EUROPEAN OPEN SCIENCE CLOUD

Open Consultation

for the
Strategic Research and Innovation Agenda (SRIA)

of the
European Open Science Cloud (EOSC)

July 2020

Open Consultation

for the

Strategic Research and Innovation Agenda (SRIA)

of the

European Open Science Cloud (EOSC)

20 July 2020

- This document sets the draft general framework for future strategic research, development and innovation (RDI) activities to be further defined in the context of the candidate EOSC European Partnership[1] proposed under the Horizon Europe Programme.
- It uses elements of the candidate Partnership document proposed by the EOSC governing bodies as well as further work by the Executive Board, in order to develop, by October 2020, a first full version of the Strategic Research and Innovation Agenda (SRIA) for EOSC.
- With the consultation launched on 20 July, the EOSC governing bodies are seeking the views and contributions of different stakeholders on the content of this document through the accompanying questionnaire. The consultation will remain open until 31 August 2020.
- The feedback obtained in the consultation process will serve as input for the SRIA. The draft SRIA will be presented at the EOSC Governance Board meeting on 1 October 2020.

1 https://ec.europa.eu/info/files/european-open-science-cloud-eosc_en

Acknowledgements

This document has been written by the EOSC Executive Board[2] and published by the EOSCSecretariat.eu[3] project, which supports the Executive Board in its activities. EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019, Grant Agreement No. 831644. The document is for the most part based on the EOSC Partnership Proposal[4] and drafts for the Strategic Research and Innovation Agenda for EOSC. Thanks are due to the members of the EOSC Executive Board Working Groups[5] for their contributions to this document.

2 https://www.eoscsecretariat.eu/eosc-governance/eosc-executive-board
3 https://www.eoscsecretariat.eu/
4 https://ec.europa.eu/info/files/european-open-science-cloud-eosc_en
5 https://www.eoscsecretariat.eu/eosc-working-groups

