

MAKING THE EUROPEAN OPEN SCIENCE CLOUD WORK: WHERE TO GO FROM HERE?

Reflections from the 2021 EOSC Symposium panel session

Sarah Jones, GÉANT, sarah.jones@geant.org, <https://orcid.org/0000-0002-5094-7126>

Rudolf Dimper, ESRF, dimper@esrf.fr, <https://orcid.org/0000-0002-7659-0882>

Ingrid Dillo, DANS, ingrid.dillo@dans.knaw.nl, <https://orcid.org/0000-0001-5654-2392>

Hilary Hanahoe, RDA, hilary.hanahoe@rda-foundation.org, <https://orcid.org/0000-0002-0328-3419>

Shalini Kurapati, Clearbox AI & Politecnico di Torino, shalini.kurapati@polito.it, <https://orcid.org/0000-0003-1492-4100>

INTRODUCTION

EOSC (The European Open Science Cloud) is a complicated beast. The vision to build an ecosystem of interoperable data and services, enabling cross-border, interdisciplinary and open research is beautiful in its simplicity. The reality of implementing this vision however is incredibly complex. EOSC is a multi-stakeholder initiative, drawing together European Member States and Associated Countries, public and private providers, and all research communities. Naturally there are differences in requirements, levels of maturity, available resources and competing objectives and ambitions. Reaching consensus and attempting to serve a significant proportion of the needs of the majority is complicated, time-consuming and demands great commitment, goodwill, collegiality and trust.

That said, there are many successful models to build on and a great deal of expertise in the community. When we look at *the Riding the Wave*¹ report which set objectives back in 2010, we see that much progress has been made. The report presented a series of visions and recommendations for 2030. While there is much yet to be done, some comfort can be taken from the fact that the vision of “All stakeholders, from scientists to national authorities to the general public, are aware of the critical importance of conserving and sharing reliable data produced during the scientific process” has significantly progressed. Another example is “EU and national agencies mandate

that data management plans be created” as part of the Funding recommendation.

EOSC has evolved from a concept to an international framework for FAIR data and the value and importance of preserving and re-using data for Open Science is now generally understood. Indeed, several research infrastructures and national initiatives are well underway in implementing FAIR and Open Science principles at a community or country level. This year France published its second National Plan for Open Science 2021-2024, which provides a coherent, multifaceted and dynamic policy. With this second plan France continues a process of sustainable transformation in order to ensure that Open Science becomes a common and shared practice, encouraged by the whole international ecosystem of higher education, research and innovation.² Among the five Science Clusters of world-class research infrastructures, the Photon and Neutron Open Science Cluster (PaNOSC) regroups analytical facilities used by extremely diverse and large user communities. These research facilities are working together to implement open data policies which will ensure that thousands of data sets will become openly accessible every year for re-use. This represents many PBs of invaluable FAIR data accessible through the EOSC.

The EOSC initiative will implement a federated ‘system of systems’ architecture that brings these domain and national initiatives together in a European-wide infrastructure. Doing so will help to

¹https://ec.europa.eu/eurostat/cros/content/riding-wave_en

²<https://www.ouvirlascience.fr/second-national-plan-for-open-science/>

address outstanding issues when responsibilities currently fall between the cracks. For example when researchers generate terabytes of data in research facilities and then lack access to the necessary compute power and tools to process it in their research institution. The EOSC ecosystem should be an interface between Research Infrastructures (RIs), High Performance Computing (HPC), a wide variety of tools and service providers, and the institutions where researchers are based and conduct their work. This will allow data to be easily stored, transferred, combined, processed, published, preserved and shared, and allow researchers to find the necessary resources to do so, independently of where they come from, delivering on the return on investment for public funds.



DATA SHARING

With the establishment of the EOSC Association in 2020 we are now well placed to offer an effective stakeholder forum and governance structure to implement an ecosystem that works. The EOSC Association will act on a political level with the European Commission and Member States, as well as representing and coordinating between the wide range of stakeholders important in making EOSC a success. Now that we have these elements in place, where do we go from here? What are the main challenges we need to overcome in order to bring this beautiful but complicated beast really alive.

