Ian Bird, LAPP
17/06/2022

European Open Science Cloud

Executive Summary

Over the last decade, the understanding of the importance of making science more open has grown, hand in hand with the recognition that scientific data is an important resource for innovation and development. The concepts of Open Science and FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) data have become important touchstones in recognising the value of science and in exploiting its results, to the eventual benefit of the broader community that is funding and supporting it.

The European Open Science Cloud (EOSC) is not a project in the usual sense, but rather an encompassing framework for Open Science and FAIR data in Europe. As such it is being developed through a number of individual projects funded by the European Commission. Several of these have run over the last several years under the Horizon 2020 Framework Programme, and in future anticipated funding under the Horizon Europe Programme. The most significant project at the moment is EOSC-Future, which builds on and integrates the results of previous projects, and importantly brings together the scientific communities and the e-infrastructure organisations that have been built up over many years. The scientific communities are represented through the five science clusters which themselves are important collaborative projects of domain-based science.

These cluster projects represent the domains of life sciences (EOSC-Life), environment (ENVRI-FAIR), social sciences (SSHOC), photon and neutron science (PANOSC), and Astronomy, Particle, and Nuclear Physics (ESCAPE). An important development is the realisation of the many synergies between these clusters and the desire of the clusters to work together to influence the direction of the EOSC and the relevant funding agencies. Although the EOSC is a European endeavour, it is important to recognise that all of the Research Infrastructures represented by the clusters are international or global in nature, and that the EOSC cannot exist in isolation, but must connect with other Open Science initiatives globally.

In addition to EOSC-Future there are several other projects running in parallel, specifically to provide resources and input to EOSC-Future. These projects are mainly extensions of the existing e-infrastructures but with specific goals related to EOSC implementation.

Although the EOSC is being implemented through these projects, which will be further described in this document, the long term oversight and direction is foreseen through the EOSC-Association, which is a legal entity established to govern the EOSC. The vision for the long-term sustainability of the EOSC is vital for the real engagement of the scientific communities, and must be a high level goal of the Association as well as of the projects and organisations involved.

This paper will describe the problems addressed by the EOSC initiatives, describe some of the precursor projects, the cluster projects and joint efforts, the EOSC-Future and associated projects, and conclude with an outlook for the future including risks and benefits.

Scientific Rationale

The vision of the European Open Science Cloud is to create a federated and open environment for European researchers, ultimately also enabling access to industry, innovators and citizens. It is, by design, multi-disciplinary and cross-domain, providing facilities to enable researchers and users to "publish, find and reuse data, tools and services for research, innovation and educational purposes"[1]. There are a number of high-level aspirations that the scientific communities are expected to contribute to:

- Seamless access to research outputs,
- FAIR (Findable, Accessible, Interoperable, Reusable) data management,
- Reliable reuse of research outputs and digital objects (data, algorithms, software, publications, etc.)

The implementation of the EOSC is a long term activity, and has been underway since 2018, with a number of projects that were designed to explore some of the baseline ideas of EOSC. As part of that, five domain-based science clusters were set up, grouping together the ESFRI projects in each domain. These science clusters are the main driving force of the EOSC from the scientific communities, and as such are extremely important, since the EOSC must be attractive to, and used by, the communities represented as part of their integrated environments in order to guarantee success of EOSC. Without that engagement of the science communities, EOSC will not be sustainable.

Thus, in addition to the high level aspirations of open science mentioned above, it is key that the EOSC provides value to the scientific community. This value can take the form of specific value-added services, such as a common Authentication, Authorisation Infrastructure (AAI) and similar fundamental services useful for all scientists; or resources provided free at the point of use to enable novel cross-discipline scientific activities that do not necessarily fit within the usual funding envelopes of individual RIs. In addition, it is important to recognise the building of synergies between disciplines and the capability of sharing not only experience and expertise, but also software and services developed by the individual science communities. There is tremendous value to be shared between communities. Ultimately, having a structure such as EOSC should enable more cost-effectiveness through common service layers that can support and evolve with the needs of all of the scientific communities. However, experience shows that care must be taken to understand what is really common and what really belongs in the domain of each application area to enable innovation and change with the evolving scientific needs.

Sharing and exchanging data between communities is a stated goal of the EOSC, and mechanisms to enable that are being developed and put in place by several of the EOSC-related projects. A current example that demonstrates the importance of this has been the

[1] https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en#documents

recent collaboration between several of the science clusters on the response to the Covid-19 crises. Such collaborations will be of benefit in many of the big societal challenges, such as climate, energy, and disease management.

In addition to the recognised scientific communities that expect to be the main customers of EOSC, there are significant opportunities to expand the use of citizen science as a research tool, as well as as public engagement in science. Making the data and software open and available, and building on many years experience in citizen science is valuable to both scientific communities and the interested scientific public. Again, sharing, and pooling expertise across communities can benefit this. At the same time this strategy will enable more scientists, not formally part of the RIs, to benefit from the data and potentially develop new approaches and new analyses. Of course this goes hand in hand with having an appropriate policy for data openness, which must foresee, where necessary, periods of embargo where the data is available only to the science community who have invested all the effort in producing it, before it is published openly.

It is widely recognised today that properly curated scientific data has a significant value, and the EOSC should engender an environment that allows that data, from all of the disciplines, to be openly available, and able to benefit European science, public, and industry. It is fully expected that industry will engage with EOSC, both as data consumers but also as providers of procured services.