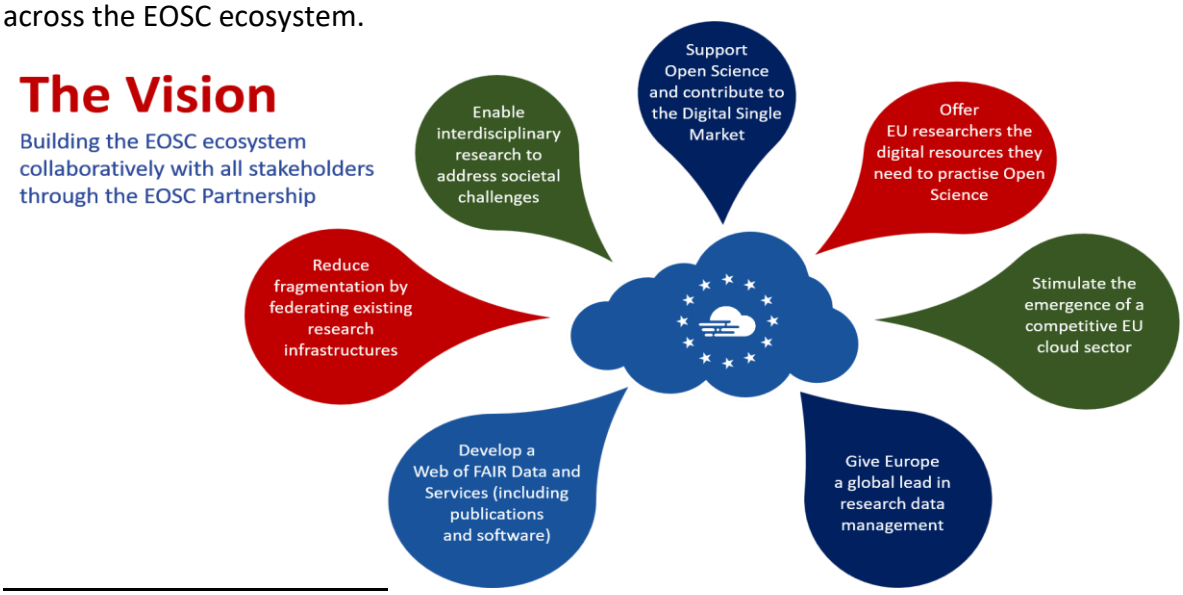
Table of Contents

1 EOSC Partnership vision and approach

The science system in Europe is in transition. It faces increasing challenges relating to trust in science and to reproducibility, and it needs to become more open, collaborative and effective. The recent coronavirus pandemic has demonstrated that in an emergency robust scientific data are of paramount importance for collaborative research and science-guided decision making. Data management and sharing are crucial for new scales of multi-disciplinary research and are expected to stimulate major industrial and social innovations while adapting cutting-edge digital technologies and interconnecting the European landscape of research infrastructures and e-infrastructures. Accelerating this transition of the European science system requires a multi-stakeholder European partnership to enhance the circulation of research data and knowledge in digital form across borders and disciplines, and to allow scientists and machines to collaborate in creating, storing, processing, finding, accessing and reusing scientific data. The new standard must therefore be the sharing and reuse of data and metadata across all scientific disciplines.

1.1 The vision

The European Open Science Cloud (EOSC) Partnership[6] will enable a trusted, virtual, federated environment in Europe to store, share and reuse digital outputs from research (including publications, data, metadata and software) across borders and scientific disciplines. The EOSC Partnership will bring together institutional, national and European initiatives, data and service providers and all relevant stakeholders to co-design and deploy a European Research Data Commons where data are findable, accessible, interoperable and reusable (FAIR). EOSC will build on the past decade of investing in Open Science projects and initiatives by the European Commission, Member States and others. This European contribution to what can be seen as a 'Web of FAIR Data and Related Services for Science' (hereafter referred to in this document as the Web of Fair Data and Services) will enhance the possibilities for researchers to find, share and reuse publications, data, and software. EOSC will stimulate and enable researchers to work collaboratively and practise Open Science as well as tackle the global societal challenges of the 21st century. The endeavour is not limited to linking datasets, federating infrastructures or aligning policies; it starts by linking people and organisations across the EOSC ecosystem.



6 https://ec.europa.eu/info/files/european-open-science-cloud-eosc_en

1.2 Main drivers

- ***The exponential growth in the quantity of research data.*** The amount of generated data is growing exponentially and is measured in Zettabytes (10^{21} bytes). The vast increase in data production equally applies to the domain of research, with the result that researchers are already unable to read (in terms of amount) or access (in the case of data) all relevant digital knowledge in their field. In particular, the underlying research data (positive, inconclusive and negative results) remain predominantly unpublished and are therefore unfindable and inaccessible.
- ***Creating interoperable data commons for data-intensive science and innovation*.** Scientific data are in dire need of openness, better handling, careful management, machine actionability and sheer reuse[7]. This cannot be realised without specifications and standards for common components to enable interoperability across the research data ecosystem. Moving to digital research outputs that are FAIR by design requires further efforts to develop, refine and adopt shared vocabularies, ontologies, metadata specifications and standards, as well as increasing the supply and professionalisation of data stewardship. The science and innovation data space envisioned by EOSC shall be properly articulated within the new sectoral data spaces foreseen in the European Data Strategy[8] servicing researchers, public authorities and the commercial sector.
- ***Bridging existing European infrastructures*.** Over the last decade or so, there have been significant investments across Europe in data-oriented research infrastructures and e-infrastructures. The outcome is a vast quantity of infrastructure components of various scales and scopes, centralised or distributed, whether generic, domain-specific or cross-disciplinary. Many of these components have not been designed initially to work together. Although progress has been made, interoperable data commons bridging these European infrastructures and connecting data silos is badly needed.
- ***Connecting research data with robust computing technologies and fast connectivity networks.*** Europe needs a robust European-scale environment for data storage, analysis and computing. Access policies for networking, data storage, computing and processing differ across Europe. This makes scientific cooperation in the EU more difficult, especially if the aim is to leverage existing e-infrastructures across countries and disciplines. Shareable research data, open data analysis tools and connected computing facilities need to become available for all researchers and research-support units.
- ***The policy drive for Open Science and open innovation in Europe.*** Open Science and open innovation are widely recognised as key transformative, enabling elements of the European RDI policy driving a renewed European Research Area (ERA). While many of the Member States and Associated Countries have policies in place regarding open access to scholarly publications, few countries have yet defined strategies regarding open/FAIR data and only some mention EOSC in their policies (although some plan to do so in the future). There is a need to progressively open up the EOSC ecosystem to allow not only European academics and researchers but also public authorities, innovators, companies and society at large to contribute to and exploit the benefits offered by open data, publications, software and associated services in Europe.

