other projects. This enables us to establish the state of the art in the domain of scientific interactive analysis and may allow WP5 to identify components or software methodologies as possible alternatives to the current ESAP implementation.

The team at CNRS has led this effort with a detailed analysis of the Rubin Science Platform (RSP), the central data access and interactive analysis tool designed and developed by the Vera C. Rubin Observatory[89]. It will provide access to 15 PB of catalogue data corresponding to 37 billion objects and to the corresponding set of 5.5 million on-sky images collected during

[81] *https://www.punch4nfdi.de*
[82] *https://eas.unige.ch/EAS2021/*
[83] *https://www.adass2021.ac.za*
[84] *https://indico.in2p3.fr/event/23987/*
[85] *https://indico.in2p3.fr/event/25290/*
[86] *https://projectescape.eu/events/european-researchers'-night-2020*
[87] *https://lanochedelosinvestigadores.fundaciondescubre.es/actividades/proyecto-escape/*
[88] *http://digital.csic.es/handle/10261/254492*
[89] *https://www.lsst.org*

10 years of operation. It is designed to scale up to serve the needs of 7000 users. The RSP is released under an open-source license, with all code available through GitHub[90].

Apart from the data themselves, nothing within the RSP is specific to Rubin. All the functionality — including Jupyter notebooks, access to tabular data, data visualization, etc; see Fig. 9 — are standard components of modern analysis environments. Data access is based on IVOA standards.

CNRS, in close contact with the RSP developers, has deployed two instances of the platform. This activity required substantial effort to acquire the necessary knowledge of this complicated system. These instances are now running stably, and have been presented in a WP5 meeting. Even this minimal deployment is enough to demonstrate the RSP's capability to easily scale-up by simply adding Kubernetes workers.

The CNRS team is now investigating how to implement a Rubin independent use case, possibly using CTA simulated data, to demonstrate how the RSP could fulfil other experiments' needs.

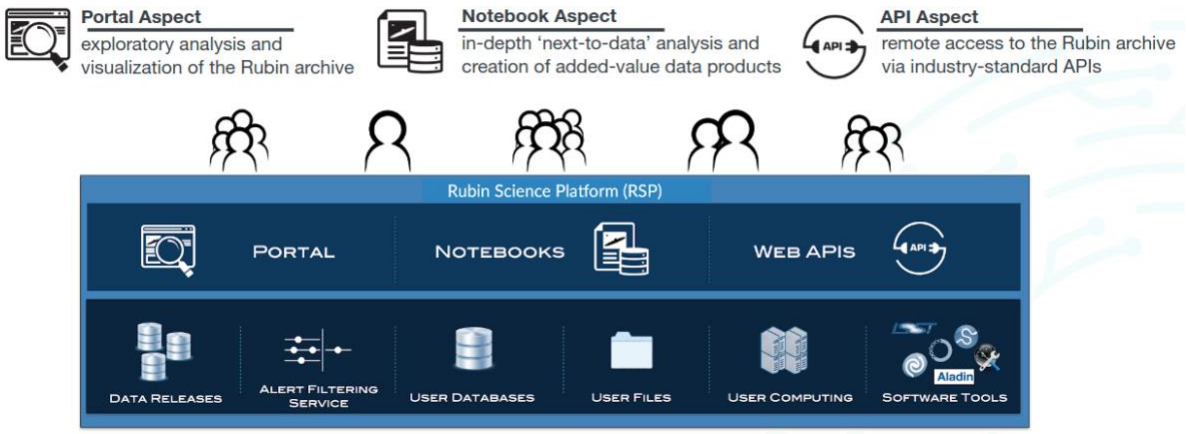

Fig. 9 Core components of the Rubin Science Platform.

1.2.5.i **Plans for future activities**

Future activities planned for WP5 focus on consolidation and integration of the extensive programme of work described above: the ultimate goal being to combine all of these threads into a single compelling service package. Specifically, this will include:

- Completion of an ongoing upgrade to the ESAP core architecture, providing additional stability and scalability, and providing new asynchronous service integration capabilities.
- Improvements to the ESAP codebase to make it easier to deploy and maintain in operational scenarios.

[90] https://github.com/lsst-sqre

- Integrate Rosetta and/or other execution engines with the ESAP core to enable a wider range of compute and analysis task types, and collaborate with WP4 and the wider VO community to develop concepts for interoperability of compute resources.
- Deeper integration with the WP2 DIOS, with a particular focus on providing easy-to-deploy solutions for providing Data Lake integration at a variety of deployment locations.
- Deeper integration with the WP3 OSSR, including more advanced searching and selecting of software and workflows within the repository.
- Improved support for batch compute services.
- Support for intelligent scheduling of compute and analysis tasks based on data type and locality and compute system availability and capabilities.
- Complete integration of the various data and analysis services which have been developed by project partners during RP2.
- Integration of ESAP and its ecosystem with EOSC

1.2.6 Work package 6 ECO (Engagement and COmmunication)

WP Lead:	Stephen Serjeant (stephen.serjeant@open.ac.uk)
WP Participants:	**OU**, Trust-IT, CNRS..

1.2.6.a **Introduction**

Activities in Work Package 6 are divided into two tasks. Task 6.1 is a conventional communications activity, with information being disseminated to the scientific communities and to the science-inclined public at large. Task 6.2 is a less traditional two-way engagement, dialogue and debate with the science-inclined public, through mass participation experiments (i.e. citizen science). This is illustrated below in Fig. 10.

