EOSC Future: Design and implementation of community engagement through Science Projects

Christos Arvanitidis, Ron Dekker, Andreas Petzold, Niklas Blomberg, Giovanni Lamanna, Rudolf Dimper, Cristina Isabel Huertas Olivares, Ana Mellado, Matthew Viljoen, Sally Chambers, Montserrat González, Sophie Viscido

Introduction

The European Union has established the European Science Open Cloud (EOSC), as the "system of systems" Infrastructure where scientists, stakeholders and other types of users can find data, analytical and other web services, publications, networks and any type or research products and outcomes to carry out their work. Multiple projects have been funded towards this vision, such as EOSC-hub, OpenAIRE-Advance, EOSC-Enhance, Science Cluster projects, and the INFRAEOSC-07 projects.

The vision of EOSC Future project is to provide an operational Platform for EOSC as an integrated environment, for the discovery, access and execution of data, professionally provided services, and open research products, to be available for use by the European researchers. This environment will provide additional functions so that the researchers and other type of users to be engaged, facilitated, trained and supported to use the EOSC resources and solutions and discover the opportunities it opens for their communities. It is expected that in the long term, this environment will free the innovation potential of the interdisciplinary and cross-domain research, especially in the new generations of the scientists, and will finally create a turning point for the modern research practices and knowledge making, which in turn, can transform the European and global scientific communities into new knowledge and innovation hubs.

The EOSC Future project has embarked towards three main directions: (a) the development of EOSC-Core and EOSC-Exchange components with interoperable data and resources; (b) the integration of data, services and other resources from the Science Cluster communities; (c) the direct involvement of users in the development of this integrated environment through the co-design, co-development and implementation of the EOSC Platform. EOSC Future is mandated to create a web of research resources that are open and/or Findable, Accessible, Interoperable, and Reusable (FAIR), to the greatest possible extend, and equipped with respective sets of services for their exploitation. The project will foster the Open Science vision in Europe, as fuelled by open data and open-source principles, by delivering a paradigm shift in the way scientists can create, share, and reuse open and FAIR data, research services, and other research products. In addition, it will extend the current EOSC model, aiming at integrating and making available the vast number of research resources available for the scientific community at large. Researchers from Europe and worldwide will be able to share, access, and analyse open and FAIR data, services, and research products or in general resources from various scientific domains, providing them with the opportunity to reproduce analytical processes and results or combine data and services for new, potentially cross-disciplinary research activities. To achieve this, EOSC Future shall aggregate and compose services provided by Research Infrastructures, e-Infrastructures, Science

Clusters and Research Organisations by leveraging, enhancing, expanding, integrating, and optimising, where necessary and possible, the outputs of past and current EOSC projects (including EOSC-hub, OpenAIRE-Advance, EOSC-Enhance, Science Cluster projects, and the INFRAEOSC-07 projects).

Concept

The ESFRI Research Infrastructures are organised in five thematic Science Clusters within the EOSC ecosystem, corresponding to the main scientific domains: ENVRI FAIR on the Environment, EOSC Life for the Life and Biomedical Sciences, PANOSC for the Photon and Neutron, SSHOC for the Social Sciences and Humanities and ESCAPE for the Astronomy & Particle Physics. These Clusters are critical to the development of the EOSC integrated environment for the provision of data, services and other research products whose integration, composition and use by their communities will undoubtedly contribute to the emerging EOSC infrastructure. One of the EOSC platform main challenges is how the Science Clusters and other stakeholders can facilitate and enhance the participation of scientists in the co-development, implementation and exploitation of EOSC and to accelerate their uptake of Open Science through EOSC.

The Science Clusters in EOSC Future pave the way on how shared projects can contribute to addressing major societal challenges for Europe and also a sound example of how Research Infrastructures within EOSC can align to support Horizon Europe's missions and facilitate Europe to take a leading role in excellent science making in modern times. They serve as advocates for the advanced capabilities of EOSC integrated environment by combining the IT technical power of the e-Infrastructures with the hypothesis-driven thinking of the Science Cluster communities. At the same time, this approach attempts to break down the multiple barriers: (a) linguistic and cultural barriers between the scientific and engineering (IT technical) communities, as well as between the scientific communities of the different Clusters (cross-domain) and even between the communities of the same Cluster (interdisciplinary); (b) current technical barriers in having access to computational power, graph systems and other resources of technical nature in order to be able to work with big data, execute models of global relevance or connect information from multiple resources; (c) current scientific barriers, which emerge from the fact that the communities between the Clusters and even within the same Cluster often work in isolation on the same topic.