7 'Realising the European Open Science Cloud: First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud', 2016.

8 'A European strategy for data', COM(2020) 66 final, 19.2.2020.

1.3 Value of a European Partnership approach

A European Partnership is identified by the current EOSC governing bodies as the most effective instrument to overcome fragmentation and to provide a framework for collaboration and pooling of resources at European, national, regional and institutional levels. An open and inclusive European Partnership will help ensure directionality (common vision and objectives) and additionality (complementary commitments and contributions at all levels). A Partnership helps to provide a focal point and a framework to reach consensus amongst those committed to achieving results. By agreeing on a common Strategic Research and Innovation Agenda (SRIA) reflecting the views of the EOSC communities, the Partnership will demonstrate a shared vision and operational objectives for the next seven years and beyond. A new legal entity (the EOSC Association) is currently being established to act as the focal point for the EOSC Partnership. The EOSC Association will be founded as a not-for-profit Belgian Association (AISBL), open to interested stakeholder organisations that comply with the published Rules of Participation.

1.4 Guiding principles to implement the EOSC vision

- ***Multi-stakeholder approach.*** The EOSC vision is not limited to linking datasets, federating infrastructures or aligning policies; it starts by linking multiple stakeholders throughout the data lifecycle and across the European research ecosystem. These typically include the Member States (for instance, through national organisations having received a mandate from their ministry to represent their national research landscape), research-performing organisations (including universities and research centres), research-funding organisations, research infrastructures in all domains and e-infrastructures (e.g. related to storage, computing and communications), research libraries, research associations, international research centres, as well as (at a later stage) other entities from the public and private sectors.
- ***As open as possible, as closed as necessary.*** The initial focus of the implementation will be on open data that are as FAIR as possible. This will allow the deployment of reliable services and applications that will become examples to demonstrate the benefits of Open Science. In parallel, emphasis will be placed on enabling new types of incentives and skills to share data openly, improve FAIR data management, and develop literacy and data stewardship skills. As EOSC matures, it will gradually expand towards more complex and specific challenges in areas where FAIR data cannot be kept open by default. Data protection, for example General Data Protection Regulation (GDPR) compliance, will have to be ensured in order to extend the reach of EOSC to disciplines in which personal data are processed.
- ***Towards a Web of FAIR Data and Related Services for Science*.** The FAIR guiding principles and related metadata standards will bring the minimum soft overlay required to enable common implementation and form a Web of FAIR Data and Services. At the same time, it will offer maximum freedom of implementation and will allow participation from all stakeholders, including research (e-)infrastructure providers, research-funding organisations, research-performing organisations and businesses.
- ***Federating existing research infrastructures*.** The landscape of European research infrastructures has been shaped by the planning of pan-European projects in all domains of research through the European Strategy Forum on Research Infrastructures (ESFRI) and is largely reflected in national roadmaps. Preliminary findings of the ongoing mapping of this landscape by the EOSC governing bodies have confirmed that a majority of Member

States and Associated Countries are making significant investments in both pan-European and national infrastructures that could in principle contribute to the EOSC federated data ecosystem. The ESFRI Roadmap[9] represents an investment of 20 billion euros and the operational costs of the European research infrastructures are in the order of 10 billion euros per year. Roughly 15% of the resources of a typical research infrastructure are directly invested in data management, storage and computing. Dedicated e-infrastructures provide data curation, services and computing power to the research system.

