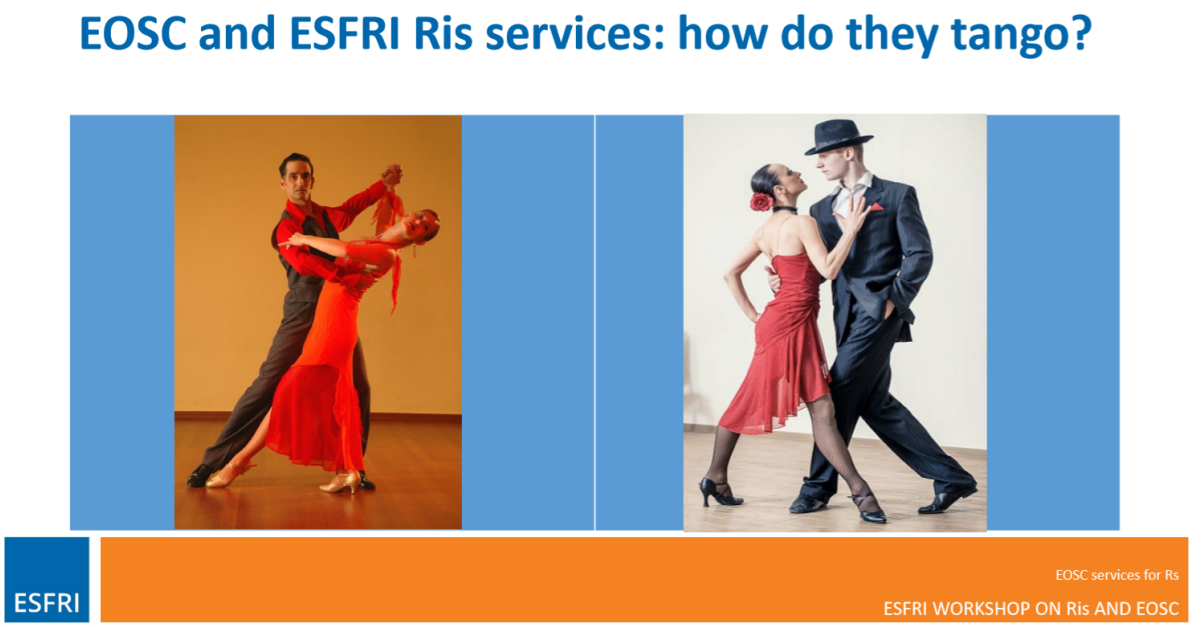
ESFRI WORKSHOP ON RIs AND EOSC
How do they tango?

30 January 2019, London

Summary note

Fotis Karayannis, StR-ESFRI



Objective

The main objective of the workshop was the liaison among ESFRI, ESFRI RIs and EOSC stakeholders, in order to align activities and ensure the optimal integration of ESFRI RIs in developing the EOSC ecosystem. This covered, among others, sharing existing good practices and experiences in Open Science policies, data sharing and reuse, FAIR principles adoption and strategies for federating RI services. The objective was broadly met with more than 150 participants joining the workshop from all the above stakeholders, including all major projects, initiatives and governance boards of horizontal and vertical (cluster) EOSC projects, ESFRI RIs and ESFRI itself, all with high-level representatives.

EOSC updates (governance, horizontal-vertical projects, European-national)

The EOSC has been progressing following a series of agreed policies and strategies and the Strategic Implementation Roadmap and already the first elements of the EOSC ecosystem have been launched in November 2018, including the EOSC portal. The EOSC portal should be understood as a one-stop-shop framework where all services and data can be found and accessed; this does not mean however a single point of access, rather an ecosystem with multiple entry points, European, national or thematic. In fact, although the portal started with onboarding some first horizontal services, further vertical (thematic) services for research infrastructures are planned (facilitated by the EOSC cluster projects), along with national EOSC ones. A call for proposals for national/regional EOSC initiatives is open and further projects