For this reason, the approach adopted by the Science Clusters has evolved into interdisciplinary and cross-domain Science Projects, the interrelations of which can be seen in the following Figure.
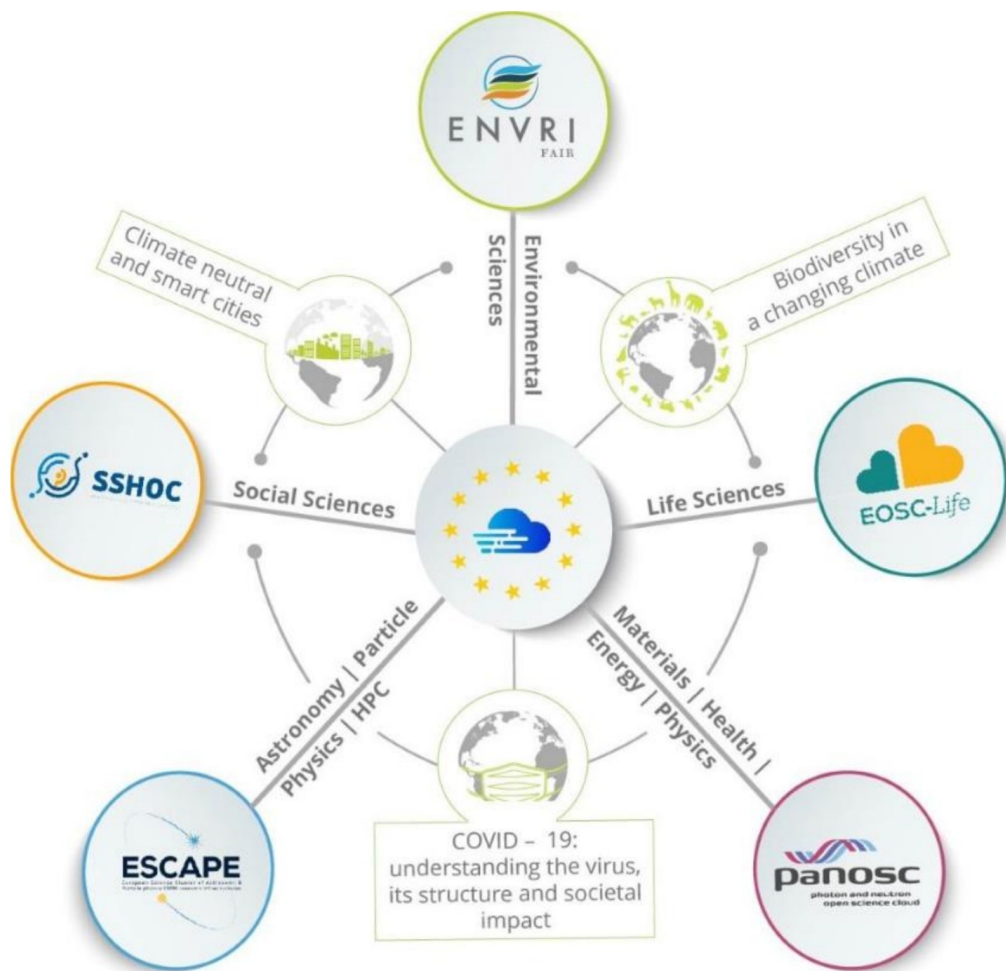
Figure 1. Cross-cluster Approach of Science Cluster Projects under EOSC Future (courtesy of Andreas Petzold and collaborators).

Accordingly, the ten Science Projects adopted to run under the EOSC Future project address the following scientific challenges:

Climate change and its impact on the Earth system and society:
1. Assessing the *impact of climate change on biodiversity and ecosystems in Europe* by creating interdisciplinary access and analytical tools for big datasets from genomics to *in situ* and satellite borne environmental data and connecting the analytical framework and federated access to relevant data infrastructures via the EOSC Portal.
2. *Dashboard on the state of the environment* to provide science-based information on environmental boundary conditions for informing society stakeholders on short-term to long-term developments

therein by means of a dashboard of environmental indicators with integrated data and analytical workflows.

3. Addressing the EU research and innovation missions '*Climate Neutral and Smart Cities'* by combining resources from two clusters encompassing Citizens and Sustainable Environment.

COVID-19 Pandemic - health, societies, and environment:

4. *COVID-19 metadata findability and interoperability in EOSC* by combining Social Science and Humanities (SSHOC) and Life Sciences (EOSC-Life) to address how to make the metadata of data objects managed by these Research Infrastructures interoperable.

5. *Imaging COVID-19 data in EOSC - COVID-19* as a demonstrator to reduce the number of deaths and illnesses related to COVID-19 by accelerating research with findable and interoperable COVID-19 metadata, and with COVID-19 data access in the cloud with streaming-based access to image data.

6. *Tracing biostructures* through data-intensive processing and AI-supported image classification.

7. Investigating the *dynamics of biological processes,* in particular the entire viral life cycles or assembly of macro-molecular complexes (COVID-19) by neutrons and X-rays, based on a high-efficiency AI machinery for rapid interpretation of the results.

Science questions meet with e-Infrastructures:

8. Increasing our *understanding of dark matter* as a major but still widely unknown component of our universe by performing new analyses within the involved ESFRI Research Infrastructures and collecting all the digital objects related to those analyses (data, metadata, and software) on a broad platform that will be ultimately hosted on the EOSC Portal.

9. Increasing our *understanding of the extreme universe and gravitational waves* by building the transversal environment to provide frontier AI, analysis methods and a cloud-based analysis dashboard that allows users to exploit the e-Infrastructure services.