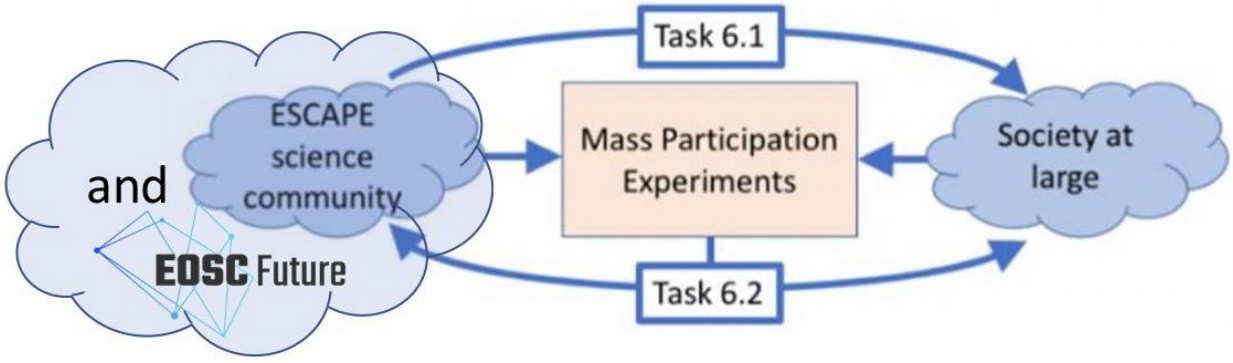

Fig. 10 Schematic representation of the WP6 ECO activities.

A recent extension of our tasks has been to incorporate engagement not just with the ESCAPE science community, but with the wider EOSC-Future communities. This is discussed in more detail below.

The WP6 PDRA at the OU, Dr Hugh Dickinson, was promoted to a permanent position at the OU starting 1 Sept 2021. Hugh will remain involved in ESCAPE, albeit at no staffing cost to the project. We have recruited a new WP6 PDRA, Dr James Pearson, who started at the OU on 25 October 2021.

1.2.6.b **Organisation**

The organisational structure of work package 2 is following the two tasks below:

- Task 6.1: Communication activities
- Task 6.2: Mass participation experiments

1.2.6.c **Deliverables and Milestones**

During RP2	
Deliverables	**Milestones**
D6.4: Citizen Science Experiments	**MS36**: First Citizen Science workshop
D6.5: Promotional education animation videos	**MS37:** Second ESCAPE Citizen Science Workshop

1.2.6.d **Task 6.1: Communication activities**

Up to Month 31 (August 2021), WP6 has had the following communication impacts:

- 1300 social media members
- 140 newsletter subscribers
- 259K impressions on Twitter
- 72.4K page views on website
- 31.3K sessions on the website
- 19.2K users on the website
- 50 news articles
- 14 videos.

The ESCAPE website has been redeveloped into a responsive and dynamic site, with new sections on ESCAPE in Open Science, ESCAPE & EOSC-Future, Dark Matter, Extreme Universe and Gravitational Waves, Publications, Jobs, and ESCAPE Sixty Second Adventures - Artificial Intelligence" - Uses of AI / Machine Learning pages. There are new updates on "ESCAPE EAB", "About us" and "Organisation" pages. We have also created a "quote section in the homepage" for endorsements about the project from senior community figures.

WP6 has coordinated communication activities with other research infrastructure clusters (PANOSC, SSHOC, EOSC-Life, ENVRI-FAIR, FAIRsFAIR), particularly through our recent involvement in the EOSC-Future project. Joint face-to-face communications meeting in abeyance since the pandemic, but resulting virtual presentations at the EOSC Symposium 2021, the Open Science Fair 2021 (in collaboration with REINFORCE H2020 project) and other meetings. We have also run an Open Science Fair debate on how to increase the subject knowledge of non-subject-specialists in EOSC to be published (in preparation).

Finally, we have created new Sixty Second Adventures animations, to illustrate and communicate ESCAPE's impact, particularly in the domain of Artificial Intelligence. These videos will comprise part of our deliverable in Month 36, but are already available ahead of schedule on our website.

1.2.6.e **Task 6.2: Mass participation experiments**

Scientific research is entering the Big Data era. Forthcoming large infrastructure experiments like SKA, High Luminosity LHC, and the Vera Rubin Observatory promising petabyte-to-exabyte-scale datasets. Citizen science is a well-established technique that is increasingly prevalent in many scientific communities and enables analysis of large and complex datasets in ways that are difficult to automate. The Zooniverse[91] is the world's largest and most popular online platform for citizen science research. It currently hosts over 50 live citizen science projects and engages over 1 million registered contributors around the world. We have therefore focussed the bulk of our crowdsourced data analysis effort around the Zooniverse.

We are currently ahead of schedule for Deliverable 6.6, "Citizen Science Experiments with Embedded Educational Resources", due in Month 42. We have already launched the SuperWasp Black Hole Hunters project[92], which has been extremely successful and has generated over 630 thousand classifications to date. The science goal of the project is to identify examples of self-lensing in black hole binary systems, through rare peaks in the light curves of the SuperWasp variable star monitoring project (a precursor to the Rubin LSST project, one of the explicit ESCAPE facilities). We were also very fortunate to secure the promotion of this project through inclusion on the print item for the BBC television series Universe, broadcast in the UK this autumn. A second citizen science project, provisionally titled "Galaxy Zoo Hyper Suprime Cam", is under development. This project aims to ask volunteers to mine images from the eight-metre Subaru telescope in Hawaii, in its imaging survey of the premier deep field for the forthcoming Euclid space telescope. As with the SuperWasp project, this is intended as preparatory work for the Rubin LSST project.

1.2.6.f **Joint Cross-Work Package Activities**

WP6 has been working closely with Work Package 5 on the ESCAPE Science Analysis Platform (ESAP), specifically in integrating ESAP with the Zooniverse citizen science platform. ESAP users are currently able to query the Zooniverse back-end database, load the results into their ESAP Shopping Basket, and then send them to analysis services for further processing. Further details of ESAP progress is given in the WP5 Technical Updates for this reporting period. Note that this activity is not included in the Description of Work, but was carried out with a view to facilitating a better integration of the ESCAPE services including citizen science.

