Project Title	European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructure
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	04.02.2019
Duration of Project	42 Months
Project Website	https://projectescape.eu/

White Paper - ESCAPE Work Package 3 (OSSR): Achievements and Future Prospects

Work Package	WP3 OSSR
Lead Author (Org)	Kay Graf (FAU)
Contributing Author(s) (Org)	Jutta Schnabel (FAU), Gareth Hughes (CTAO), Thomas Vuillaume (LAPP), Mark Kettenis (JIVE)
Due Date	30.05.2022 - M40
Date	10.06.2022
Version	1.0

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

Version	Date	Authors and Contributors	Notes
0.1	27.05.2022	Kay Graf	Setup document and plan
1.0	10.06.2022	Kay Graf, Gareth Hughes, Mark Kettenis, Jutta Schnabel, Thomas Vuillaume	Implement comments from partners

Disclaimer

ESCAPE - The 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.

Acronym list

EOSC:	European Open Science Cloud
ESCAPE:	European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures
ESFRI:	European Strategy Forum on Research Infrastructures
ESF/RI:	ESFRIs and major RIs as projects within ESCAPE
FAIR:	Findable, Accessible, Interoperable, Reusable
OSSR:	Open Science Software and Service Repository (ESCAPE WP3)
RI:	Research Infrastructure
VO:	Virtual Observatory
WP:	Work Package

Table of Contents

1. Introduction

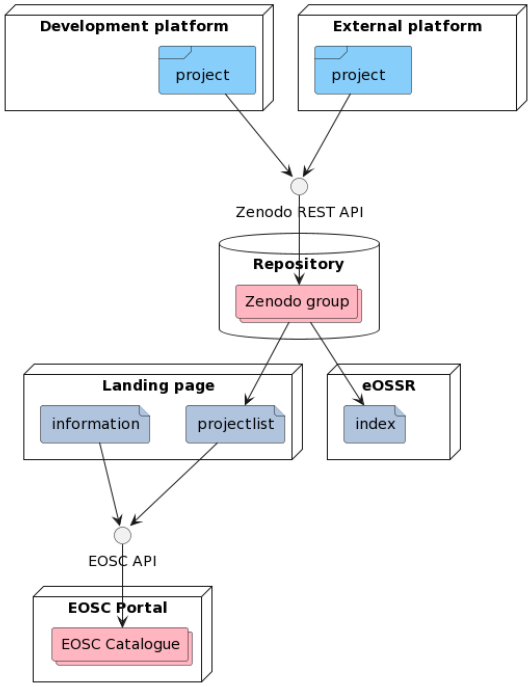
Activities in ESCAPE Work Package 3 - OSSR - are broadly divided into three major areas:

1. Support a community-based approach for continuous development, deployment, exposure and preservation of domain-specific open-source scientific software and services in the global context of the EOSC catalogue of services - the OSSR itself;
2. Enable open science interoperability and software re-use for the data analysis of the ESCAPE ESFRI projects based on *FAIR* principles;
3. Create an open innovation environment for establishing open standards, common regulations and shared software libraries for multi-messenger/multi-probe data;
4. Educating stewards for FAIR software by knowledge transfer, collection of best practices and software schools.

This document is structured as follows. Section 2 briefly summarizes the overall vision for the OSSR. The current status of, and major achievements in, the OSSR implementation is described in Sec. 3. Finally, Sec. 4 considers possible future directions, including both specific technical work and a consideration of follow-up activities with wider potential impacts.

2. The OSSR Vision

The OSSR vision is to establish a trustable, sustainable repository for software and services and to foster collaboration on the co-creation of high-quality, open-source software for open science. Contributing to the OSSR enhances software quality through guidelines and recommendations for software contributions, enables the development of interoperable use of the software through extended metadata and increases findability of software through OSSR integration in the EOSC and other research environments. It thus makes software a first class citizen of open science and the EOSC.



3. Achievements to Date

Architecture

The OSSR is based on collecting software in the form of a code repository or container image in a Zenodo community[1], which is connected to a specialized landing page and searchable via a python client library.

The landing page[2] is the entry point of users to the OSSR products, as well as to other services within the ESCAPE EOSC cell. It also contains links to documentation and training materials.
The development platform[3] provides a common place to gather the common developments, ideas,

[1] https://zenodo.org/communities/escape2020/
[2] http://purl.org/escape/ossr
[3] https://gitlab.in2p3.fr/escape2020/wp3

ESCAPE - The 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.

guidelines and templates for the community, as well as a platform for new developments if required by an institution/group without access to another solution. It showcases the full software lifecycle up to the publication in the repository. The development platform is not to substitute the development platforms already used by each institution/group, the technical enhancements necessary to create project links to the OSSR can be equally well applied on the respective native development platforms.
The repository backend of OSSR is the ESCAPE2020 Zenodo community[4]. It holds the long-term archived open science projects developed in the platforms. Each OSSR record is required to have an additional codemeta.json-file to hold the extended metadata that describes the software.
To facilitate easy access, the eOSSR Python library[5] 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 and functions to help developers automatically contribute to the OSSR, in particular using their continuous integration.
Integration of the OSSR in aggregating portals will increase the visibility and findability of the OSSR entries. To this end, the OSSR will be integrated in the EOSC portal. Additional portals will be considered, and already the OSSR has been integrated to re3data[6], the registry of research data repositories.