In terms of how the scientific communities will exist within the EOSC environment, the goal is to enable science by providing a common service and support layer, and providing common tools to enable the open publication of data. Thus it should be seen as a collaboration between sciences and e-infrastructure providers, and a federation of services and catalogues coming from each RI or cluster. For example, data catalogues of individual domains would not be “absorbed” into a common EOSC catalogue, but rather publish metadata to enable high level searches so that consumers can find data sets of interest. The EOSC Portal[2] provides a single entry point to negotiate and explore the EOSC ecosystem. Of course, however, the science communities will continue to work within their own environments, existing workflows, etc., while hopefully using more and more common services, and also contributing their services or software to the common pool. EOSC should not require disruptive change to the science activities, but rather offer interesting value to be integrated and used.

It is very important that the direction and strategy of the EOSC is guided by the needs of the scientific communities. For this reason the evolution of the science clusters, and their collaboration will be essential to enable the long term success of the EOSC ecosystem. There are still many questions to be answered around the long term sustainability of EOSC and how it will be funded, but it is widely recognised that a structure such as this is vital for the long term support of science through openness and engagement with stakeholders.

[2] https://eosc-portal.eu

First generation EOSC projects

Today the development of the EOSC is in the hands of the EOSC-Future[3] and associated concurrent projects, discussed later. However, it is important to realise that the development of the vision of EOSC and aspects of open science have been supported and developed by a number of precursor projects. Here we mention a few of the relevant ones that have built services in preparation for EOSC and that are being integrated through EOSC-Future.

EOSC is a natural development of the European landscape of Grid and e-infrastructure projects, as well as the ESFRI science projects. Early development through the "European Datagrid" project, followed by "Enabling E-Science in Europe" (EGEE), were driven by the novel data and global collaborative computing requirements of particle physics, but other sciences, notably life-science, quickly recognised the benefit of such common and collaborative structures. It was important to recognise that these efforts were coordinated with many national and international e-infrastructures. In France that structure is "France Grilles"[4]. That series of projects evolved into the EGI[5] project and later a permanent structure, in step with the national initiatives, also evolving technologies from early "grid" ideas to clouds for example.

Another notable project was OpenAire[6] which focussed on developing tools for open science. A number of other projects were also funded to develop various aspects of the EOSC. Here we give a brief overview of these, and their relevance to the EOSC and its evolution.

EGI

The EGI foundation was established in 2010 in Amsterdam, and is the coordinator of the EGI federation. The EGI Foundation offers a federation and management platform that enables service providers to harmonise interfaces and connect to a common hub. The foundation team engages with research communities to understand the demand, to simplify the access and to drive innovation together. The Federation consists of national initiatives providing compute, storage, and networking to their national scientific users, but also collaborating as part of the federation to support international scientific collaborations.

EGI provides a number of base services for the federation, many of which are used by the national infrastructures. These include the following:

- Operations coordination and support - providing a collaborative helpdesk and ticketing system to the entire federation, and support for the ongoing operation of services and facilities. This also includes services for monitoring of operations, and accounting of resources used.
- Service management coordination - implementing processes for improving service delivery, based on industry standards.
- Security coordination - providing a service to ensure common security policies are implemented across the federation, enabling the necessary level of trust; and

[3] https://eoscfuture.eu
[4] http://france-grilles.fr
[5] https://egi.eu
[6] https://www.openaire.eu/

providing an incident response coordination team and expertise drawn from the federation. This also includes the essential security policy development and evolution.
- Federation of the national and thematic identity providers, to manage users and attributes.
- A large scale cloud and storage environment, that acts as a single entry point to the federated resources opened by the national collaborating infrastructures.

Many of these services have been developed and evolved over close to 20 years through the various grid projects to a very mature state. It is anticipated that many of these will be integrated into the EOSC.

France (IN2P3) has been involved in EGI and the previous grid projects since the very beginning in 2001.

OpenAire

The mission of the OpenAire project was to encourage openness and transparency in scholarly communication, and to facilitate innovative ways to communicate and monitor research. It has done this through a number of actions, including:

- Aligning policies through national collaborators
- Producing guidelines and sharing common practices, standards, and protocols as well as sharing content and expertise to help build a sustainable open research ecosystem.
- Facilitating interoperability, and linking research to enable discovery, transparency, reproducibility.
- Training, researchers, content providers, policy makers, citizens in open science practices.

One of the key outputs of OpenAire is the Zenodo[7] service, which was designed as a catch-all repository for EC funded research, and was launched in 2013. Zenodo is an open access repository that accepts all research outputs from any field of research, with any file format. It promotes peer-reviewed research and curates the uploads that are visible. It assigns DOI to make all of the data citeable. Communities can create their own spaces within Zenodo to allow them to create their own collections which are open and citeable. It also enables the use of different licensing policies. Today Zenodo is widely recognised as a very important resource for publishing open publications and associated research data. It will be a key ingredient in the EOSC infrastructure.

EOSC Pilot

The EOSC Pilot project[8] which ran from 2017 to 2019 was set up to support the first phase of developing the EOSC. The objectives were:

- Propose and trial governance frameworks for the EOSC and contribute to the development of European open science policy and best practice;

[7] https://zenodo.org
[8] Https://eoscpilot.eu

- Develop a number of demonstrators functioning as high-profile pilots that integrate services and infrastructures to show interoperability and its benefits in a number of scientific domains; and
- Engage with a broad range of stakeholders, crossing borders and communities, to build the trust and skills required for adoption of an open approach to scientific research.

Its important outputs included a recommendation on an EOSC architecture, and a roadmap for a service portfolio. These aspects were later picked up by dedicated working groups in preparation for EOSC implementation.