- ***Machines and people*.** Computers have long surpassed individuals in their ability to perform pattern recognition over large datasets. However, actionable knowledge and translation of its benefits to society will still be handled by humans for decades to come. As computers have become indispensable research assistants, everything published by science needs to be made understandable to these computers. In concert, scientists need to form what may be called social machines in order to be able to visualise, analyse and predict when facing complex problems. The availability of data that are FAIR by design and of shared application programming interfaces (APIs) will allow new ways of collaboration between scientists and machines to make the best use of digital objects of any kind. All these machine-run algorithms need to be transparent to the researchers, to be effective accelerators of reliable data analysis and results.

1.5 Overall timeframe

Three stages of implementation are foreseen, serving the following two major aims[10]:

- deploy European Open Science Cloud operations to serve EU researchers by 2025;
- open up, connect and deploy EOSC beyond the research communities to the wider public sector and the private sector from 2024 onwards.

Stage 1: period 2021–2023

Creating the **European Open Science Cloud operations (EOSC-Core)** to provide authentication and authorisation infrastructure (AAI) and other necessary core functions of the Minimum Viable EOSC.

Stage 2: period 2024–2025

Expanding the Minimum Viable EOSC with **access to added-value services, applications and tools (EOSC-Exchange)** supporting the full cycle of scientific workflows.

First pilots/demonstrators on linking EOSC beyond the research communities to the wider public sector and the private sector from 2024 onwards, for addressing societal challenges.

Stage 3: period 2026–2027 and beyond

Deployment of the **Web of FAIR Data and Services**, including the EOSC-Core, EOSC-Exchange and other framework conditions for interoperability and machine actionability of data. Connection of the European contribution to a Web of FAIR Data and Services to other Open Science commons across the world.

Continuous support to enhance the ecosystem of the Web of FAIR Data and Services for the research community.

9 https://www.esfri.eu/esfri-roadmap-2021

10 'A European strategy for data', COM(2020) 66 final, 19.2.2020.

1.6 Synergies with other initiatives: boosting the impact of EOSC through collaboration and alliances

The EOSC Partnership is a key partnership in a wide network of other partnerships, initiatives and infrastructures that form the digital and research ecosystems in Europe and globally. Thus shared activities should be undertaken and alliances built in order for Europe to optimise the outcomes of its efforts for the benefit of the research community and society at large.

European partnerships and missions

The EOSC Partnership aims to accelerate the transition to more effective Open Science and open innovation with new opportunities for multi-disciplinary research, frontier science and data-intensive science. Therefore, the Partnership has a unique transversal role (cross-pillar, cross-cluster) in Horizon Europe, bringing the potential for collaboration with most of the proposed European partnerships.

- All vertical partnerships can benefit from a successful development of EOSC as it will equip them with minimum, rigorous standards and protocols and maximum freedom of implementation to share and reuse data and other digital objects across their domains.
- The EOSC Partnership will join forces with partnerships working on specific technologies, such as the European Partnership for High-Performance Computing and the European Partnership on Artificial Intelligence, Data and Robotics.

European data spaces[11]

The EOSC Partnership will also seek synergies with the European data spaces. While EOSC focuses on federating research data infrastructures, the research data made accessible through EOSC will be relevant for several data spaces (e.g. health and the Green Deal). EOSC will build primarily on FAIR data. Where appropriate, access to such data will be provided.

Open Science organisations in other parts of the world

Open Science has become a shared goal across the globe. While Europe is in a leadership and pioneering position, there are many other initiatives around the world.