As part of our involvement in EOSC-Future, we have extended the remit of our milestone Second ESCAPE Citizen Science Workshop to research domains beyond ESCAPE. As a direct result, we now have crowdsourced data mining projects at an advanced stage of development in the domains of the SSHOC and ENVRI-FAIR research infrastructure clusters.

[91] *https://www.zooniverse.org*
[92] *https://www.zooniverse.org/projects/hughdickinson/superwasp-black-hole-hunters*

1.2.6.g **Events, outreach and communication**

ESCAPE has participated in many events and dissemination activities in this reporting period, listed below in summary form. Joint face-to-face communications meeting have been largely in abeyance since the pandemic, but ECO have made virtual presentations at EOSC Symposium 2021, Open Science Fair 2021 (in collaboration with REINFORCE H2020 project) and other meetings. At the 2021 Open Science Fair, a debate was co-organised by ECO on how to increase the subject knowledge of non-subject-specialists in EOSC, to be published (in preparation).

- 2020-09-29 PRACE-CERN-GEANT-SKAO kick-off workshop on High Performance Computing https://indico.cern.ch/event/952623/timetable/
- 2020-09-30 Ask The Experts session https://www.eventbrite.co.uk/e/ask-the-expert-session-things-that-go-bump-in-the-night-sky-tickets-121216219977
- 2020-10-06 2nd Workshop on the connection of ESFRI Research Infrastructures (RIs) to the European Open Science Cloud (EOSC) https://www.esfri.eu/esfri-events/2nd-esfri-ris-eosc-workshop-research-infrastructures-shaping-eosc
- 2020-10-07 KM3NeT Collaboration Meeting .
- 2020-10-13 PHIDIAS: Steps forward in detection and identification of anomalous atmospheric events https://www.google.com/search?client=firefox-b-d&q=PHIDIAS%3A+Steps+forward+in+detection+and+identification+of+anomalous+atmospheric+events
- 2020-10-26 Second WP5 workshop to analyze prototype performance & F2F meeting https://indico.in2p3.fr/event/22482/overview
- 2020-11-16 Realising the European Open Science Cloud https://www.sshopencloud.eu/realising-european-open-science-cloud
- 2020-11-20 107th Plenary ECFA meeting https://indico.cern.ch/event/966397/
- 2020-12-01 AGU 2020 Fall meeting https://www.agu.org/fall-meeting
- 2020-12-02 ESCAPE 1st Citizen Science Workshop https://indico.in2p3.fr/event/21939/registrations/
- 2020-12-09 2nd ESCAPE WP2/DIOS workshop https://indico.in2p3.fr/event/22693/
- 2021-01-12 PaN ESCAPE Data Management Workshop https://projectescape.eu/events/pan-escape-data-management-workshop
- 2021-02-04 EOSC France European Days. https://eoscfrance.sciencesconf.org/342848
- 2021-02-08 WP4 "1st school on science with interoperable data" (Mark Allen) https://svo.cab.inta-csic.es/svoMeetings/index.php?mid=54&action=page&pagename=Meetings/SVO_thematic_network/First_ESCAPE_School/Presentation
- 2021-02-15 2nd WP5 workshop to analyse prototype performance .
- 2021-03-05 NuPECC Meeting Liverpool-on-the-web https://indico.ph.tum.de/event/6784/
- 2021-03-08 IWAPP Workshop - Innovative Workflows in Astro and Particle Physics https://indico.in2p3.fr/event/20424/registrations/2376/
- 2021-04-13 "On the Future of Data Centers and eScience Institutes

- Celebrating LIneA's 10th Anniversary" https://workshop2021.linea.gov.br/
- 2021-04-13 WP4 Technology Forum 2 https://indico.in2p3.fr/event/23481/
- 2021-04-19 The ESFRI Clusters at RDA House of Commons https://projectescape.eu/events/esfri-clusters-rda-house-commons
- 2021-05-17 CHEP2021: 25th International Conference on Computing in High-Energy and Nuclear Physics https://indico.cern.ch/event/948465/contributions/4348789/
- 2021-05-24 The IVOA Interoperability Meeting 2021 https://indico.ict.inaf.it/event/1441/
- 2021-06-07 ESCAPE Summer School https://indico.in2p3.fr/event/20306/ and https://projectescape.eu/events/escape-summer-school-2021
- 2021-06-11 ESFRI Science Clusters' Long Term Commitments to Open Science https://projectescape.eu/events/esfri-science-clusters-long-term-commitments-open-science
- 2021-06-16 EOSC Symposium 2021 https://www.eoscsecretariat.eu/eosc-symposium-2021
- 2021-06-28 European Astronomical Society Annual Meeting https://eas.unige.ch/EAS2021/
- 2021-09-15 ESCAPE and EOSC Future 2nd Citizen Science Workshop https://indico.in2p3.fr/event/24676/registrations/
- 2021-09-22 Open Science Fair 2021 https://www.opensciencefair.eu/2021/workshops/citizen-science-open-science-challenges-and-opportunities-for-collaboration
- 2021-10-04 ADASS XXXI https://www.adass2021.ac.za/
- 2022-02-20 Second ESCAPE Virtual Observatory School https://projectescape.eu/events/second-escape-virtual-observatory-school

1.2.6.h **Plans for future activities**

Future activities planned for WP6 partly focus on extending the work described above, following the Description of Work, for example:

- Promotional and educational videos (deliverable D6.5, M36) and citizen science experiments with embedded educational resources (deliverable D6.6, M42)
- Webinars covering each of the ESCAPE work packages / services
- News articles about the ESCAPE (Test) Science Projects
- At least two further newsletters
- Interview about the ESCAPE summer school

However, **our broader objective is to better integrate the crowdsourced data analysis (citizen science) into the wider EOSC ecosystem,** exactly in line with the recommendations of the External Advisory Board, and this will be a significant focus of our future activities.

