



**1. Define the EOSC reference architecture** to map out the functionality and interfaces offered by the core services that will make Europe's existing research data infrastructures interoperable and accessible via a single mechanism. Absolutely fundamental to delivering the European Open Science Cloud, this reference architecture will describe what federation actually means in practice - how data and software tools from many different scientific disciplines can be made accessible to all kinds of researchers. To that end, the architecture will have to encompass generic horizontal services, such as a portal, authentication, authorisation and security services, as well as the catalogue of EOSC services for data storage, management and analytics, simulation and visualisation and distributed computing (see working group 2 described below). As outlined in the European Commission's staff working document, EOSC shared resources need to be developed to cover all the aspects of FAIR data:

- Findable, through e.g. catalogues of data/ services and metadata;
- Accessible, through e.g. Persistent Unique Identifiers, Data Management Plans;
- Interoperable, through e.g. interoperable standards and common metadata; -
- Reusable, through e.g. common IPR and legal provisions (e.g. Creative Commons).

The data and services providers who need to connect to the core EOSC services should review this architecture.

**2. Complete a full definition of EOSC services** based on an analysis of how research is being conducted today and the on-going advances in cloud services. Building on the results of the 'EOSC reference architecture' working group, this working group would select implementations of the core services from amongst the candidates available. Clearly this step is critical to the implementation of the EOSC and should be a priority to ensure the cloud is a practical tool that makes scientific research easier and quicker. As part of this effort, this working group should consider how the existing trans-European public e-infrastructure can best collaborate and utilise federated services to serve a wider range of European researchers and scientists, whilst also investigating synergies with commercial partners in the research and education sphere to bolster the EOSC service catalogue. In particular, it should consider how the EOSC can best harness the latest technologies and services provided by commercial cloud providers. The working group should also investigate how the service catalogue can be complemented by e-infrastructure providers within the European Data Infrastructure (the EDI).

**3. Detail the rules of participation for the EOSC**, starting with rules for data providers, service providers and end-users. The key stakeholders won't fully participate in the EOSC until they know what commitments they

will need to make and the rules that they will need to comply with. These rules should seek to ensure that the EOSC balances the interests of the different stakeholders, while complying with European laws and ensuring research data is available on a FAIR (findable, accessible, interoperable and reusable) basis. All of the federated research data infrastructures and commercial service providers will need to comply with the Rules of Participation. This working group could also explore how European legislation could be amended to encourage scientific research to be opened up.

**4. Establish how the EOSC will be integrated into global scientific research** - identify policies and specifications to allow the EOSC to interact with similar structures in other regions of the world to enable open science on a global level. Integrating the EOSC into the science clouds being developed by third countries is crucial because scientific research is global - meaningful progress generally depends on sharing data and insights across international borders. The EOSC needs to be seamlessly integrated into the global network of scientific data/information exchange. At the same time, this working group will need to consider how to achieve reciprocity and ensure that all parties accessing EOSC services comply with European data protection laws. Still, by moving early to open the EOSC up to the world, the EU should encourage third countries to also make research data FAIR.

**5. Identify the optimum legal and governance structures for the EOSC beyond 2020**, supported by a funding model that takes into account the output from the 'business model and incentives' working group (see number 7). Until it has a robust legal and governance structure, the EOSC's impact will be limited: the executive body will need to be empowered to award contracts, make investments and procure services in a timely manner. To that end, the body's procurement processes and legal position will need to be clear and transparent. The governance model for the second phase (2021 onwards) should take into account experience gathered during the first phase through a consultation with all participants.

**6. Establish a detailed segmentation of the EU scientific research community** - a market analysis for each segment to understand the nature and scale of the services they require from the EOSC, as well as how such services are currently provided and funded. As proposed by the EU Competitiveness Council, this analysis could be underpinned by a map of national research data infrastructures and initiatives in the Member States that can be federated. Ideally the map should become a form of compendium with an overview of the current status, contents and scale of the EOSC, enabling the EOSC Board to monitor the progress of the EOSC.

7. **Define the business model and incentives** for each segment of EOSC users. The initial priorities for the EOSC have to be driving participation and usage. Like most private businesses, the EOSC will probably need to operate at a loss (be subsidised) in its early years to ensure its proposition is appealing to both the data providers and the data users. However, once it is benefitting from network effects, the EOSC should seek to implement a sustainable business model that ultimately enables it to become self-financing. To do that, the EOSC will need to be highly relevant for the private sector, as well as the public sector, providing valuable data sets and tools at an attractive cost. This working group should map out how the EOSC can develop a compelling business proposition. Note, this commercial proposition should allow for citizen science carried out by independent researchers to draw on the resources of the EOSC.