- Existing Open Science commons initiatives include the Australian Research Data Commons (ARDC), the African Open Science Platform (AOSP), the US National Institutes of Health (NIH) Data Commons, Canada's National Data Services Framework (NDSF), the China Science and Technology CSTCloud and the Association of Southeast Asian Nations (ASEAN) initiative.
- In the thematic realm, examples are the Australian Biocommons, the Data Commons for Food Security and the Commonwealth Scientific and Industrial Research Organisation's (CSIRO's) Managed Data Ecosystem.
- The Research Data Alliance (RDA) has become a recognised actor in the development of standards and good practices for managing research data. The Committee on Data (CODATA), as the standing committee of the International Science Council, works closely with the RDA and GO FAIR working groups and implementation networks, while the World Data System focuses on high-quality data services with an emphasis on repositories and their certification.

11 'A European strategy for data', COM(2020) 66 final, 19.2.2020.

1.7 Synergies with other programmes

The EOSC Partnership can play a strategic role in supporting the EU policy priority of having a Europe fit for the digital age. Therefore, at European level, synergies with other programmes, in particular the Digital Europe Programme and the Connecting Europe Facility, in areas such as the development of digital skills, standards or tools, should be exploited in order to pool the necessary resources and avoid overlaps.

Synergies with the European Structural and Investment Funds[12] should also be exploited in order to leverage the investments through structural funds (such as for the Smart Specialisation Strategy sectors) in research data infrastructures that can be federated as part of the EOSC ecosystem. Moreover, a majority of Member States are making significant investments in national infrastructures and Open Science programmes that could, in principle, be federated as part of the EOSC ecosystem. There is, therefore, a strategic European value in improving the alignment and compatibility of national Open Science policies and national plans for data infrastructures in the EOSC context.

12 https://ec.europa.eu/regional_policy/en/policy/what/glossary/e/esif

2 Strategic Objectives and Action Areas

Three main *Strategic Objectives*[13] and fourteen *Action Areas* have been defined to accelerate the development and deployment of the EOSC ecosystem in accordance with the vision, principles and timeline described in Section 1. These Strategic Objectives and Action Areas are described below.

2.1. Strategic Objectives

The Objectives Tree below sets the foundations for the EOSC Partnership intervention logic along three main Strategic Objectives.