The current ESAP data discovery portal provides functionality to identify Zooniverse classification data and mark it for subsequent download and processing by interactive or batch analyses launched by the ESAP. However, ESAP does not currently provide tools that enable

straightforward generation and management of citizen science projects on the Zooniverse platform. Many of the experimental data generated by the ESCAPE partners and accessed via the ESCAPE data lake are complex, large in volume, and present significant analysis challenges that, despite being difficult to automate, could form the basis of exciting and engaging citizen science projects. Exposing data via the Zooniverse platform would not only enable valuable scientific analyses that are difficult to automate, but would also promote ESCAPE and ESCAPE science to the general public.

To minimise any perceived barriers to entry for researchers wishing to launch citizen science projects, we will develop a suite of easy-to-use tools that can be used to automatically create, initialise and manage new projects on the Zooniverse platform. Our tools will streamline interaction with the Zooniverse back-end (Panoptes) and advanced aggregation and retirement engine (Caesar) via their respective REST APIs. We will also provide generic template workflows that can be straightforwardly adapted to integrate Zooniverse projects with automated analysis pipelines and Deep Learning models. In particular, we will provide a template active learning workflow that can be used to optimize the generation of Zooniverse volunteer labelled-training data to rapidly train or retrain Deep Learning algorithms.

Conversion of experimental data into high-level “subjects” (images, videos etc.) that can be successfully interpreted and analysed by non-expert volunteers often requires substantial effort from researchers who manage citizen science projects. To streamline this process and reduce the workload of running a successful citizen science project, we will also provide tools and template workflows that generate attractive subject data, upload them to the Zooniverse servers and manage them effectively as the volunteer classifications accumulate.

Although we will focus on the data types and analyses typical of the ESCAPE partner experiments, we anticipate that the diversity of use cases we encounter and facilitate will be sufficient to make the tools and workflows we develop readily applicable to numerous other experimental datasets that will be generated as part of the EOSC Future initiative.

We intend to create:

1. Notebook and documentary materials demonstrating web-interface based and programmatic (scriptable) Zooniverse project management including project and workflow creation, subject creation and upload, adding training and feedback to subjects,
2. Notebook and documentary materials demonstrating integration with the Zooniverse’s Caesar engine for advanced aggregation and efficient subject retirement.
3. Notebook demonstrating how to integrate Zooniverse projects with existing machine learning frameworks and combine volunteer classifications with machine learning predictions.
4. Notebook and documentary materials demonstrating how to set up an active learning framework to continuously train machine learning models using volunteer classifications of optimally selected subjects.

1.2.7 Ethics Requirements

Work package 7 lead by CNRS sets out the 'ethics requirements' that the project must comply with. Deliverable D7.1 POPD Requirement 1 was submitted during previous reporting period (RP1). No update has been made during the this reporting period (RP2).

1.3 Impact

The information on expected impacts as described in the DoA is still relevant and up to date.

1.4 Access provisions to Research Infrastructures

1.4.1 Trans-national Access Activities (TA)

1.4.2 Virtual Access Activities (VA)

1.5 Resources used to provide access to Research Infrastructures

Not Applicable.

2 Update of the plan for exploitation and dissemination of result (if applicable)

The plan for exploitation and dissemination of result do not require an update.

3 Update of the data management plan (if applicable)

The data management plan does not require an update.

4 Follow-up of recommendations and comments from previous review(s) (if applicable)

The first project review took place in November 2020. Here are elements from the Review Report.

(A)

- *Project achieved most of its objectives and milestones for the period with relatively minor deviations. ESCAPE delivered exceptional results with significant immediate or potential impact (even if not all objectives mentioned in the Annex 1 to the GA were achieved).*
- *The project so far has contributed to the idea de EOSC but its impact is limited to the participant ESFRIs and facilities (RIs), not to the astrophysics and particle physics researchers in general.*

Follow up:

- The participation of RIs is indispensable and fundamental to fulfil the expected genuine scope of the project as it was considered in the H2020 call to which ESCAPE proposal was successfully submitted. However, ESCAPE together with the ESFRI-EOSC Task Force have been acting to enhance the researchers' involvement in EOSC. Now, the TSPs with their association with two corresponding JENAA actions (1 and 2) aim at involving into the development of the ESCAPE VRE the entire concerned community, astrophysics and particle physics researchers in general, that will be asked to participate in shaping further the open science EOSC cell that our project is delivering.
- The extremely successful training and citizen-science events that have taken place online despite COVID-19 have allowed to strengthen the support to users less equipped/experienced as well as fellow citizens.

(B)

- *Demonstrate better the need of creating a multiple of similar services. For example, what is the added value of accessing data from different services and data bases. During the mid-term review we have seen that one can access different set of astronomy data from multiple sources - telescopes archive, data lake, ESFRI science analyses platform.*

Follow up:

- The ESCAPE VRE implementation demonstrates the declination of FAIR Science at the different level of openness and expertise, while it adds up significant degree of freedom in interoperability of data, rewarding of scientists' contributions and education opportunities. Although every single ESFRI project in ESCAPE remains responsible for curating and publishing their respective data products. The detailed mechanism put in place would vary from project to project, depending on the types of data being generated and the needs of their community.

(C)

- *Increase ESO/European Extreme Large Telescope engagement.*

Follow up:

- ESO's participation in ESCAPE is focused mainly on the scope of WP4, with strong links and interest on several other work packages (Data Lake, Science Platforms). This is both for its current observatories, namely La Silla Paranal and ALMA, and, crucially, the Extremely Large Telescope (ELT). During the first phase of ESCAPE, ESO was very engaged in delivering according to the agreed milestones. Specifically, EOS is now wrapping up its main deliverable, i.e. D4.5 of WP4 as well as provided extensive requirements to WP5 both for ALMA and the VLT and, in perspective, the ELT. Interest in WP2/Data Lake is also confirmed as high. ESO focus there is not so much of I/O of large amounts of data, which is not an issue for the ELT to the level that it is for CERN or SKA, but to integrate VO query capabilities in Rucio/Data Lake. There is focus on the topic now after the discussions at the last E-GA meeting back in September 2021, where it was highlighted as potential issue. Finally, I suppose ESO representative joining the E-MST as E-GA Chair as well as ESO directorate support to the new long-term perspective of ESCAPE can also be regarded as increased engagement from ESO in our cluster as well as in EOSC.