EOSC-Hub

The EOSC-Hub project[9] ran from 2018 until December 2020, and was designed to build a service catalogue to simplify access to a range of products, resources, and services coming from the pan-European and international organisations, and to encourage adoption of standards for interoperability of compute, storage, data and software. It also aimed to consolidate existing e-infrastructures, expanding capacity and improving services, as well as broadening the access to researchers, educators, business and more.
The project developed the initial EOSC portal and marketplace, and a platform for industrial collaborations with EOSC - the Digital Innovation Hub. It also developed a number of aspects that are picked up in the EOSC-Future project as enablers of the EOSC structure, including:

- Rules of participation enabling service providers to integrate their services into the EOSC portal
- Service management system
- Interoperability and integration guidelines
- Catalogue of services for researchers
- Training materials

EOSC Enhance

The EOSC Enhance project[10] funded by the EC ran until end 2021 and was tasked with developing the vision for the EOSC. During the lifetime of EOSC Enhance, project partners developed and improved the functionality of the EOSC Portal, further augmenting the catalogue of services, and connecting independent, thematic data clouds for the benefit of users and service providers across Europe.

These objectives were achieved via wide-ranging stakeholder consultation and in close collaboration with the EOSC Governance. Users, service providers and existing EOSC-related projects played an active role in the co-design of the portal specifications.

With a focus on facilitating the interoperability and discoverability of services and resources across scientific disciplines, the project applied the emerging EOSC Rules of Participation, using specifications, guidelines, tools, and APIs to link service and resource providers.

The key outputs of the project were:

[9] https://eosc-hub.eu
[10] https://eosc-portal.eu/enhance

- EOSC portal catalogue of resources through merging and integrating previously existing resources
- EOSC portal training materials
- APIs to enable providers to upload resources and to retrieve content from the portal
- Developed the portal onboarding process
- Developed a first marketplace to show the contents of the catalogue and allow exploration of service offerings

Together the outcomes of these projects from the e-infrastructure consortia form part of the basis of the EOSC infrastructure that EOSC Future integrates to build a first consolidated prototype EOSC.

The Science Clusters

Of course, the EOSC is really about building a collaborative and open environment for Science. As such one of the important developments has been the actions to build and enhance the clusters of domain-based science. Some of these existed prior to the funding call, but others developed as a result of the funding opportunity. These so-called "Science Clusters" represent a huge proportion of European scientific research (some 50,000 researchers or more), and a strong representation of the scientific activities that will potentially benefit from EOSC.

In this section we briefly describe the clusters, and the coordinating activities related to them.

ENVRI-FAIR

The ENVRI-FAIR project is a collaboration of the large Environmental Sciences community, comprising 26 research infrastructures, networks and projects. Together they cover different aspects of the Earth system, of which 14 are in the current ESFRI roadmap. The infrastructures and interests are shown in the Figure below. Additional research infrastructures and other research infrastructure communities operating within the environmental landscape are engaged and utilizing the ENVRI-FAIR developments through the ENVRI community-related activities within this project.

ENVRI-FAIR will implement, further develop and test tools, methods, and services at the research infrastructure level with strong synergies across the subdomains and at the ENVRI community level. It fosters the coherent development of the leading research infrastructures towards synergies, complementarity and interoperability and supports upcoming research infrastructures according to the principles of the ENVRI community developed in the previous ENVRI and ENVRIplus projects.

An important outcome of the project is the ENVRI-hub, which will be a central gateway to environmental data and services offered by the European environmental research infrastructures. The data offered through the hub will be interoperable across the Earth system disciplines and therefore easy to use for interdisciplinary environmental research. The data will be open and free to use by anyone. Users of the ENVRI-hub will be also able

to use the Virtual Research Environments and do their science computing directly inside the hub.

Users should be able to test the first demonstrator of the hub at the end of 2022. ENVRI-hub will be part of the EOSC ecosystem, and accessible from the EOSC portal.