are expected in this area. In addition, the EOSC Governance has been agreed, with an Executive Board and a Governance Board from MS and the EC. All these developments have been presented in the EOSC status update session (Session 1) along with ESFRI position. In particular, the ESFRI Chair underlined the need for EOSC to adopt a participatory and subsidiarity principle for all RIs getting into the EOSC "Commons". And in the first part of the second session, the major horizontal projects showcased their achievements and future plans, including EOSC-Hub (as the main EOSC horizontal project integrating already a series of other sub-projects and services), eInfraCentral (responsible for the EOSC catalogue of services), FREYA (on persistent identifiers for data) and EOSCSecretariat.eu (supporting the EOSC governing boards).

ESFRI RIs and new EOSC cluster projects

In parallel, the ESFRI RIs have been or will be producing research datasets which need to be made FAIR + Reproducible. With the EOSC Cluster projects, electronic needs (e-needs) at broader scientific domains represented in the ESFRI Strategic Working Groups are aggregated into groups, and this makes it easier to get their services and data integrated into the EOSC ecosystem. All these efforts and their plans towards EOSC, making their data FAIR and working on Data Management Plans have been presented in the different ESFRI strategic domains, namely ENVRI-FAIR (Environment) EOSC-LIFE (Health and Food), ESCAPE (Astronomy and Particle Physics) and PANOSC (on Photon and Neutron Sources under Physical Sciences and Engineering), and SSHOC (Social and Cultural Innovation). Research Infrastructures, Projects or Landmarks possibly established as ERICs or International Organisations are involved in the clusters. Cluster projects will last for four years. Regarding their common e- or other related needs, the main areas identified in the corresponding session are research data management and related Data Management Plans (DMPs) and policies, making data and services FAIR, along with related tools and services available in the catalogue of services, preservation and sustainability of data and services and finally training requirements, mainly around research data management and EOSC services.

EOSC and ESFRI RIs/clusters: how do they tango?

A first panel discussion revealed some hot topics on how to best coordinate and harmonise EOSC and ESFRI RIs/cluster projects. Prioritisation of requirements for ESFRI RI projects was one such topic. Answers included that a very basic requirement is capturing metadata for all data and services. And working on making them FAIR, possibly prioritizing among the FAIR areas and developing related tools. The new FAIRisFAIR project will be also dealing with this area of FAIR data and prioritisation. Another requirement is training on data management and the FAIR aspects, not only for the RI managers, but also for the RI users who create the data. On the other hand, some RI communities are more advanced and have already operational services, which can be easily integrated within the EOSC framework. So, such low hanging fruits should be exploited. In addition, good practices inside and across clusters should be communicated, so that they can be worked out and adopted by others. A final point was that

sustainability needs to be kept in mind from the very beginning, while designing and implementing the EOSC ecosystem, to avoid the "near death experience" that many EU projects face after building something useful.

Key points from the break-out sessions

The breakout sessions dived deeper into some of the key EOSC areas, namely:

- Services: good practices and sustainable business models
- Policies: Open Science, FAIR data, Reproducibility
- Architecture: technologies, platforms, interfaces
- Skills and training

Services

- EOSC does not start in a green field, as there are several services already available from vertical or horizontal projects.
- More effort is required in the area of data ownership in the EOSC portal/marketplaces.
- Further support and encouragement for the usage of DOIs and access identifiers.
- KPIs are needed to track performance and assess the services, but they have also to be usable.
- Data analysis is quite a complex issue; sometimes a bottleneck; data should be curated immediately after creation. In addition, in order to achieve reliable and reproducible analysis, reproducibility is essential also for the related tools.
- The cost of data curation needs to be understood; what is the operational level of this cost and how to fund it.
- Besides thematic approaches, in many cases there are national structures offering services, which need to be better understood and be coordinated and finally integrated with the European and thematic ones.