5 Deviations from Annex 1 and Annex 2 (if applicable)

5.1 Tasks

Following Deliverables and Milestones have been delayed due to COVID-19. These delays have been communicated and approved by the EC project office and also recorded in an Amendment (AMD-824064-27 and AMD-824064-32).

Deliverables:

	Deliverable Name	Responsible	Initial due date	Last Due date
D1.1	Final Integration Event	WP1 CNRS	M40 May 2022	M46 November 2022
D1.2	Final Blueprint report	WP1 CNRS	M42 July 2022	M48 January 2023
D3.4	Establishing of innovation competence group	WP3 FAU	M18 July 2020	M26 March 2021
D3.5	Thematic training event - first school for software development and deployment in the EOSC	WP3 FAU	M24 January 2021	M29 June 2021
D3.8	Thematic training event - second school for software development and deployment in the EOSC	WP3 FAU	M36 January 2022	M44 September 2022
D4.3	First Science with interoperable data school	WP4 CNRS	M16 May 2020	M27 April 2021
D4.4	Intermediate analysis report of VO data and service integration into EOSC	WP4 CNRS	M18 July 2020	M22 November 2020
D4.5	Prototype demonstrator for value-added archive services	WP4 CNRS	M18 July 2020	M36 January 2022
D4.6	Second Science with Interoparable Data school	WP4 CNRS	M35 December 2021	M41 June 2022
D4.7	Final Analysis Report on integration of VO data and services into EOSC	WP4 CNRS	M38 March 2022	M42 July 2022
D4.8	Final analysis report on use of IVOA standards for FAIR ESFRI and community data, WP4 and best stewardship practices for value- added data	WP4 CNRS	M40 May 2022	M44 September 2022
D5.3	Performance assessment of initial Science Platform prototype	WP5 NWO	M24 January 2021	M43 August 2022

Milestones:

	Milestone Name	Responsible	Initial due date	Last Due date
MS6	E-EAB evaluation + Acceptance of periodic report 3 + Final project review (incl. Lessons learned) Means of verification - Review of deliverables of M21 - M42, final project review report	WP1 CNRS	M42 July 2022	M48 January 2023
MS11	Extension of the data lake to efficiently serve data to external compute resources providers	WP2 CERN	M30 July 2021	M35 December 2021
MS12	ISO 16363 certification process underway in core data centres	WP2 CERN	M32 September 2021	M41 June 2022

MS15	Establishment of innovation competence group	WP3 FAU	M21 October 2020	M26 March 2021
MS17	Progress of common software and service proposition	WP3 FAU	M27 April 2021	M32 September 2021
MS18	Final workshop to evaluate the outcome of WP3 with respects to the main objectives of the call and define the necessary future steps	WP3 FAU	M40 May 2022	M46 November 2022
MS19	Software and Service Repository online	WP3 FAU	M42 July 2022	M48 January 2023
MS24	Hands-on workshop for data providers	WP4 CNRS	M29 June 2021	M35 December 2021
MS29	Initial science platform prototype	WP5 NWO	M18 July 2020	M22 November 2020
MS30	Deployment of initial set of ESFRI software on prototype platform	WP5 NWO	M20 September 2020	M22 November 2020
MS31	Second WP5 workshop to analyse prototype performance	WP5 NWO	M26 March 2021	M31 August 2021
MS32	Integration of Science Platform with OSSR repository	WP5 NWO	M28 May 2021	M32 September 2021
MS33	Integration of Science Platform with Data Lake expanded prototype	WP5 NWO	M30 July 2021	M34 November 2021
MS34	Delivery and integration of new ESFRI visualization and analysis tools	WP5 NWO	M36 January 2022	M37 February 2022
MS35	Final WP5 ESFRI user training workshop on the Science Platform	WP5 NWO	M38 March 2022	M40 May 2022
MS36	First Citizen Science workshop	WP5 NWO	M18 July 2020	M30 July 2021
MS37	Second Citizen Science workshop	WP6 OU	M30 July 2021	M33 October 2021

Also, to mention that following Deliverables and Milestones were submitted after scheduled date. Delays did not have any impact on other tasks listed under annex 1 description of work.

	Milestone /Deliverable Name	Responsible	Due date	Submission date
D3.5	Thematic training event - first school for software development and deployment in the EOSC	WP3 FAU	M29 June 2021	M30 July 2021
D6.4	Citizen science experiments with embedded educational resources (midterm)	WP6 OU	M24 January 2021	M27 April 2021
D3.6	Mid-term technology WP3 project progress report	WP3 FAU	M24 January 2021	M25 Febuary 2021
MS9	Second WP2 workshop to analyse the performance of the pilot, prepare D2.2	WP2 CERN	M22 November 2020	M23 December 2020
MS23	Progress and priorities at IVOA	WP4 CNRS	M22 November 2020	M23 December 2020

5.2 Use of resources (not applicable for MSCA)

End 2020 and 2021 were again marked with COVID-19 changing situations. The main issue faced was the diffiulty to forsee, anticipate and organise travels and events. This explains for every partner the amounts of Other Direct Costs budget still not used at that time of the Project. It was noticed during our september 2021 General Assembly and we took some time during October and November to discuss it with Partners. This budget not used would be for many the occasion to catch up with delays in some tasks and technical activities by transfering into extra Personnel budget.