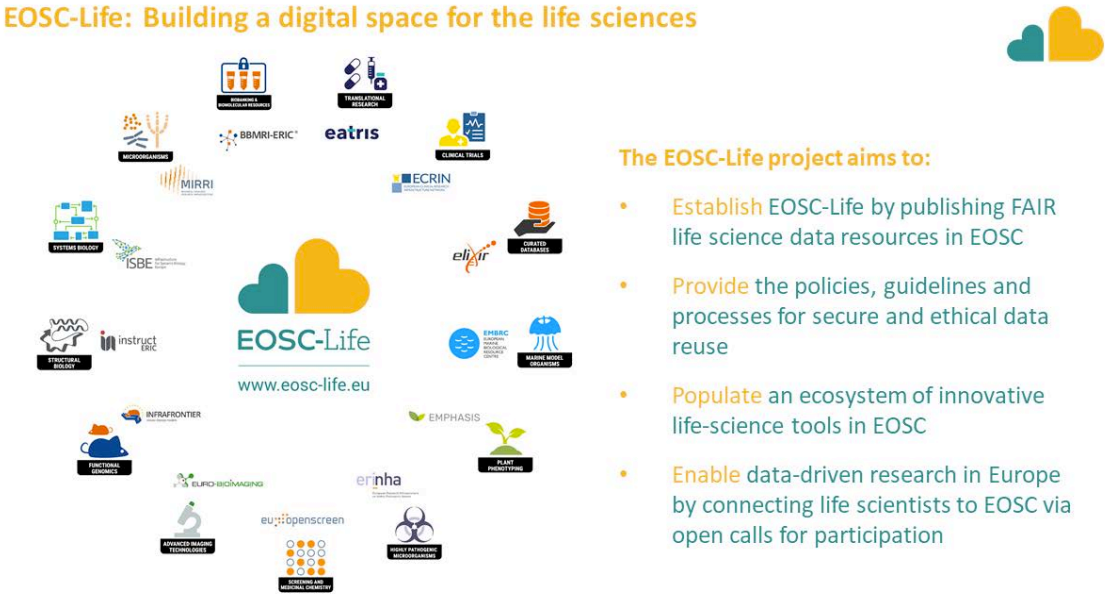
EOSC-LIFE

The EOSC-Life project[11] brings together the 13 Life Science ESFRI research infrastructures (LS RIs) to create an open, digital and collaborative space for biological and medical research. The project will publish 'FAIR' data and a catalogue of services provided by participating RIs for the management, storage and reuse of data in the EOSC.
This digital space will be accessible to European research communities. The Figure below illustrates the engaged RIs and the goals of this cluster project.

EOSC-Life will make data resources from LS RIs 'FAIR' and publish them in the EOSC following guidelines and standards. Overall, this will drive the evolution of the RI repository infrastructure for EOSC and integration of the LS RI repositories.

EOSC-Life will implement workflows that cross disciplines and address the needs of interdisciplinary science. Through open hackathons and bring-your-own-data events the project will co-create EOSC-Life with the LS user communities, providing a blueprint for how the EOSC supports wide-spread and excellent data-driven life science research.

[11] https://eosc-life.eu



EOSC-Life will address the data policies needed for human research data under GDPR. Interoperable provenance information will describe the history of the sample and data to ensure reproducibility and adherence to regulatory requirements.

EOSC-Life aims to ensure that European scientists will have access to advanced data services, technology platforms, samples and support services throughout the European Research Area, and the resulting data will be openly accessible for reuse through the European Open Science Cloud in full compliance with all ethical, regulatory and legal requirements.

The user-driven project will have open calls for user research which will allow the large user community to adopt advanced data management practices and access data integration and large-scale analysis tools in the cloud. The collaborative platform and inclusive community and capacity building will help to foster data science skills in life-science research.

Thus, by the end of this project EOSC-Life will be established as the new norm for digital biology in Europe – accessible by Europe's 500,000 life scientists.

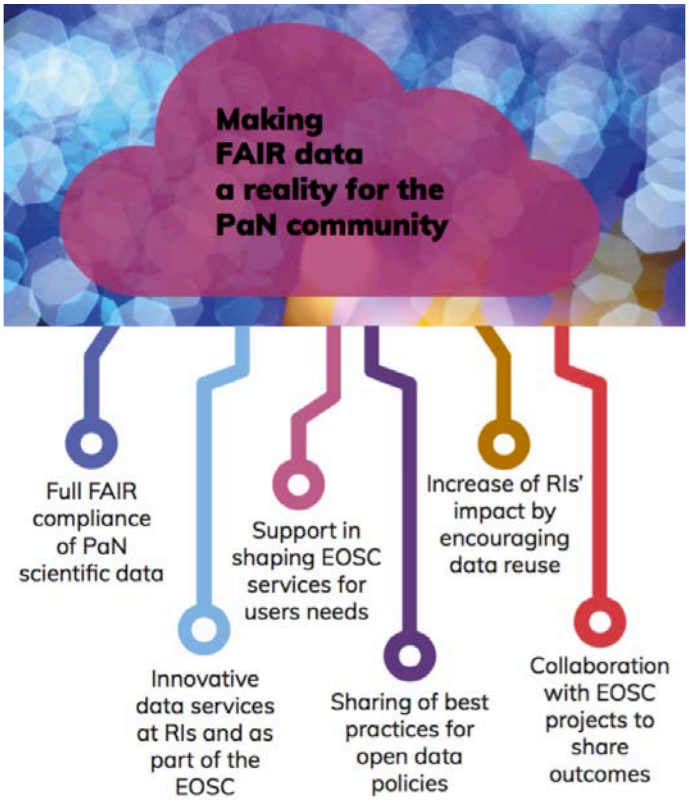
PANOSC

The Photon and Neutron Open Science Cloud (PaNOSC)[12] is a European project for making FAIR data a reality in 6 European Research Infrastructures (RIs), developing and providing services for scientific data and connecting these to the European Open Science Cloud. The objectives of the project are to:

- Participate in the construction of the EOSC by linking with the e-infrastructures and other ESFRI clusters.
- Make scientific data produced at Europe's major Photon and Neutron sources fully compatible with the FAIR principles.

[12] https://panosc.eu

- Generalise the adoption of open data policies, standard metadata and data stewardship from 15 photon and neutron RIs and physics institutes across Europe
- Provide innovative data services to the users of these facilities locally and the scientific community at large via the European Open Science Cloud (EOSC).
- Increase the impact of RIs by ensuring data from user experiments can be used beyond the initial scope.
- Share the outcomes with the national RIs who are observers in the proposal and the community at large to promote the adoption of FAIR data principles, data stewardship and the EOSC.



The project brings together the six strategic European research infrastructures (ESRF, CERIC-ERIC, ELI Delivery Consortium, the European Spallation Source, European XFEL and the Institut Laue-Langevin – ILL, and the e-infrastructures EGI and GEANT, with the goal of contributing to the construction and development of the EOSC.