This paper documents a panel discussion held at the EOSC Symposium in 2021 where a number of representatives of different communities shared their honest reflections, concerns and hopes for EOSC. It is an opinion piece that is intended to generate further discussion and community inputs to help us achieve the EOSC vision. During the event the panellists were asked three questions by Sarah Jones, the Chair:

1. *What are the top 3 challenges facing EOSC in your perspective?*
2. *What are the key priority actions you think we should be focusing on?*
3. *If you had a magic wand and could change one thing to make EOSC better, what would that be?*

Responses to the first two have been combined into a series of thematic challenges in implementing EOSC and recommendations on how we should move forward. The responses to the magic wand question are given per panellist, before reaching some shared conclusions.

Challenges in implementing EOSC

The challenges in EOSC are manifold: we need ease of access to data and services so researchers can conduct their work, and interoperability across these to enable cross-disciplinary reuse. The support aspects cannot be overlooked as this human connectivity is essential to supporting the uptake of technology. Issues of trust and long-term sustainability are also critical for EOSC to flourish.

» Access to data

Making data FAIR is a long journey which has started for many but needs continuous incentive to be pursued and for others to join. Without FAIR data there is no EOSC, and we have to ensure that everybody understands the importance of managing, preserving and curating the scientific data we are producing. The currently tiny fraction of FAIR research data has to become the majority of data produced in Europe. It is important that all EOSC stakeholders do not underestimate the effort still required to achieve a higher degree of FAIRness. We need an EOSC data search engine, comprehensive and simple enough as Google. In order to achieve this we need to federate all the

individual data catalogues under a common search API and aggregate the information from search requests into something meaningful. Naturally that means in-depth community work on mapping disciplinary metadata standards and ensuring semantic interoperability so we can integrate content across repositories. There is also a need for coordination and agreement on broader aspects of data structure and provenance to enable reuse, not just discovery.

» Access to services

The desktop PC is rarely sufficient for analysis any longer and will often act as a display rather than a computer for data processing. Flexible access to physical IT resources such as storage and compute is vital in many cases and has to cover a wide range of use cases, many of them requiring little resources, others a lot. Providing seamless access to services is complicated by differences in usage rights and challenges operating services across borders and the on-going COVID-19 pandemic has shown how fundamentally important this is for working collectively. Many providers may be funded to support a certain community of users, and have to charge for those outside of given discipline, institutional or national boundaries. There are often VAT implications when operating services outside the host country, and certain providers may not wish to offer services beyond their target group. Developing business and service models that allow providers to offer services within the EOSC ecosystem, to a restricted subset of users if required, and enabling charging where necessary is very complicated and still to be resolved. Moreover, it needs to be clear from the user's credentials to which services they automatically have access. Researchers are very mobile, frequently moving from one institution to another. Affiliation needs to be managed in the approach to Authentication Authorisation and Identity (AAI) otherwise we won't be able to manage access to data and services efficiently.

» Interoperability

Interoperability across data and services will allow us to gather new insights and open up new territories for scientific discovery by combining data showing correlations which are as of today impossible to explore. Cross-disciplinary Open Science can be seen as the ultimate goal of the EOSC. This requires the use of formal standards, protocols and APIs. It should be possible for researchers to combine datasets from different disciplines and to compose a pipeline of services for processing and analysing data without major

issues in transferring the data between tools. Moreover, institutions and research groups should not be locked into certain tools. Use of open standards and APIs will allow data to be transferred from one tool to another. Within EOSC, an interoperability framework is being defined. This will specify a blueprint for data and service providers to meet which ensures their resources can integrate within the EOSC ecosystem.

As an example, this interoperability framework will specify a series of profiles to help connect multiple approaches to AAI. The aim is to have interoperable AAls, not a one-size fits all, since different research infrastructures and services use different profiles and users may have preferences over which account and sign-in method is used. Given that users also have existing pipelines and tools which they use, EOSC needs to enhance this and plug into the back of what is already used. There should be a suite of small, interoperable tools that are useful separately as well as combined. If we adopt a micro-services approach, we need each provider to be clear about what standards they support so that composition of services becomes realistic. Focusing on interoperability and making this a core part of the Rules of Participation is therefore key. However, we also realise that interoperability is the most complex challenge within EOSC and it should be positioned as a long-term goal to make sure that we are able to profit from the low-hanging fruits first.