Also during the second reporting period, following budget reallocations were approved from the project office and included in an Amendment (AMD-824064-32):

- Budget transfer from CNRS (Other Direct Costs) to ORB (Other Direct Costs) for General Assembly 2020 reimbursement of costs : 3662 euros.
- Budget transfer from CNRS (Other Direct Costs) to OpenUniversity (Personnel Costs and Other Direct Costs) to emphasise ESAP users support and maintain their crowdsourcing projects : 47500 euros.
- Budget transfer from CNRS (Subcontracting) to Trust IT Services (Personnel Costs) for Wavefier activity added: 37500 euros.

5.2.1 Unforeseen subcontracting (if applicable) (not applicable for MSCA)

Not Applicable

5.2.2 Unforeseen use of in kind contribution from third party against payment or free of charges (if applicable) (not applicable for MSCA)

Not Applicable

6 Terminology

Term	Explanation
AARNET	Australia's Academic and Research Network
ASTERICS	Astronomy ESFRI & Research Infrastructure Cluster
ASTRON	The Netherlands Institute for Radio Astronomy
CERN	European Organization for Nuclear Research
CEVO	Connecting ESFRI projects to EOSC through VO framework
CMS	Compact Muon Solenoid
CNRS	Centre national de recherche scientifique
CS	Citizen Science
CS3MESH	Cloud Storage Services for File Synchronization and Sharing
CTA	Cherenkov Telescope Array
DG CNECT	Directorate-General for Communications Networks, Content and Technology
DG RTD	Directorate-General for Research and Innovation
DIOS	Data Infrastructure for Open Science
DIRAC	Distributed Infrastructure with Remote Agent Control
DL3	Data Level 3
DMP	Data Management Plan
EC	European Commission
ECFA	European Committee for Future Accelerators
ECO	Engagement and Communication
EEAB	ESCAPE External Expert Advisory Board
EEB	ESCAPE Executive Board
EGO	European Gravitational Observatory
ELT	Extremely Large Telescope
EOSC	European Open Science Cloud
ESA	European Space Agency
ESAP	ESFRI Science Analysis Platform
ESCAPE	European Science Cluster of Astronomy & Particle Physics ESFRI research infrastructures
ESFRI	European Strategy Forum on Research Infrastructures
ESO	European Southern Observatory
EST	European Solar Telescope
EUDAT	European Data Infrastructures
FAIR	Findable, Accessible, Interoperable, reusable
FAIR	Facility for Antiproton and Ion Research
FAU	Friedrich-Alexander University
FRB	Fast radio burst
GRB	Gamma-ray bursts
GSI	GSI Helmholtzzentrum für Schwerionenforschung

HESS	High Energy Stereoscopic System
HL-LHC	High Luminosity Large Hadron Collider
HPC	High-performance computing
HPC	High-Performance Computing
HTC	High-throughput computing
IFAE	Institute of High Energy Physics
IN2P3	Institut national de physique nucléaire et de physique des particules
INFN	Instituto Nazionale di Fisca Nucleare
IVOA	International Virtual Observatory Alliance
JIV-ERIC	Joint Institute for Very Long Baseline Interferometry European Research Infrastructure Consortium
KM3NeT	Cubic Kilometre Neutrino Telescope
LAPP	Laboratoire d'Annecy de Physique des Particules
LOFAR	Low-Frequency Array
LSST	Large Synoptic Survey Telescope
MAGIC	Major Atmospheric Gamma Imaging Cherenkov Telescopes
MIND	Management, Innovation, Networking and Dissemination
MPE	Mass participation experiment
NIkhef	Dutch National Institute for Subatomic Physics
NuPPEC	Nuclear Physics European Collaboration Committee
OSSR	Open-source scientific Software and Service Repository
PRACE	Partnership for Advanced Computing in Europe
RDA	Research Data Alliance
RuG	University of Groningen (Rijksuniversiteit Groningen)
SKA	Square Kilometre Array
SNe	Supernovae
TDE	Tidal Disruption Event
TSP	Test Science Project
VERITAS	Very Energetic Radiation Imaging Telescope Array System
VO	Virtual Observatory
WLCG	Worldwide LHC Computing Grid
WP	Work Package
XDC	eXtreme DataCloud

7 Appendix 1: External expert advisory board review and recommendations - M13 (March 2020) M31 (September 2021)

Introduction to the Report

The ESCAPE External Expert Advisory Board provides external advice and evaluation of the achievements of the project. The Board brings additional expertise to the project, comments on its progress and results, vision from thematic national research institutes, their encompassed communities as well as from the pan-European space research field, to orientate activities towards a full achievement of the goals of ESCAPE with the most inclusive approach.

In the first year of the project, WP1 had concluded the initial appointment of the (EEAB) members. They delivered a first report in March 2020. Since then, two of the five members changed, and the EEAB is currently comprised by the following colleagues:

- Christophe Arviset, ESA – European Space Agency
- Andreas Haungs, Chair of APPEC - Astroparticle Physics European Coordination committee Karl Jakobs, Chair of ECFA - European Committee for Future Accelerators
- Marek Lewitowicz, Chair of NuPECC - Nuclear Physics European Collaboration Committee Colin Vincent, Chair of ASTRONET Board - Astronomy European Collaboration

Brief details on the profile of each of the EEAB members are presented at the ESCAPE web portal. Since the renewed appointment, the EEAB was invited to the H2020 ESCAPE progress meeting as well as the second ESCAPE General Assembly meeting held online on 29 September 2021. Unfortunately, the schedule was delayed, so a dedicated EEAB meeting was not possible. Nevertheless, the event provided the information to the EEAB members on the project objectives as well as the status of the project since the kick-off and the last general meeting. The EEAB had a closed meeting early November 2021 where a first discussion on this report had started. In this document, we present an executive summary of the review of the EEAB members on the deliverables and activities since the last report in March 2020 as well as an assessment of the implementation of the recommendations of this last report.