The mission is to contribute to the realization of a data commons for Neutron and Photon science, providing services and tools for data storage, analysis and simulation, for the many scientists from existing and future disciplines using data from photon and neutron sources. To achieve this aim, the exchange of know-how and experiences is crucial to driving a change in culture by embracing Open Science among the targeted scientific communities. This is why the project works closely with the national photon and neutron sources in Europe to develop common policies, strategies and solutions in the area of FAIR data policy, data management and data services.

SSHOC

The Social Sciences and Humanities Open Cloud (SSHOC)[13] was a project funded by the EU framework programme Horizon 2020 and united 20 partner organisations and their 27 associates in developing the social sciences and humanities area of the EOSC. The stakeholders are shown in the Figure.

During the 40-month lifespan of the project, (from January 2019 to April 2022) SSHOC worked to transform the previous social sciences & humanities data landscape with its disciplinary silos and separate facilities into an integrated, cloud-based network of interconnected data infrastructures.

The project has delivered an SSH Open Marketplace, in the spirit of the EOSC catalogue of services, but for the Social Sciences and Humanities. The marketplace provides exploration and access to tools and services, training materials, publications, data, and workflows. It also brings together a number of existing data catalogues from the participating RIs.

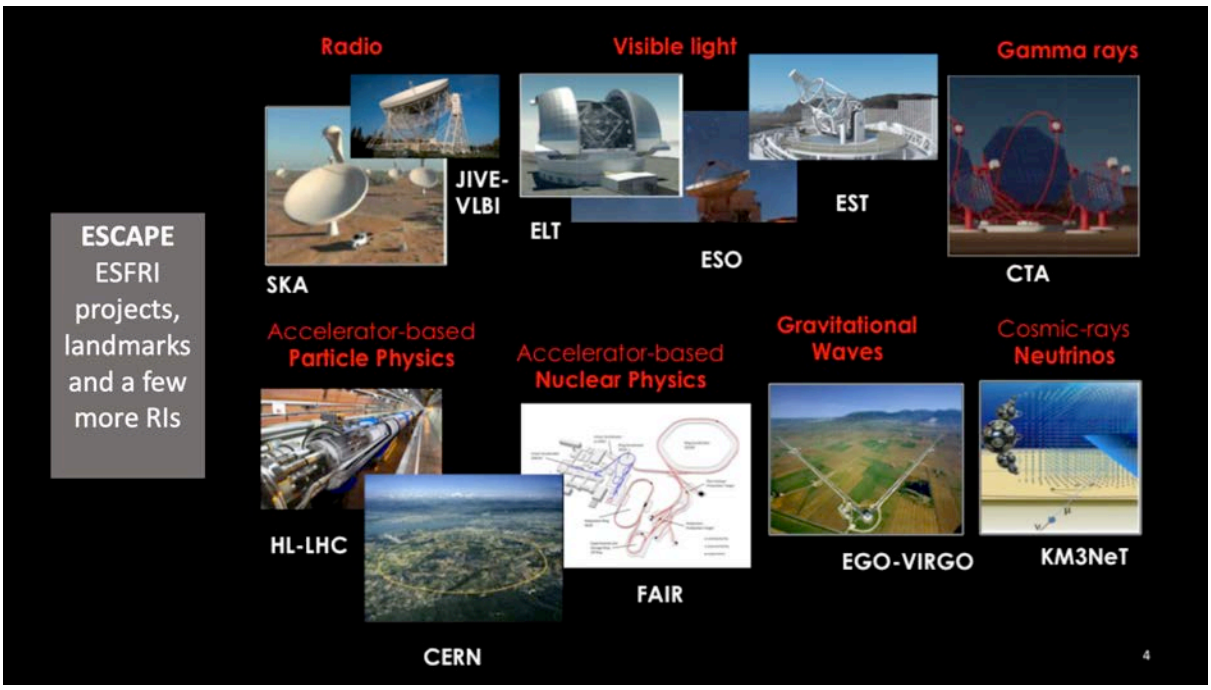
ESCAPE

The European Science Cluster of Astronomy and Particle Physics ESFRI RIs (ESCAPE)[14] project brings together the ESFRI and other world-class research infrastructures in Astronomy, Astroparticle Physics, Gravitational Wave and Particle and Nuclear Physics in a single collaborative structure to create a domain-specific “cell” of the EOSC. ESCAPE is a

[13] https://sshopencloud.eu
[14] https://projectescape.eu

cluster of ESFRI projects with aligned challenges of data-driven research, with demonstrated capabilities in addressing various stages of data workflow and concerned with fundamental research through complementary approaches. The engaged RIs are shown in the Figure.

ESCAPE aims to produce versatile solutions, with great potential for discovery, to support the implementation of EOSC thanks to open data management, a cross-border and multi-disciplinary open environment, according to FAIR principles. The ESCAPE foundations lay on the capacity building of the earlier ASTERICS[15] project work towards enabling interoperability between the facilities, minimising fragmentation, encouraging cross-fertilisation and developing joint multiwavelength/multi-messenger capabilities in astronomy, astrophysics and particle astrophysics communities.