Upgrade Cluster services to other domains :

10. *Access Management for distributed Research Infrastructures* by facilitating access to and exploitation of instruments and privacy-sensitive data via the expansion of an existing tool within Life Sciences to other domains, starting with the Social Sciences and Humanities domain.

The portfolio of resources contributed by the ten Science Projects includes data, software and services in the following broad categories:

- Data
  - Biological and environmental data from Research Infrastructure repositories, including data from global aggregators such as GBIF and OBIS;
  - Data for marine physics, chemistry, bathymetry, geophysics, and biology originating from more than 700 organisations, from national marine monitoring and scientific research programmes and cruises, concerning the European seas and global oceans;
  - Data from European Social Science Data Archives, including thematic archives on COVID-19, Circular Economy, Cancer, Climate Change, Economic Crises, and Infectious Diseases;
- Services, including cataloguing
  - Data repositories of all contributing Research Infrastructures;
  - Metadata catalogues;
  - Data products;
  - Data vocabulary services;
  - European Languages Social Sciences Thesaurus;
  - Data directory services;
  - Software catalogues;
- Workflows & processes
  - Pipelines for data chaining originating from environmental monitoring structures;
  - Data discovery through data APIs, including metadata explorer services;
  - Data reduction, analysis and visualisation services for research data including Jupyter Notebooks and software and advanced services for image processing;
  - Cloud orchestration services and workflow management services;
  - Data management services for data transfer, content delivery to arbitrary compute services, policy driven data storage and replication storage designed to support and manage arbitrary size data sets up to many Exabytes;
  - Facility management services.

Design and Implementation

The Science Projects in EOSC Future address the need to widen the user base and to foster increased use of the EOSC Platform by research communities, either by composing services out of existing scientific content and e-Infrastructure services or by bringing in additional services they use for their

research into the EOSC integrated environment. As a prerequisite, the services need first to be properly integrated in the EOSC platform. The scope is to move beyond the topic-specific competence centres and to consider cross-domain, interdisciplinary Science Projects. This approach facilitates the growth of EOSC by involving, connecting, and linking with more research communities and scientific disciplines, preventing the creation of silos, de-fragmenting communities and supporting Open Science.

The most demanding challenge needed to met by the end of EOSC Future project is the cultural change of the current solitary working pattern of the scientific communities towards multidisciplinary and cross-domain research practice in order to address global societal demands (e.g. impacts from Climate Change, Covid-19, etc.).

Accordingly, large, cross-domain and mature Science Projects have been designed and implemented by the Science Clusters communities are in the process of being integrated into the EOSC environment. As such, they are flagships of community engagement, and they tackle critical scientific and societal challenges. From the EOSC Future project perspective, the activities carried out include:

(a) Preparation of Large-sized Science Projects: right from the start of the project, scientific and technical coordinators from the five Science Clusters, in collaboration with the e-Infrastructures, initiated the dialogue to ensure that the current EOSC ecosystem of services is adequate and enhanced with the ones these Projects bring in. The criteria for the successful implementation of the Science Projects include, among others: (i) Eligibility - the SPs are open to participation by research communities (RPOs: universities and research institutes) and not-for-profit providers from outside the consortium. The selection of a catalogue of services will imply participation in publication and promotion of the SP results. (ii) Contribution to EOSC: cross-domain and composability features, contributing to EOSC; the feasibility of integrating the proposed services into the EOSC ecosystem; the type of services these proposals ask from EOSC (e.g. security, monitoring, AAI); (iii) Quality of research process: the quality of the proposed research, from the hypotheses to be tested all the way to the interpretation of the results; the quality of the proposed incoming services, based on elements like standardisation, protocols, best practices, efficient use of resources, semantics, data access (FAIR- compliance); the anticipated added value of the proposed services and their workflows, beyond those making the original request; (iv) Relevance: there will be a preference for service catalogues that address the EU focus areas for research and societal demands, including infectious diseases, the European Green Deal, circular economy and EC Missions like climate change, health,

agriculture; the user communities they bring and attract; gender issues; potential for Open Science and open innovation.

(b) Monitoring and Evaluation of SPs. This activity is being applied by means of specific achievements, summarised in the form of Milestones and Deliverables by each SP. Periodical reviews are performed by the community and by the Management Board of the EOSC Future Project by means of pre-defined Key Performance Indicators (KPIs). This is planned to continue after the project as the Science Clusters will uptake the results of the Science Projects.

(c) Publication and promotion of SP results. A first round of publications of the SP outcomes, that will continue after the EOSC Future project.

The Special Collection of EOSC Future Science Projects papers

The availability, applicability and exposure of the data, services and other research products, developed by the Science Projects (SPs) need to be actively communicated to the scientific and academic societies, as well as other types of potential stakeholders. Accordingly, a series of peer-reviewed publications have been combined together and are presented in this first EOSC Future Special Collection. This Collection includes the SPs Templates, which summarise the design, planning and implementation activities to be carried out in each of them. The Templates are well structured in 6 sections: (a) general information; (b) description of the work; (c) implementation plan; (d) the likely impact; (e) engagement plan; (f) a list of services to be made available from EOSC.