EEAB Review

In this section we present the summary of the EEAB review, which was developed and approved by all the EEAB members. The review is focusing on the project progress since the last report in March 2020:

Overall H2020-ESCAPE is challenging but is proceeding with adequate speed. After the first half of the project period, the ESCAPE teams have established an excellent portfolio overview of the landscape of ESFRIs and RIs involved in astrophysics, astroparticle physics, nuclear phyiscs and particle physics, and of their computing and software infrastructure profiles.

Synergies and similarities as well as different levels in the tools of the individual research areas were identified. The aim now is to further develop the strengths of each community and make them available to the others. However, there is also a risk that certain communities will not be taken along in certain aspects. ESCAPE's strategy to counteract this is to use User Cases or Test Science Projects (TSP) to create a manageable cross-cutting example and make it accessible to users.

This approach is advocated and supported by the EEAB, which therefore sees the TSPs as an essential and critical element to ensure the ESCAPE will deliver useful and efficient services to serve its beneficiaries, i.e. the users of ESCAPE.

The project is properly implementing the guidelines of the EOSC and will be key to giving it meaning for the wide community of astro- and particle physicists. At this stage, there is good progress in all areas with many of the objectives well underway and no particular cause for concern.

After the first year of activities, the interfaces between the ESCAPE Work Packages have become more and more important for the success of the whole project. There were several initiatives for the project management to work out the interconnections between the WPs. These initiatives were very welcome but should be intensified in the future. With a view to a successful ESCAPE project, it is essential to continue to promote and establish coordination across adjacent disciplines that share similar data-related challenges and that have overlapping scientific ambitions.

Review and recommendations on the work and deliverables by work packages:

WP1 Management, Innovation, Networking and Dissemination (MIND):

WP1 MIND has a strategic role to play, by connecting ESFRI projects to the EOSC and setting a basis for potential new collaborations within the EOSC future landscape. Special attention needs to be paid to the networking activities (EOSC stakeholders, Research Infrastructures, National Consortia, ESFRIs...), to be expanded and reinforced in order to identify new directions and development plans after January 2023 (end date of the ESCAPE project).

Coordination among WPs needed to be improved in the last year, and still needs to be increased. As each WP is led by a partner coming from a particular community, the partners of the other communities need to be engaged.

> *REC: WP1 should continue to increase coordination among WPs to ensure full engagement of all communities in all WPs.*

Two TSPs have been presented but without providing details how they will make use of the various infrastructure provided by the ESCAPE WPs.

> *REC: a dedicated report should be provided to explain how both TSP (Dark Matter and Gravitational Waves and Extreme Universe) will make use of infrastructure provided by the DIOS, OSSR, CEVO, ESAP and how they will enable engagement through ECO.*

In the course of developing conceptual data management plans for the research fields involved, AAI and the handling of data licences are an important topic. MIND is here in close

contact with both, e-infrastructures and research infrastructures, which is welcomed by the EEAB.

> *REC: a short report should be provided on best practises in data management plan aspects (in particular AAI, GDPR) across the participating research fields.*

WP2 Data Infrastructure for Open Science (DIOS):

Since the last review, a data lake pilot with 10 storage points has been deployed with strong involvement from the RIs and experiments. The AAI infrastructure is also deployed. Registration as an ESCAPE IAM member is required to access the data lake, but this restriction might be softened in the future. It is to be noted that this is a prototype for federating resources existing in different places. To be able to scale to Exascale, scientific collaborations will have to deal on their own with the storage and computing parts. The service that deals with data management is Rucio (CERN). A concern is raised for the current ability of the data management services to deal with rich, structured metadata for querying a science archive. It further needs to be assessed if Rucio is the right solution for all, or new services will need to be developed by the collaborations.

The DIOS data lake is based also on the CERN Rucio platform which seems very adequate for particle physics, and some indication of its use for SKAO and LSST future project, but full demonstration of its ability to deal with astronomical data (and associated metadata) coming from other partners (e.g. CDS, INAF) was not clear. More connection with the CEVO WP would be required to ensure proper interconnection between the data lake and data access in particular for astronomical data.

The next step is a notebook integration providing data lake as a service. We recommend an in-depth analysis for this step, by considering the users' profile and software needs, the current tools already used by the various communities and the requirements (technological, financial and decisional) for an end-user Notebook to be implemented and integrated with the data lake. Synergies with external projects and new communities should also be evaluated, to foster continued development and take-up beyond the current ESCAPE project.

There is also the question on the maintenance and further development beyond ESCAPE. The data management ecosystem (the data lake) should remain one of the core aspects of any follow-up programme of ESCAPE.

> *REC: DIOS should provide details and concrete examples on how ESFRI data currently in their science archives can efficiently be stored and accessed through the ESCAPE data lake.*
>
> *REC: DIOS and CEVO WP should work more closely together to demonstrate full interoperability between the ESCAPE data lake and data / metadata storage and access through VO protocols.*
>
> *REC: DIOS should list the data required to implement the 2 TSPs, where they will be stored and how they will be accessed.*

WP3 Open-source scientific Software and Service Repository (OSSR):

Following the recommendations from the last report, there has been an increase in the level of communication between WP3 and WP4 via the 'onboarding' strategy of the IVOA software. A very successful machine learning training activity (summer school) with more than 1000 participants was held in 2021. We encourage pursuing the software onboarding procedure for the partners and its extension to other interested communities, by defining clear standards required for integration to OSSR. To leverage the cross-fertilization between the ESFRIs, common approaches (libraries, data format, etc.) should be clearly formalised. The multi-messenger approach is crucial and we stress the importance of developing innovative methods for data analysis and machine-learning algorithms, together with a common approach for the software development and implementation techniques.

WP3 has already identified a good list of software which have been made part of the OSSR and should continue to do so, in particular in identifying the list of software required to support the two TSPs.