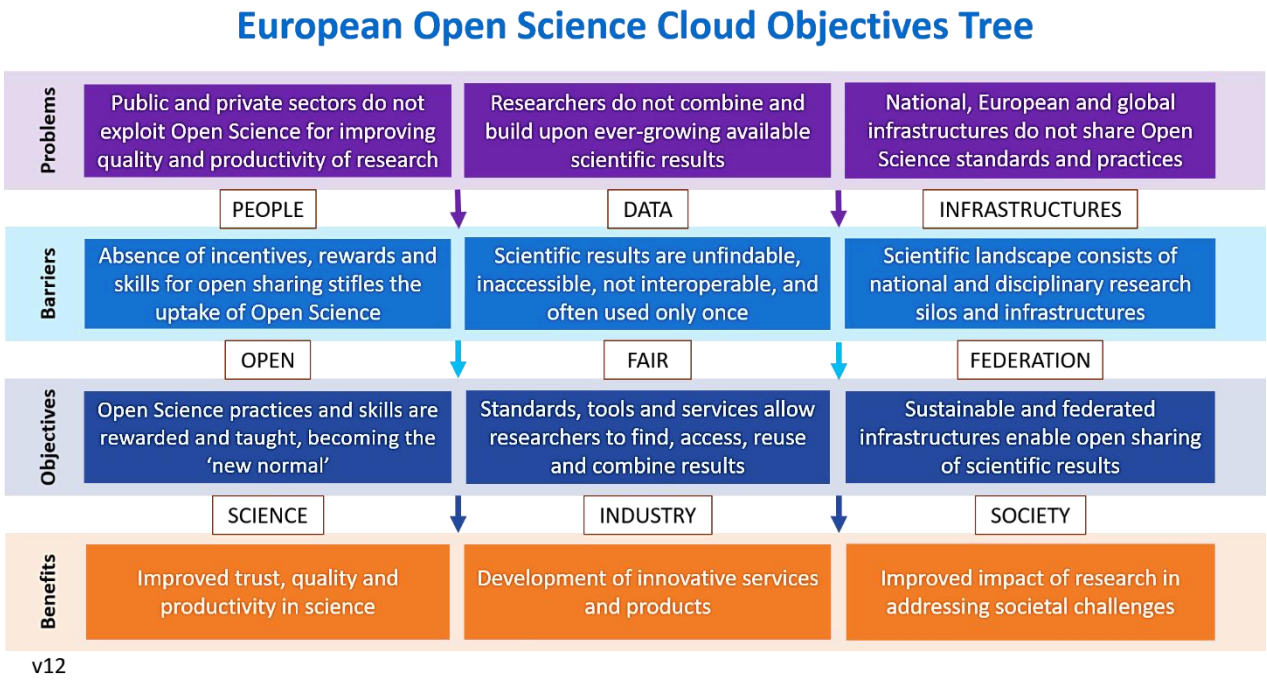


Figure 1: EOSC Objectives Tree

Strategic Objective 1

Open Science practices and skills are rewarded and taught, becoming the 'new normal'

Main milestone. The EOSC ecosystem underpins the reward of Open Science practices and data stewardship, thus improving trust, quality and productivity in science.

Scope. A key goal of the EOSC Partnership is to help move the research enterprise in Europe towards an Open Science model as there is a political will towards the notion of open research and many European countries are implementing national programmes that are aligned with the European Commission Recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information[14].

EOSC will be established as the EU-wide infrastructure for open research. Assuming the European Commission and national funders will require researchers to use EOSC-approved infrastructures, this will lead to output in line with the 'new normal'.

13 For a more detailed description of the Strategic Objectives see the EOSC Partnership proposal where they are named 'General Objectives'.

14 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018H0790

Strategic Objective 2

Standards, tools and services allow researchers to find, access, reuse and combine results

Main milestone. Making data and any other digital research object (such as algorithms, tools, and workflows) as FAIR as possible across all European research infrastructures will be key for sharing/reusing research results.

Scope: The FAIR principles[15] provide high-level answers to the challenges involved in managing research data:

- ***Heterogeneity.*** Data are defined by the scientific discipline that produces or studies them. They originate from equipment, surveys or any other data production/collection process. Data are also the results of processing steps taken after the raw-data collection. As a consequence of this multi-faceted landscape, a very wide variety of data frameworks have been designed by research communities.
- ***Granularity.*** While raw data may be produced in large volumes, users may not need access to the raw data itself but only to the results of post-processing steps. Deciding at what stage data should be preserved and made available to users is also specific to the scientific domain.
- ***Versioning.*** Data are collected over time. The sampling rate, and the importance of preserving information on the sequence of events, also depend on the domain.
- ***Disambiguation.*** When it comes to representing data in digital form, the choice of identifiers is a problem in itself. One identifier may point to multiple objects, introducing another level of complexity for the user.
- ***Diffusion prevention.*** On the other hand, multiple identifiers may point to the same object, introducing a different level of complexity.

The availability of data that are FAIR by design will allow scientists to make the best use of new data by leveraging the power of machines. FAIR data, being machine-actionable, allow the development of software services, applications and tools that deliver the requisite information for scientists to optimise their research.

Researchers are increasingly reliant on computational and machine-assisted support to deal with research data as a result of the increase in the volume, complexity and creation speed of that data. There is thus a scientific and policy consensus that research data must be made machine-actionable, when applicable, to allow computational systems to find, access, interoperate and reuse the data.

Digital objects must be described with rich metadata, assigned a globally unique persistent identifier, and be released with a clear and accessible usage licence. There is an onus on researchers to adopt relevant community standards and select appropriate data services that enable digital objects to be discovered and retrieved using standard protocols, applicable for both humans and machines. Research communities need to define data standards, sharing agreements and services to enable FAIR digital objects.

