Project Title	European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures
Project Acronym	ESCAPE
Grant Agreement No	824064
Instrument	Research and Innovation Action (RIA)
Topic	Connecting ESFRI infrastructures through Cluster projects (INFRA-EOSC-4-2018)
Start Date of Project	2019-02-04
Duration of Project	42 Months
Project Website	www.projectescape.eu

White Paper – ESCAPE Work Package 5: Achievements and Future Prospects

Work Package	WP5, ESFRI Science Analysis Platform
Lead Author (Org)	John D. Swinbank (ASTRON)
Contributing Author(s) (Org)	
Due Date	2022-05-30 (M40)
Date	2022-05-26
Version	0.2

Dissemination Level

- [X] PU: Public
- [ ] PP: Restricted to other programme participants (including the Commission)
- [ ] RE: Restricted to a group specified by the consortium (including the Commission)
- [ ] CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Revision	Date	Description
0.2	2022-05-26	Draft for distribution to WP5
0.1	2022-05-04	Early draft shared with VRE team

Disclaimer

ESCAPE – European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

List of Abbreviations

CEVO Connecting ESFRI Projects to EOSC through VO framework.

CTA Cherenkov Telescope Array.

DIOS Data Infrastructure for Open Science.

EOSC European Open Science Cloud.

ESAP ESFRI Science Analysis Platform.

ESCAPE European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures.

ESFRI European Strategy Forum on Research Infrastructures.

IVOA International Virtual Observatory Alliance.

OSSR Open-source scientific Software and Service Repository.

PID Persistent Identifier.

SAMP Simple Application Messaging Protocol.

SKA Square Kilometre Array.

VO Virtual Observatory.

VRE Virtual Research Environment.

WP Work Package.

ESCAPE – European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

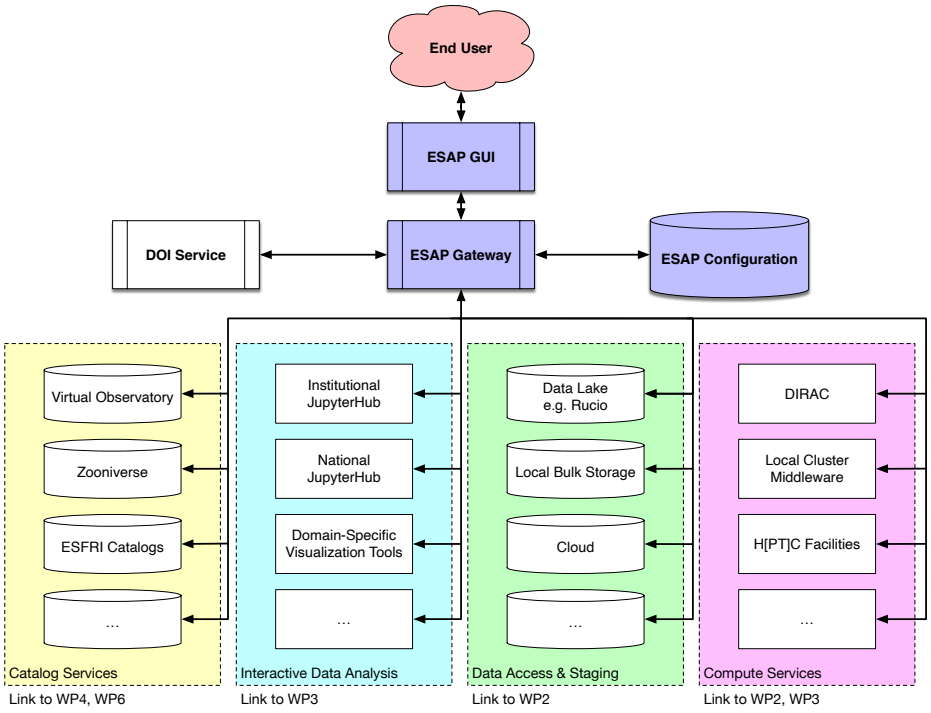

Figure 1: ESAP provides a single, consistent, interface and point of access to a variety of services drawn from a range of providers. Links to the other work packages ESCAPE are indicated.

1 Introduction

Activities in ESCAPE 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 and across EOSC.

This document is structured as follows. Section 2 briefly summarizes the overall vision for the ESAP system, describing what is being built, and why it is useful. The current status of, and major achievements in, ESAP construction are described in §3. Finally, §4 considers possible future directions, including both specific technical work on the existing ESAP codebase (§4.1) and a consideration of follow-up activities with wider potential impacts (§4.2).

2 The ESAP vision

ESAP is an *science platform toolkit*: an integrated set of software components which ESFRIs, ESCAPE project partners, and other groups can use to rapidly assemble and deploy platforms that are customized to the needs of their particular user communities and which integrate their existing service portfolios. These various deployed instances of ESAP then provide the key interfaces between the services delivered by the ESCAPE project and the wider scientific community.