> *REC: OSSR should list the software that will be required to support the two TSPs.*

Regarding the software itself, each domain inevitably has its particular needs, tied closely to details of the instruments and data formats, which make quite a few of the higher level tasks quite specific to each domain area. We feel that cooperation with the HEP Software Foundation (HSF) would be very useful to exploit synergies. In addition, there is a rather generic aspect to the challenge of scientific computing on modern hardware (CPUs, GPUs, FPGAs) where a lot of knowledge can be usefully exchanged. Common low level projects that would allow an engagement with HEP/HL-LHC should be considered.

> *REC: OSSR should engage with the HSF and e.g. with the Software Institute for Data Intensive Science (SIDIS), which is a European focussed software engineering initiative.*

We appreciate a lot that the two TSPs are looking at demonstrating analysis pipelines in a multi-science project and how that machinery integrates with the data lake. We encourage ESCAPE that the work remains focussed on real use cases and that the work continues to be done with a direct participation of the experiments / users. Also here the HSF could be involved providing input and feedback, e.g. for reproducibly sustaining a complex workflow.

WP4 Connecting ESFRI projects to EOSC through VO framework (CEVO):

The objective of this WP is to implement the FAIR principles for the astronomy data in order to support Open Science. Latest developments include the connection of the VO registry to the EUDAT B2FIND. The priorities of the RIs regarding the IVOA standards have been assessed, nevertheless we recommend continuing close discussions and exchanges between the research infrastructures needs and the IVOA standards updating, to ensure these needs are fulfilled. Also, we recommend to pursue a tight collaboration with the other WPs, and mostly with WP5, which can benefit from input provided by the new IVOA communities, e.g. Astropy, a Python library for astronomy, or Gammapy, a similar tool for astroparticle physics.

WP4 continues to be very active in defining interoperability protocols mainly aimed at dealing with astronomical data and metadata format and exchange. Further interactions with some other WPs needs to be reinforced to ensure that the VO protocols are compliant with other areas of ESCAPE.

REC: See REC on DIOS/CEVO in the DIOS section: CEVO should identify the list of VO protocols that will be required to support the two TSPs

REC: CEVO shall identify which VO protocols will be required to support the two TSPs.

WP5 ESFRI Science Analysis Platform (ESAP):

The ESAP is designed to be a science platform toolkit that adapts to the needs of the RIs. A prototype is already available. It is important to have in mind that the WP5 will only deliver the ESAP, not operate it. Currently it is not clear how this issue will be solved in the future. We therefore recommend some directions are provided by the WP5 teams at the end of the project, and details about the technical requirements to be able to operate the ESAP. Also, as ESAP is not intended to store bulk data, it is very important to carefully plan and implement the content delivery layer and its interface with the Data Lake.

WP5 describes the ESAP as a hub of decentralised science platforms, which on one side will indeed provide some flexibility in implementation, while on the other side, how all these platforms will be interconnected and their interface with OSSR and DIOS is still not totally clear. In addition, how ESAP will make use of EOSC resources would also need to be further described.

REC: ESAP should describe how the science platforms will interconnect between themselves and how will they interface with OSSR and DIOS

REC: ESAP should describe how it will connect to EOSC.

Furthermore, to help understand how ESAP will support end users in practice, it would be really useful to understand how ESAP will actually support the two TSPs.

REC: ESAP should describe how the science platforms will support the two TSPs

WP6 Engagement and Communication (ECO):

The plan to include training in citizen science activities is excellent and provides an added value for the EOSC communication projects, by attracting new users and engaging more with the experts, helping them switch from citizen science tools to professional ones. ECO provides a very interesting list of community engagement citizen science projects, although it is not clear how they actually benefit from the newly developed ESCAPE infrastructure.

REC: ECO should detail how its actions benefit from the existing ESCAPE infrastructure provided by DIOS, OSSR, CEVO and ESAP and/or how ECO's needs are used as inputs to the other WPs work plans.

Another comment is concerning citizen science, which has already been very successful in a number of areas of astronomy and is well-connected to non-physics endeavours. The EEAB sees a great chance for visibility, but feels that the work here is treated a little too peripheral to the main aims.

> *REC: ECO should give some more thoughts to how the work underway can be transformed into citizen science activities.*

Summary and next steps

The EEAB recognises the good progress made by the ESCAPE project and only needs to ensure it addresses a few course corrections, mainly around coordination and communication. We reiterate the enormous importance of the engagement of the entire research community in the digitisation of the research field and the provision of the appropriate tools and services. Therefore, a structured communication between ESCAPE and the researchers, developers and the infrastructures, as well as users and the public is important. In particular, the links to the infrastructures are an important part of this - we need to get SKA, HL-LHC, FAIR, CTA, etc. to see ESCAPE developments as critical to their success.

The meanwhile selected Test Science Projects (TSP) are seen to be fundamental for the success of ESCAPE. TSP are critical to test the central tools under development (data lake and software catalogue, e.g.), while at the same time having a functional role to connect researchers to the EOSC.

In general, the ESCAPE project, together with its partners, has to verify how to sustain the EOSC vision in the ESFRIs/RIs after the ESCAPE project. To ensure this, the EOSC principles must be pro-actively embedded in the research and organisation culture of the communities. This is a major but visionary challenge for the ESCAPE project, but for which it is well aligned.

The EEAB supports the plans of the ESCAPE management to think about the sustainability of this important project and to initiate first measures in this direction. The establishment of an MoU between the directorates of the research infrastructures to operate a sustainable "ESCAPE cluster platform" beyond the duration of the current H2020 Grant is indeed a first step. This can also imply the inclusion of new RIs. Important as well is a coherent approach for sustainability in every of the five Science Clusters. The EEAB strongly recommends to actively follow these plans and supports all measures to reach a coherent, sustainable and FAIR Big Data Science of the research field. To give more emphasis to this point, we have drafted a letter of support (see appendix), which is available for appropriate action.

Finally, a comment on our own behalf: The project management should ensure that at ESCAPE Progress meetings, the External Expert Advisory Board has enough time to elaborate their recommendations during a dedicated closed session. The preparatory documents should be available for the EEAB at least 10 days prior to the meeting.