Onboarding

Partners in the ESCAPE project and the wider science community are encouraged to onboard their scientific software, public datasets (limited in size), container images, or repositories with full analyses environments to the OSSR[7]. They are requested to complete an onboarding process which involves the curated presentation of their project and upload of the contribution to the Zenodo community, triggering a short review process. Although currently focusing on ESCAPE-related projects, contributions are encouraged from any related research field, as is already exemplified by a cooperation with the DMA project in the Helmholtz research program Matter.

Content

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

[4] https://zenodo.org/communities/escape2020
[5] https://escape2020.pages.in2p3.fr/wp3/eossr/
[6] https://www.re3data.org/repository/r3d100013827
[7] https://escape2020.pages.in2p3.fr/wp3/ossr-pages/page/contribute/onboarding/

ESCAPE - The 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.

Best practices

In order to collect best software practices, a workshop was organized with contributions from a wide field of scientific software experts[8]. The outcome of the workshop is made available in the OSSR and used to inform contributors about recommendations on software development. Focus topics include software interoperability, development strategies, licensing practices and software solutions. In the Guidelines and rules of participation to the ESCAPE OSSR[9], this lead, amongst others, to recommendation of an open source license, the requirement to provide documentation alongside the software, and ensure software provenance through adequate version control.

Transmitting best practices and know-how

Producing FAIR software in practice requires know-how. The ESCAPE Data Science schools organised by OSSR transmit such knowledge and educate software stewards. During these schools, scientists in the field of astronomy, astro-particle and particle physics are taught the necessary ingredients for their software to become a part of open science by experienced code custodians. The 2021 special online edition welcomed more than 1000 registered participants. Following the FAIR paradigm and as an example of good practices in code development, all the school material is openly available online, including scientific programme, agenda and links to all contributions (software repository, notebooks, presentations and recordings).[10]

Cooperation

An exemplary effort to foster cooperation is given in the ConCORDIA project to produce CORSIKA turnkey containers for various use cases in astroparticle physics and linked research fields shared for research conducted with primary and secondary detectors underwater or in ice and for muons in the low atmosphere. It involves setting up the CORSIKA containers and running test productions in various computing centres and assessing and certifying the quality and physics relevance domain of the simulation.

Cross-fertilisation

Although cross-fertilization occurs on various levels in the work package, a special focus was put on the establishment of an Innovation Competence Group which was finalized by the organisation of an all-hands meeting, implemented as an online workshop IWAPP - Innovative Workflows in Astro & Particle Physics[11].

The construction of the innovation group started with a series of 12 dedicated meetings of the Focus Group 3 and Task 3.4 of OSSR, where the different ESCAPE partners presented their activities in terms of innovation in data management, software and data analysis, in particular regarding the use of artificial intelligence and especially in the form of deep-learning techniques for querying large data archives, pre-processing data, object classification and parameter inference. The group continues to be active on specific topics, especially in regard of establishing a common project in the context of EOSC-future.

[8] https://escape2020.pages.in2p3.fr/wp3/wossl/
[9] https://escape2020.pages.in2p3.fr/wp3/ossr-pages/page/contribute/guidelines_ossr/
[10] https://doi.org/10.5281/zenodo.5838436
[11] https://indico.in2p3.fr/event/20424/

ESCAPE - The 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.

4. Future Goals

This section reflects the future work within the OSSR context.

4.1 Technical Developments

The current OSSR architecture, see Section 3, provides a robust and flexible repository implementation. However, there are opportunities for further enhancements that cannot be completed within the scope of the ESCAPE project, as:

- extending the CodeMeta schema beyond the current implementation to keep track of further developments within the repository (see also 4.2);
- extending the eOSSR library allowing for enhanced searches, additional development platforms and archives;
- Support for a seamless integration with the ESAP and Virtual Research Environment and the full support of the outcomes of the ESCAPE/EOSC-Future Test Science Projects.

4.3 Sustainability

The long-term sustainability of the OSSR service was a design requirement from the beginning. Several choices have been made so that the individual components have an impact also in the future even if OSSR as a service itself should not be maintained, this is e.g. for the simple interfaces between the different components, the metadata standard and the archive in Zenodo.

Several actors in the current OSSR focus group on software and service collection have expressed their interest to join a *Curation interest group* that continues the onboarding - on a lower level - also after the ESCAPE project. More importantly, the maintenance of OSSR is also one of the goals of the recently funded ESCAPE collaboration.

4.2 Future integration and cooperation

One major future goal is that a repository in the form of the OSSR - or future developments of it - become a natural part of the EOSC Exchange Layer or even the EOSC core services and the interoperability between repositories part of the EOSC Interoperability Framework. This will be a medium-term goal for the future integration of OSSR.

For this, the EOSC architecture integration should follow the recommendations of the EOSC working group on *Scholarly infrastructures for research software*[12] and the current *EOSC Task Force on Infrastructure for Research Quality Software*.

Repositories as the OSSR should be sustained as essential part of the infrastructure for scientific software. The collaboration also with other communities should be further pursued and the architecture harmonised. This should be a mid-term future effort between the different EOSC Clusters or their follow-up organisation.

In a long-term vision, an open-science software repository could be lead as a foundation.

[12] European Commission, Directorate-General for Research and Innovation, Scholarly infrastructures for research software: Report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS, Publications Office, 2020, https://data.europa.eu/doi/10.2777/28598

ESCAPE - The 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.