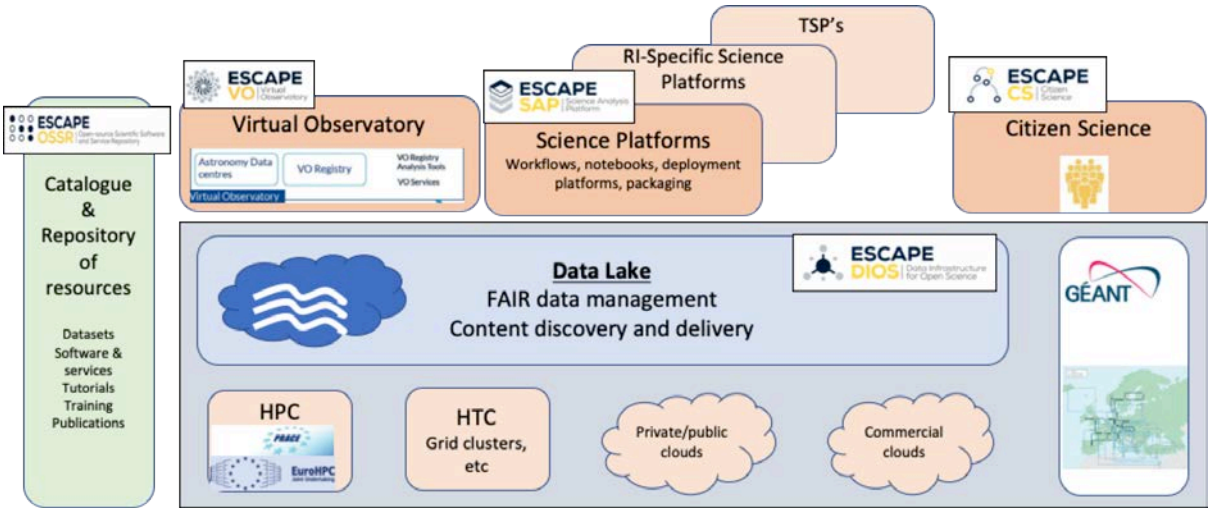


The output of ESCAPE is a domain-specific “EOSC cell” to be integrated in the global EOSC infrastructure and to help in setting up the related services. Such an EOSC cell is composed of five main components:

- ESCAPE Software Repository: the repository of scientific software services of the research infrastructures concerned by the ESCAPE.
- ESCAPE ESFRI Science Analysis Platform: a flexible science platform for the analysis of open access data.
- ESCAPE Virtual Observatory: astronomical high-level products archive and related services.
- ESCAPE Data Infrastructure for Open Science: a scalable federated data infrastructure as the basis of an open access data service for the ESFRI projects within ESCAPE and concerned by Exabyte-scale data volumes.
- ESCAPE Citizen Science: an open gateway dedicated to the public through Citizen Science and communication actions.

These are illustrated below:

[15] https://asterics2020.eu



The ESCAPE activities and relationships; user communities access services and resources through the RI-specific platforms, Virtual Observatory, Test Science Projects (TSP), and as citizen scientists.

ESCAPE provides deep training, education and capacity building programmes for the new generation of scientists and engineers that fully exploit ESFRI and EOSC facilities, to ensure the requirements and service features are properly understood and uptaken.

The main goals of the project are to:

- Improve access to data and tools to unlock innovation for society at large.
- Provide data according to FAIR principles to increase researchers' efficiency thanks to scientific data interoperability and establish new methodological approaches and rules for quality certified data and science tool sharing.
- Build a European cross-border and multi-disciplinary open innovation environment for research data, knowledge and services, while connecting EOSC and ESFRI.
- Facilitate interdisciplinary and networked research between different sciences, through research infrastructure ecosystem and by supporting data publishing, analytics, computational capacity, virtual analysis environments and workflow systems.
- Create of economies of scale, through the adoption of common approaches for data management
- Educate and train the scientific and wider user communities, to ensure the up-take of ESCAPE's results.

Evolution of the Science Clusters

The five science clusters described above have worked together closely throughout the lifetime of these cluster projects, and in particular in guiding the setting up of the EOSC-Future project, where they are all key stakeholders and participants. In fact the clusters have found the community structuring enabled by the cluster constructs themselves and acting together as a broad science community to have been extremely beneficial. In all cases the five clusters have a plan for long term sustainability of their cluster, and in addition a collaborative structure among the clusters is also foreseen and welcomed by all.

- The Life Sciences have had a strategy board in place for many years, and a collaborative MoU between the RIs since 2015. The intention is to maintain that collaborative structure following the end of the current EOSC-Life project.
- The Environmental Science Community has had in place for several years a Collaborative Framework for pan European Environment research infrastructures, together with a Steering Board of European environmental RIs. They are preparing a MoU to be signed by all RIs for a long term collaborative structure.
- The SSHOC project has ended, but they have put in place an MoU to form the SSH Open Cluster between the SSHOC ERICs and other RIs, and will maintain the marketplace and the collaboration started in the cluster project.
- For the PANOSC community they have a consortium agreement already in the context of LEAPS (League of European Accelerator-based Photon Sources), which acts as a strategy board and engages the RIs in PANOSC.
- ESCAPE has put in place a new collaboration agreement which has been signed by all of the RIs and International Organisations participating in the cluster project. The cluster will transition to the collaboration at the end of the cluster project funding. The ESCAPE name will be kept.

All of the five clusters will be open to new RIs in their domains, often using the ESFRI roadmap as a guideline for membership, but not limited to that. None of the clusters have set up legal entities, but their long term structure is based on MoU or Collaboration Agreements, as with many scientific collaborations.

There is a strong intention that the clusters will collaborate together, and act in concert towards ESFRI and the EC, in particular to guide policies and strategies for EOSC and related topics affecting all of the scientific communities in Europe.

It must also be remembered that almost all of the RIs in all of the clusters have an international and often global aspect beyond Europe. It is essential that we ensure that strong leadership and action towards open science in Europe is done in collaboration with international partners.