» Trust and sustainability

While it is important to start small with EOSC and make one component or service aspect work at a time, moving beyond proof of concept to sustainable, production level services is critical to ensure researchers trust that the tool will stay around and they can rely on it. The data within EOSC also needs to be of a high quality to be reliable and should be preserved so it can be referenced and reused in the long-term. Long-term preservation and sustainable services are however challenging and costly to achieve. As a result they are still areas of work where much more attention is needed. Certification models like CoreTrustSeal were recommended for adoption in EOSC by the Turning FAIR into Reality Expert Group report and should be pursued. We need to be able to trust the services which store our data for the long-term, as well as measuring the quality and FAIRness of the data itself. Having certified repository services that guarantee a good quality of FAIR implementation within their area of expertise could be one route to ensuring ongoing trust.

» Enablement: technology AND people

The human infrastructure that supports the use of technology is critical and often overlooked. Cloud service providers have dedicated service teams that set-up and optimise cloud infrastructure for deployment and computing scale up. Offering 'last mile' support to help research groups put services in place is incredibly important. This is already done well by individual infrastructure providers, but will need to be done in an integrated, coordinated way for EOSC. We can't simply list services and assume users know what they need and how to combine them. Some facilitation, either through a centralised helpdesk or via intermediaries such as data stewards and lab managers who work closely with research groups and could act as a broker or liaison with EOSC, is recommended.

Recommendations on how to implement EOSC

The methods and approach we use to address the aforementioned challenges is also important. Given that EOSC is a multi-stakeholder initiative, serving many research communities and a diverse set of needs, a modular, iterative approach that ensures inclusivity and places researchers at the centre is recommended.

» Place researchers at the centre

The primary mantra repeated throughout the panel discussion was that we need to base service development on real researcher requirements and co-design together with them. The "If we build it, they will come" approach will never work well for such a grand vision and heterogeneous landscape. We need direct input from scientists that work with data on a day-to-day basis to ensure that the policies and services fit their needs. Researchers also have an incredible amount of expertise and could offer fresh insights to help achieve the EOSC vision. We must ensure responsiveness in this collaboration, and respect their existing working methods and environments.

» Be modular - iterate and test

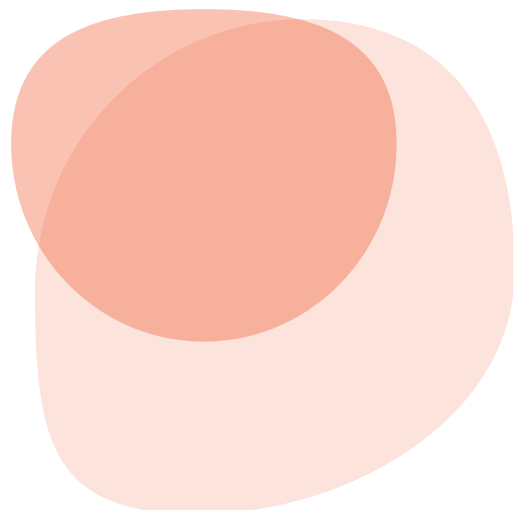
Another common theme was that we need to move away from (re)definition and proceed with implementation and testing. We should get

something out there asap and let users touch and feel it. Making one small thing work well and releasing it, then developing incrementally to build on this, can create a huge impact. This is far better than waiting on several linked components to be ready and potentially risk mis-managing expectations. Many feedback loops should be foreseen, and openness on how that feedback was processed and acted on to build trust. Overall the panel argued for a lean approach, releasing regularly and gathering feedback to accelerate EOSC's progress, create a sense of community and build a platform that serves research community needs.

The panel also issued a plea for more convergence and a move away from the 'not invented here' syndrome which leads to much replication and redevelopment of similar services in the Open Science field. Building on open source approaches and stimulating the research community at large to use and improve existing tools, standards and registries is recommended.