In order to meet this goal, ESAP has been designed to be flexible and adaptable to the particular needs of each user community. This is achieved my abstracting the details of heterogeneous underlying infrastructures

ESCAPE – European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

away from users, a task that is achieved by ESAP's modular, plugin-driven architecture: an ESAP instance is integrated with its surrounding environment by enabling and configuring an appropriate selection of plugins, and new capabilities are easily added by writing new plugins.

While this plugin based system makes ESAP infinitely reconfigurable, its base level of functionality includes:

- Data discovery and retrieval from a range of archives and data repositories;
- Exploration and discovery of relevant tools within the ESCAPE software repository;
- Access to a range of compute and analysis services provided both by project partners and by other facilities;
- Orchestration of data, services, and software to help users create and access research environments that meet particular needs.

This model is shown schematically in Fig. 1, which illustrates the range of services that ESAP can help the user access. The flexibility of this approach means that instances of ESAP can be deployed at a variety of scales, from providing services to just a few users within a small project, to supporting major pieces of infrastructure.

3 Achievements to date

3.1 Platform capabilities

The core ESAP system — the user interface, the API Gateway, and a selection of plugins connecting it to key services — are now mature and well tested. This includes:

Core data management capabilities
As the user accumulates and manipulates datasets, they accumulate results and carry them with them through the system. Subsequent analysis tasks augment this data collection. Ultimately, it can be persisted for future references and, if appropriate, published for wider use.

Data archive interfaces
The user is able to use ESAP to search for and discover data in a wide range of archives. While this certainly includes archives which are built upon open standards — and, in particular, VO interfaces as described below — it is possible to incorporate bespoke and non-standard data sources by writing appropriate plugins. For example, plugins are currently available that provide access to data collections from Apertif[1] and Zooniverse[2], amongst others. Note that the ESAP interface adapts appropriately depending on the data collection being queried, and even makes it possible to search multiple semantically-related archives simultaneously.

Interactive data analysis facilities
The user can spawn interactive analysis jobs based on Jupyter[3] running at a range of different computing facilities. These jobs can integrate with the other ESCAPE work packages, as described below.

[1] https://alta.astron.nl
[2] https://www.zooniverse.org
[3] https://jupyter.org

ESCAPE – European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

Data Lake integration (DIOS, WP2)
ESAP provides direct access to query the Data Lake, discovering data and working with it directly within the core ESAP interface. Further, the *Data Lake as a Service* system provides direct integration of the Data Lake with the interactive analysis environments available through ESAP.

Software Repository integration (OSSR, WP3)
The ESCAPE OSSR collects software from across the various ESFRIs associated with the project. Deep integration between ESAP and the OSSR makes it possible for ESAP users to discover this software and dispatch it — together with their chosen datasets – directly to selected analysis environments.

Virtual Observatory protocol support (CEVO, WP4)
ESAP provides pervasive support for VO standards, developed in conjunction with ESCAPE WP4. Users can locate and access data for processing in ESAP using VO systems, and even leverage the IVOA SAMP standard to have ESAP communicate with applications running on their personal computer!

Managed database support
ESAP's (prototype) managed database support makes it possible to fetch results from multiple remote catalogues into a local database that can be used for advanced analytics functionality like cross-matching diverse catalogues.

3.2 Software packaging and support

The WP5 team have packaged and integrated a wide variety of scientific analysis software, covering a variety of the ESCAPE ESFRIs, in forms that are designed for re-use and ultimately for integration with the OSSR. Space constraints make it impossible to list all of the various initiatives in this white paper, but illustrative examples include:

- Publication of solutions to the Square Kilometre Array (SKA) Science Data Challenge to the OSSR;
- Packaging and containerization of the R3BRoot and CmbRoot[4] tools;
- A reproduction package for the scientific analysis of Hickson Compact Group 16 described in Jones et al. (2021)[5];

3.3 Infrastructure deployment

Deployment and provisioning of infrastructure for the long term is not within the scope of WP5. However, the work package does support a number of software installations which have provided limited services to end users while acting as proofs-of-concept for the ESAP system. In particular, these include:

- ESAP core instances deployed at ASTRON and SKAO;
- JupyterHub[6] and/or BinderHub[7] systems deployed at CSIC, JIV-ERIC, FAIR, and RuG.

[4] https://www.r3broot.gsi.de and https://redmine.cbm.gsi.de/projects/cbmroot
[5] https://ui.adsabs.harvard.edu/abs/2019A%26A.632A..78J/abstract
[6] https://jupyter.org/hub
[7] https://binderhub.readthedocs.io/

ESCAPE – European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

These systems have been instrumental both in integrating ESAP itself and in testing scientific workflows within the associated user communities.