EOSC-Future and Related Projects

In 2021 a new project was launched with the goal of bringing together and integrating all of the results and developments in the EOSC ecosystem to date, and building a first full implementation of the EOSC itself. That project is EOSC-Future[16], and it brings together all of the significant stakeholders: the five science clusters, and the five e-infrastructures: EGI, EUDAT, OpenAire, GEANT, and the Research Data Alliance (RDA).

EOSC Future will build on the existing baseline for the European Open Science Cloud to deliver a platform with a durable set of user-friendly components that are designed for the long haul. It will adopt a system-of-systems approach to the EOSC platform, linking together other research portals, resources and services to respond to the data needs of a wide range of researchers.

[16] https://eoscfuture.eu

One way to think about EOSC is as a fully operational web of data and related services founded on FAIR protocols, principles and standards for accessing interoperable datasets. In practice, it will work with key stakeholders to ensure a smooth user experience, developing:

- EOSC core, the set of enabling services needed to operate the EOSC
- EOSC exchange, registering resources and services from research infrastructures, other EOSC projects and science clusters to the EOSC and integrating them with the EOSC core functionalities
- the EOSC interoperability framework will provide guidelines for providers that want to integrate services or data into EOSC

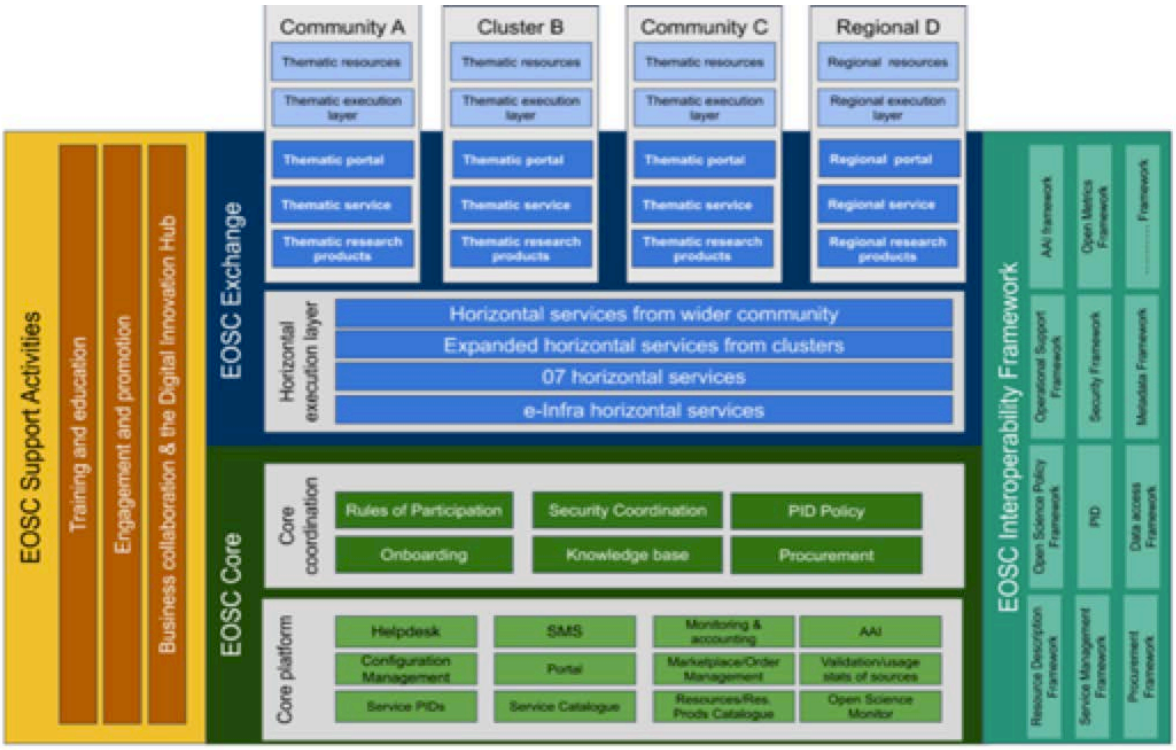
It will engage with users throughout the different development stages to make sure the EOSC matches researchers' needs and is intuitive. We'll also provide support and training to make sure users can make the most of the EOSC platform.

The main components of the strategy are the following:

1. Delivering and operating the EOSC-Core, developing and operating key components of the platform and integrating them with a growing set of services to realise EOSC-Exchange
2. Expanding EOSC-Exchange with resources from across scientific disciplines, delivering common services for scientific workflows and integrating resources from national and pan-European e-infrastructures and research infrastructures
3. Scaling up capabilities to deliver an EOSC Execution Framework, evolving the EOSC architecture and promoting composability through the Interoperability framework to deliver a user-friendly execution environment where users can combine and process data
4. Increasing European scientific impact by enabling interoperable resources for cross-domain research that can address complex scientific challenges
5. Promote the participation of the commercial sector in EOSC, facilitating encouraging SMEs and industry to deliver missing functionalities and enable commercial innovation
6. Support and train users and providers through learning materials and training to increase the uptake of EOSC resources and open science practices
7. Engage with the EOSC communities and end users to co-design and co-create the EOSC platform
8. Align with the strategic vision for EOSC by liaising with the EOSC Association and its members, EU member states and associated countries and other initiatives

An important ingredient of the project is the 10 innovative science projects that the science clusters will deploy into the prototype EOSC infrastructure. These are cross-domain, and in several cases cross-cluster, and are designed to demonstrate new capabilities that EOSC can bring, in particular to being able to publish and reuse scientific data and services.