» Ensure inclusivity

The language we use is often overlooked and we underestimate its effect on inclusiveness. Specialist terminology, acronyms and jargon are common in the Open Science and FAIR environment and can alienate certain stakeholders. We may easily fall into a pattern of working in a small in-crowd of relatively well informed people who speak to each other in what is almost a coded language that prevents others from engaging. We run a real risk of losing parts of Europe that are still less connected and the long tail of science not represented by European Research Infrastructures and clusters. Indeed even the basic definition of EOSC is still unclear for many. Harmonising on a central vision and using clear language that everyone can understand is seen as a key task for the Association.



Personal reflections

The final question to the panel was what they could change if they had a magic wand to make one thing in EOSC better. The reflections below offer personal insights.



Sarah Jones, EOSC Engagement Manager, GÉANT, and EOSC Association Board Director

In my perspective we placed the cart before the horse in forming EOSC, as we only introduced the governance structures after many projects were funded. It remains challenging to offer effective coordination and steer initiatives. The EOSC Association now has a key role to play in delivering an effective stakeholder forum and working in partnership with the service providers to meet research community needs, which should be at the core of EOSC developments.



Rudolf Dimper, IT Advisor, European Synchrotron Radiation Facility (ESRF), Grenoble, France

If we define EOSC as a tool for carrying out research and serving researchers, my magic wand would ensure that researchers take control, request services and functionalities, and use those in their daily work. This would trigger an avalanche of requests for “more” and ensure that Open Science becomes the “new normal” in Europe and beyond. EOSC would make a significant contribution to reinforce trust in the scientific endeavor, and I think most of us know how important this is nowadays where obscurantism is poisoning our societies.



Ingrid Dillo, Deputy Director, DANS, The Netherlands

I would use my magic wand to materialise the EOSC by combining the best of the public and private world. A commercial player would probably guarantee a quick and agile delivery of an undoubtedly attractive EOSC end-product. But there might be issues with openness and consensus. In the public realm we build an EOSC that is open, inclusive and based on community consensus, which is incredibly valuable. But it comes with a price as well. It would be great if

we could diminish the complexity of the process and the enormous amount of talking and red tape which makes progress very slow. So with my wand I would create an open and inclusive EOSC based on community consensus, with a wonderful ease-of-use and get it done before the end of the year.



Hilary Hanahoe, Secretary General, Research Data Alliance (RDA)

If there was more global transparency, more coordinated communication we would be in a position to minimize the continued duplication of efforts and to engage the many researchers who still know nothing about EOSC and its potential benefits. Use the Research Data Alliance to showcase and advance EOSC which in turn supports open research and open science across the globe for the benefit of society. The massive COVID-19 research, developments and discoveries on a global scale over the past 18 months is all the proof that we need to show that this is essential for the future of science and, hence, society.



Shalini Kurapati, Co-founder and CEO, Clearbox AI and Open Science Fellow, Politecnico di Torino

If we could go back in time and rename and re-strategize the vision and goals of EOSC, so that it would be both an inclusive and an exigent player in the global Open Science ecosystem. I would engage end users from the very beginning to co-create it not only to be useful but also indispensable for their FAIR journey. I would also emphasise the need for a clear and sustained communication strategy to highlight its strengths such as trustworthiness, accessibility of resources to researchers, quality seal, data sovereignty, stewardship and related societal benefits.

Conclusion

This paper, based on the panel discussion held during the EOSC Symposium in 2021, reflects on the status of EOSC and what the panelists think is needed to achieve rapid progress to make EOSC usable and useful for European researchers. The technical developments undertaken by past and current projects to build EOSC need a bottom-up consolidation where individual components are quickly released and exposed to researchers using them. Early feedback will allow enhancements to those components and ensure that an incremental approach creates trust and inclusiveness. There is a key role for the EOSC Association to play in making this change. It requires strong leadership and a commitment to purpose to shift the culture to the collective, consensus-building approach spoken about in the opening. The core functions of EOSC have to be kept simple - simple to use and simple to understand. The authors fully understand that this is very challenging for a "system of systems" with such diverse stakeholders and interests. But now is the time to prove that the concept of EOSC is valid and achievable and helps researchers in doing their work for the benefit of our society.