4 Future goals

This section considers future work in the ESAP context. We consider first the natural technical evolution of the codebase — areas in which technical work beyond that scoped for ESCAPE could have substantial impact. Then we discuss the wider impacts that future support for ESAP could have on the EOSC and ESFRI community.

4.1 Technical development

The current ESAP system, as described in §3, provides a robust and reasonably fully-featured environment. However, there are many opportunities for further enhancements that cannot be completed within the scope of ESCAPE. This include:

- Improved support for provenance tracking through the ESAP data management system, including robust minting of Persistent Identifiers (PIDs) for published results.
- Support for sharing data between users of the platform, enabling collaborative workflows.
- Support for persistent development environments, in which users are able to store the state of their session and return to it in future.
- Richer understanding of the links between data products, enabling features such as presenting science products together with the calibration data used to generate them.
- Federation between distributed ESAP instances.

In addition, ESAP remains endlessly adaptable to integrate with new external service offerings. Many opportunities for these arise: from the management of OpenStack[8]-based virtual machines to integration with workflow- or function-as-a-service systems.

4.2 Wider prospects

This section addresses the potential wider impacts of ESAP: how can continued development of the system continue to further the promise of EOSC and ESCAPE?

4.2.1 Development of an ESCAPE/EOSC Virtual Research Environment

The ESAP vision, as described in §2, provides the basis for a true *Virtual Research Environment*: an on-line collaborative system that provides all the tools and services that scientists required to conduct high-impact research. The flexible and scalable design of ESAP means that it can evolve into a role in which it provides seamless access to all of the various software and services which are developed by ESCAPE, by its successors, and across the whole EOSC ecosystem. However, assuming such a role requires two major ongoing activities.

First, ESAP will require continued maintenance and integration effort. The platform is stable, solid and extensible, but — as with any network connected service — emergent issues will need to be resolved as they arrive. Further, while the plugin-driven ESAP system is extensible to address a wide range of use cases, development

[8] https://www.openstack.org

effort will be required to integrate new service plugins and to extend the ESAP core to support new service types when necessary.

Secondly, the ESCAPE project provides no long-term structural support for delivery of ESAP as an operational service. Within the context of the project, various ESCAPE partners have deployed infrastructure in support of ESAP development or in support of particular ESFRI use cases. However, these are inappropriate for providing high-availability services to a more general community: that would require not only dedicated infrastructure, but also support and maintenance services.

4.2.2 Sustaining ESAP in support of ESFRI development

Several of the ESCAPE-affiliated ESFRIs are currently at a stage in their development where they are making strategic choices about the long-term delivery of services to end users. For example, the Cherenkov Telescope Array (CTA) is currently evaluating how to provide user-facing services in the context of its upcoming *Science Data Challenge 2*, while SKA is in the processing of beginning prototype development for its network of *Science Regional Centres*, which will be observatory's primary means of delivering data to end users. Both of these infrastructures have been heavily involved in ESCAPE in general and in the development of ESAP in particular; it would be natural for ESAP technologies to play a major role in their future plans.

No individual ESFRI will find it advantageous to assume responsibility for a legacy software system, even if that system provides useful functionality. The value of ESAP to these projects is therefore dramatically increased if it is not simply a collection of source code, but an actively maintained and supported project with a clear governance model and a sustainable future.

4.2.3 Common standards for science platform development and interoperability

The last several years have seen an explosion of heterogeneous development in the field of "science platforms" (broadly defined). Tens or hundreds of projects generally providing some form of interactive analysis environment with access to bulk storage and computing systems are being rolled out across a multitude of research domains[9]. This ecosystem is disjointed and fragmented: although many of these projects are individually very technically advanced and provide compelling functionality, they are specialized and difficult to apply in other domains.

The ESAP experience suggests a way to move beyond this fragmentation of platforms into disconnected systems serving individual communities or infrastructures. Instead, ESAP proposes a model of common standards and interconnectedness among science platform efforts. By using ESAP, Individual projects or communities which need particular capabilities can build on a common, interoperable technological basis, customizing and extending it to address just their particular needs.

Moving beyond that: as more platforms become centred around common technical standards, we can move to adopt common standards for access not only to data — as pioneered by the Virtual Observatory in the ESCAPE context — but to compute and other resources. This process has already started: members of the ESCAPE team have engaged with our colleagues in the IVOA to begin exploring concepts for the *ExecutionPlanner* system[10], which would standardize access to computing capabilities in much the same way as IVOA recommendations harmonize access to data.

[9] For example, consider CANFAR Skaha, Rubin Science Platform, ESA Datalabs, NOIRLab Astro Data Lab, SciServer, and many others.

[10] https://github.com/ivoa/ExecutionPlannerNote

ESCAPE – European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

Ultimately, by building upon the experience developed in ESAP and CEVO, one can imagine working towards a future federated network of science platforms, build on common standards, and providing the lowest possible barrier to entry to new service or data providers making their offerings available to a wide range of researchers.

ESCAPE – European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.