15 https://www.go-fair.org/fair-principles/

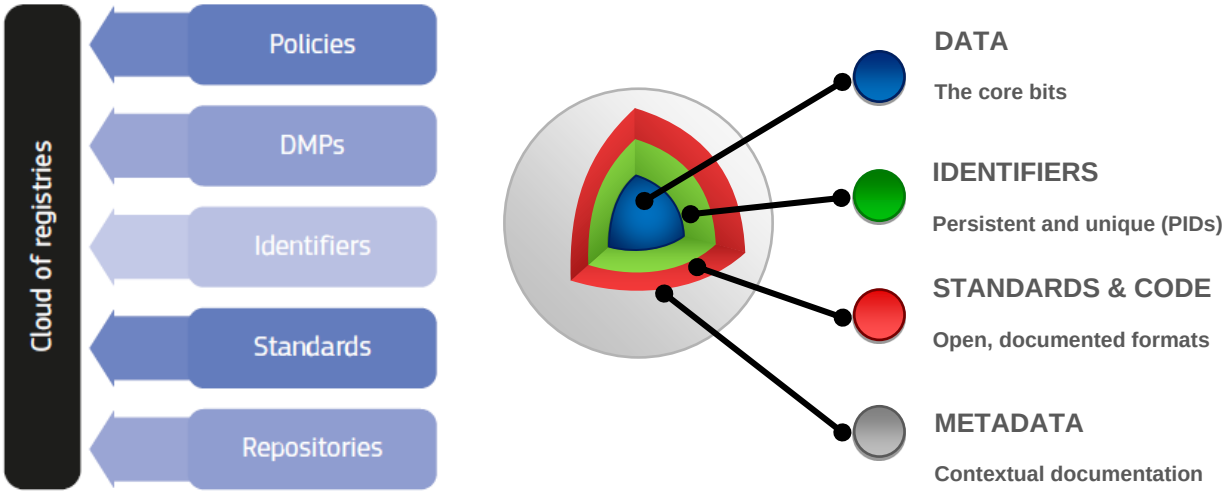


Figure 2: FAIR ecosystem

The strength of the FAIR principles is in defining a set of common characteristics required in all digital objects, irrespective of type, discipline and content. This enables machines to act across a broad set of content, enabling interdisciplinary research. Many aspects of the FAIR principles, however, addrerss community-specific standards and practices. The principles will be applied differently according to the needs and requirements in the different fields of knowledge. Crosswalks and brokering are needed to support interoperability across the standards of multiple disciplines. The interoperability frameworks should be articulated in common ways and adopt global standards where relevant. Intelligent crosswalks, brokering mechanisms and semantic and other cutting-edge technologies such as artificial intelligence should all be explored to break down silos and allow cross-disciplinary data exploration, analysis and visualisation.

Strategic Objective 3

Sustainable and federated infrastructures enable open sharing of scientific results

Main milestone. EOSC will operate a federation of infrastructures, forming a Web of FAIR Data and Services, supporting research by the academic, private and public sector. Building on existing research data infrastructures, EOSC will grow through a series of iterations. Each iteration will add more functionalities and services for a wider user base and satisfy a broader range of use cases.

Scope: The system will be based on three layers: (1) the federating core (or **EOSC-Core**), (2) the **federation** of existing and planned research data infrastructures, and (3) a service layer comprising common services and thematic services (**EOSC-Exchange**).

(1) The **EOSC-Core** assembles all the basic elements to operate and provide the means to discover, share, access and reuse data and services in a reliable manner. These elements address key technical, cultural and policy decisions of EOSC and they must be maintained over the long term.

(2) The FAIR principles and metadata standards enable the **federation** of existing and planned research data infrastructures, adding a soft overlay to connect them and forming a Web of FAIR Data and Services. This is a distributed system that needs an underlying framework

based on commonly agreed, minimum standards and maximum freedom to operate with agility, whilst still ensuring global and interdisciplinary interoperability.

(3) The **EOSC-Exchange** builds on the EOSC-Core to ensure that a rich set of services (common and thematic), exploiting FAIR data and encouraging its reuse, are available to publicly funded researchers. It is expected that rivalrous services, such as those that store, preserve or transport research data as well as those that compute against it, will be made available via the EOSC-Exchange. Service providers that participate in the EOSC-Exchange will be required to conform to predefined Rules of Participation.