8. **Implement a process to enable total quality management of data** throughout its lifecycle via the certification of digital repositories that can preserve data for the long-term and store it securely, where necessary. There will be a need for funding/incentive mechanisms that ensure these depositories can safely store data for years and even decades, depending on the scientific discipline involved. The starting point should be an analysis to determine the most cost-effective way to preserve scientific data for the long term. At the same time, the EOSC will need to reconcile individuals' legitimate right to privacy with the need for behavioural data to fuel progress in the social sciences, the humanities, medicine and healthcare.

9. **Define the long-term roadmap of the EOSC** and how it will develop. The EOSC will need to be guided by a clear strategic plan that will give clarity and certainty to all potential stakeholders. This roadmap should paint a compelling and inspiring picture of what the EOSC will deliver, and spelling out how it will engage with the private sector. This will help attract the attention of the many commercial companies that have yet to engage with the EOSC. In the interests of transparency, the roadmap should be published and revised on an annual basis and be divided into three sections identifying activities to be performed in the short-term (1 year), medium-term (2-3 years) and long-term (4-5 years).

10. **Develop a plan to address the relevant gaps in skills and education** so that end-users can make full use of the EOSC. One of the biggest barriers to the medium to long-term success of the EOSC will be researchers' ability to manipulate and analyse the data it captures and makes accessible. Therefore, this working group should focus on how to ensure the research community has the data science expertise and services it needs, while considering how advances in artificial intelligence, robotics, quantum computing and the Internet of Things could change the skillsets required in future. For

example, some future research projects may need to be guided by people with a strong understanding of ethics. In practice, this is likely to mean running education programmes for both academics and students.

#### **Additional considerations**

- The EOSC Executive board should ensure that working groups are complementary, innovative and delivery-focused.
- Where appropriate, the working groups should build on the preparatory work performed by the EOSC-pilot and EOSC-hub, such as their draft versions of the reference architecture and rules of participation.
- The new groups also need to work with existing implementation projects (such as eInfraCentral and EOSC hub). Where appropriate, the board should identify and reduce areas of remit overlap and any potential duplication of work between the working groups and current/future implementation projects.
- The number of working groups should be limited - a large number of working groups significantly increases the risks of silos developing. Each working group should have a tightly defined objective. Once that objective has been completed, the group should cease to exist.
- The working groups should aim to complete their function within two years, paving the way for the EOSC to have as much impact as possible for 2021.
- Initially, to limit the time and scale of work to be performed, the EOSC should be established as a 'minimum viable ecosystem', which can be expanded during a second phase (after two years) taking into account feedback on initial usage and participation.
- The board should identify inter-dependencies between working groups and the duration and scheduling of their terms should reflect these inter-dependencies, as well as the duration of current/future implementation projects.
- Where the work of the individual working groups will be inter-related, the progress of each group should be communicated to all others.
- Working group chairs should meet frequently and the EOSC could organise events with a wider participation twice per year.
- Transparency will be necessary to ensure trust amongst EOSC participants so the output of all the working groups should be openly published.

**Contact:**

David Pringle, Special Advisor, Science|Business  
[david.pringle@sciencebusiness.net](mailto:david.pringle@sciencebusiness.net)

# SCIENCE | BUSINESS<sup>®</sup> NETWORK

Bringing together industry, research and policy

## Industry

Amgen	Nickel Institute
Elsevier / RELX	Novartis
Frontiers	Pfizer
GE	Sanofi
Google	Total
Huawei	Toyota
Microsoft	

## Academia

Aalto University	Trinity College Dublin
ESADE Business & Law School	TU Berlin
ETH-Zurich	University College London
Imperial College London	University of Amsterdam
KTH Royal Institute of Technology	University of Birmingham
KU Leuven	University of Bologna
Norwegian University of Science and Technology (NTNU)	University of Eastern Finland
Politecnico di Milano	University of Luxembourg
Polytechnique Montréal	University of Pisa
Sorbonne University	University of Twente
Tallinn University of Technology	University of Warwick

## Public organisations

Barcelona Supercomputing Center	Eureka
Business Finland	Fraunhofer
CERN	Innovate UK
Centre National de la Recherche Scientifique (CNRS)	Innovation Norway
The COST Association	Republic of South Africa - Department for Science and Technology
European Investment Bank	

## Associations

ACM Europe Policy Committee  
ATTRACT  
European University Association (EUA)  
Photonics 21  
The Association of Commonwealth Universities (ACU)  
The Guild

## Group members

Hospital Saint Joan de Deu  
Deusto International Research School (DIRS)  
ICHOM - International Consortium for Health Outcomes Measurement  
GEANT  
Simmons & Simmons  
Tataj Innovation

[www.sciencebusiness.net](http://www.sciencebusiness.net)