Policies:

- Open science and data FAIRness come with a burden and fatigue for the community and this is a challenge.
- The cost of NOT having Open Science/FAIR data may be much higher than the cost of having FAIR data.
- Policies alignment is required both within and across Clusters, starting with some core policies. Good practices and recommendations, along with related training and certification (of repositories and services) are again vital.
- The use of DOIs is still limited and needs to be encouraged (in citing data).
- Metrics/KPIs for FAIR data and related impact assessment is needed.
- Dealing with sensitive data is also an issue that would need to be tackled in EOSC.

Architecture:

- On Authentication and Authorisation Infrastructures (AAI):
  - The aim is to have a common/generic AAI infrastructure model for all communities; Communities with an existing community AAI can connect to the EOSC-hub-e-Infrastructure project proxies and gain access to generic e-Infra services; Communities that don't operate their own AAI service can make use of either dedicated or multi-tenant deployments of AAI services operated by e-Infrastructures.
  - What about communities, which want free searches without any credentials? Not an issue. Some identity management and AAI capabilities still required to onboard and upload the datasets.
- On Interoperability across domains
  - There are still many different data catalogues with different APIs and user interfaces. A possible solution (as per the PANOSC paradigm) is to implement a common API searchable across sites.
  - Data interoperability problems need to be first handled at the discipline level (narrower level).
  - Data sharing is key to making Open Science work
    - Related schemas, ontologies, vocabularies and tools always help; and mapping services. Such work may be tedious for scientists and can be taken up by IT people; but such a career profile may be missing.
  - Open standards contribute towards sustainability.
  - RIs need help adopting right solutions and adapting them to domains; better communication is needed across domains, esp. on good practices.
  - Sharing Open Data should be one of the major services of EOSC.

Skills and training:

- Newcomers, including new RIs and researchers, need to be trained in data FAIRness; This is relevant for both the thematic and the national level.
- Many EOSC projects already provide training; the topic is very inclusive.
- Coordination is needed in training across the different levels to optimise resources, as the cost of skills development and training is also relevant.
- Some good practices include: the introduction of training and skills in PhD programmes; train the trainers; recognition and rewarding is needed.
- Several RIs and clusters are working on actions around training.

Follow-up discussion – Conclusions

In the discussion that followed a main point raised was that the discussion between ESFRI RIs and EOSC should be an evolving process; the workshop was only the start and more structured work, possibly via working groups or targeted meetings is needed. On a point on the EOSC

portal and whether it should have a single point of access, it was made clear that it is an ecosystem rather a single portal with a single point of access. There may be different views of the portal with multiple entry points, European, thematic or national/regional. The idea of evolving the catalogue of services into a real marketplace where resources/services can be consumed was also proposed for future work (connecting demand with supply), still requiring substantial effort. The role of the national EOSC building blocks was also highlighted as important and that each country needs to get organised for EOSC.

Workshop conclusions

- The workshop was very useful bringing together the horizontal projects-e-infrastructures with domain (thematic) ones and clusters
- The feedback received in the workshop and through the on-line tools was very useful
- There were already some good ideas on how the EOSC vision and principles can be translated into daily messages and actions, and what can be done in the immediate future
- Sustainability is important but can't be solved immediately. EOSC is not owned by the EC. It is the responsibility of the MS with the support from the EC. ESFRI has a lot to contribute and a lot of experience; information and experiences need to be shared with the EOSC Governance Board. And ESFRI can contribute a lot around these, especially in the Governance and Business Plan. The EOSC Executive Board should interact with ESFRI and explore synergies in helping EOSC implement its roadmap and deliver on the timeline.
- Next steps beyond this workshop: identify more detailed areas for further targeted meetings and workshops. To work on a timeline and related events, also by the EOSC Secretariat and StR-ESFRI. These workshops can meet some objectives and timelines. An invitation to the related stakeholders and further discussion is needed for identifying the follow-up topics. This was the first workshop of this kind but should not be the last. We all need to engage in this cooperation.
- Final statement from the workshop: You can be richer by sharing!