The project has designed an architecture for the EOSC, based on outputs from many EOSC working groups that had been in place previously. This architecture is shown below.



The EOSC architecture: common services and facilities in the core; an Exchange layer of contributed services from the RIs or other projects; support structures; and an interoperability framework. The science communities and clusters are integrated, but also have thematic-specific resources and workflows.

The core, the support services, and interoperability framework will be useful to all communities. The Exchange layer is where science communities can onboard services they have developed which are potentially of broader use. In addition, the Exchange layer provides access to compute and storage resources provisioned through several related projects, specifically set up to provide resources. Additional details on the architecture can be found on the project web.

As an example of how a cluster may integrate with the EOSC architecture, we can consider the ESCAPE plan. ESCAPE will publish infrastructure software such as the Data Lake (exascale data management system), and other project developed software into the Exchange layer. This will be done through the integration of the metadata of the ESCAPE software catalogue. This is the strategy of all of the clusters. Thus the Exchange layer can federate all of the research software and services produced in the clusters. As time goes on we anticipate additional common services to be provided, where such are interesting to a broad range of communities. An obvious first such service would be a generic (Jupyter) notebook service, or common citizen science frameworks.

Items, such as the science analysis platforms, Virtual Observatory service, and RI workflows remain in the domain of the RIs, but should make use of common services from the core (such as common AAI, support services, accounting, security, etc.).

ESCAPE will publish its established standards into the interoperability framework. That framework can be used, in particular to enable the cross domain use of data sets.

Importantly, the provision of resources (compute, storage) funded through the provisioning projects will enable new cross-discipline projects to be deployed. Such resources will of course be an attraction for users to the EOSC. Long term mechanisms for procurement of resources and services are an aspect of the EOSC that EOSC-Future together with the EC must develop during the lifetime of the project. A long term goal is to open the service and resource provision and operation to procurement tenders open to industry as well as the EOSC community.

EOSC Governance, and the EOSC-Association

The governance of the EOSC is a tripartite one. The actors are the EU represented by the Commission, the EOSC Association, and the EOSC Steering Board representing the Member States.

The EOSC Association is a legal entity set up to govern the EOSC, and currently has more than 200 members and observers. The EOSC Association has a role in coordinating between the projects funded to implement EOSC, and to set up Task Forces to inform strategy on specific topics. The members of the Association are Research Funding organisations, research performing organisations, service providers, and others such as intergovernmental organisations. Thus the Association should represent the research community and its stakeholders, and in the long term should maintain a coherent strategy across specific projects and funding calls. The Association should guide the implementation of the EOSC, run advisory groups, set policies, and create an environment for networking.

Most of the institutes that make up the research infrastructures and science communities are members of the Association.

Future Developments and Outlook

The current situation has the EOSC-Future project together with its partners and stakeholders, integrating and developing the outputs of the precursor projects, and building a coherent first full implementation of the EOSC. The science communities, via the clusters, as partners in EOSC-Future are starting to deploy their Science Projects within that environment with the goal of demonstrating novel capabilities in open science.

Many aspects of how the EOSC will develop and evolve are still to be understood. In particular it will be important to understand how the sustainability of the EOSC in terms of funding will be managed. It is being implemented through project funding in Horizon 2020 and Horizon Europe framework programmes. However, that does not guarantee sustainability, and ultimately that will only come if the EOSC demonstrates to the science communities that it is really beneficial and attractive to users. Thus it is essential that the scientific communities drive the direction and implementation. Without responding to needs and adequate take-up the long term is not certain.

The potential strength of EOSC is its collaborative aspect, and the provision of a common set of core services required by all. This can be a huge cost benefit. In addition, the structuring of the scientific community and sharing of experience, expertise, as well as

research products and software can be very important. The Exchange layer can capture that. Common strategies and policies for open and FAIR data, open access, and open science is a huge benefit to the science communities and potential data consumers.

Risks

There are 2 main threats to the success of the EOSC. The first is a lack of take up by the user communities. This may be because the EOSC does not respond to user needs, or does not appear to provide sufficient benefit. The mitigation of that is to ensure that the science communities, primarily the clusters today, but later a broader range of stakeholders via the EOSC Association, can steer the development of the technical implementation. Also it must be recognised that scientific computing strategies of the sciences is a rapidly evolving environment. There is no such thing as being able to specify “requirements” for extended periods. The requirement is to be able to evolve as the experiments evolve and needs change. If the infrastructure cannot adapt rapidly there is a significant risk of communities diverging from common services. It is important that the common layer be carefully developed with this in mind. It is also important to ensure that there are no barriers to being able to effectively use the infrastructure.

The second main threat is a lack of a clear long-term funding model. Today the individual RIs are funded to do science by their national funding agencies. Additional funding may be available in national e-infrastructures. That funding is usually subject to a rolling planning with a several year outlook. If the communities are to commit to EOSC, it is going to be important to understand how the funding model will evolve, how the infrastructure will be supported in a sustainable way, with adequate resources for continual evolution, and how physical resources will be funded and made available. While project funding is useful to specific developments or changes, it is a very costly and time consuming model for a long term infrastructure. A funding model that integrates and addresses all of these aspects is going to be essential, for both the operators and the science users.

In conclusion, the outlook is very positive, as there is a clear desire from both the funding agencies and the science communities to be open and collaborative. The advantages of that from many aspects are well recognised. That desire must be balanced by the need of the science communities to produce world-class science - the infrastructure must support that, and enable new activities.