2.2. Action Areas

The EOSC governing bodies have identified fourteen Action Areas (AA) to help deploy the EOSC ecosystem. Some are more technical in nature, others relate more to social dimensions (financial, legal, educational, cultural, policy). They are listed under the respective headings Implementation challenges and Boundary conditions in Table 1 below and described in the following sections.

NOTE: *Please note that the numbering of the Action Areas (AA1, AA2, etc.) is not related to their level of importance. They are numbered so that they can be referred to in the accompanying questionnaire.*

Implementation challenges	Boundary conditions
AA1: Identifiers AA2: Metadata and Ontologies AA3: FAIR Metrics and Certification AA4: Authentication and Authorisation Infrastructure AA5: User Environments AA6: Resource Provider Environments AA7: EOSC Interoperability Framework	AA8: Rules of Participation AA9: Landscape Monitoring AA10: Business Models AA11: Skills and Training AA12: Rewards and Recognition AA13: Communication AA14: Widening to the Public and Private Sectors

Table 1: Action Areas

AA1 – Identifiers

The persistence of the identity of digital objects and stability of references to those objects are essential to the European Open Science Cloud. Only if researchers can be assured that digital objects (including publications, data and software resources) do not alter over time and are continuously accessible via linking mechanisms can a trusted distributed research ecosystem that supports verifiable and reusable research be sustained. The use of persistent identifiers (PIDs) has been specifically recognised within the FAIR principles as a key feature supporting the findability and accessibility of research objects. PIDs therefore form a stable, trusted structure which can be used to make the research infrastructure a reliable source of verifiable and reproducible research. EOSC should seek to support a shared policy for the use of PIDs both for the management and analysis of data, and also for the publication, curation and tracking of research outputs.

AA2 – Metadata and Ontologies

Metadata and ontologies are essential to realising Open Science, and thus are an important topic that needs to be addressed by EOSC. Metadata and ontologies have evolved organically over time, addressing the needs of individual communities and sub-communities. Because of these community-specific drivers, to date an overarching, coordinated approach to metadata and ontologies for scholarly resources has for the most part been missing.

Interoperability is thus the biggest gap that EOSC needs to address with regard to metadata and ontologies, and the path towards better interoperability is through the development of governance structures for how metadata and ontologies are used within EOSC. This governance should be built primarily around existing discipline-based communities but needs to be coordinated across these communities within EOSC, and with activities outside of EOSC, for example in the Research Data Alliance (RDA).

The work that these governance structures coordinate should include registries that describe metadata schemata in a standardised and machine-actionable way, better researcher-focused tools and services working with these metadata, crosswalks between existing metadata schemata, and training and documentation. The drivers for all work regarding metadata and ontologies should be use cases from and adoption by the researcher community, and the work should be based on existing infrastructure and communities.

AA3 – FAIR Metrics and Certification

The definition of criteria for FAIR potentially has very significant consequences, especially if metrics are used to decide on participation or funding. It is thus essential to use neutral fora for this purpose, and to seek international agreement, as has been done in the RDA FAIR Data Maturity Model Working Group[16]. The implementation of FAIR can only be achieved in an ecosystem. Research artefacts are made FAIR by the services in which they are created, discovered and reused. FAIR data maturity depends on the capabilities and trustworthiness of services such as repositories and PID provision.

Significant work has been devoted to certification of data repositories, but these frameworks need to be aligned with FAIR and applied to other types of services. Existing work on FAIR metrics and certification should be extended to ensure applicability across disciplines and support their implementation. FAIR assessments must be inclusive and take the specific research context and needs into account.

AA4 – Authentication and Authorisation Infrastructure

The purpose of authentication and authorisation infrastructure (AAI) in EOSC is to support the FAIR principles for data and services while enabling high-trust collaborations to be established and maintained with little or no friction to the end user.

As federated AAI provides trusted identity information and allows scalable management of roles and rights, it is a key concern to the security and trust of any collaboration. The AAI use cases in a complex ecosystem like EOSC are often hard or even impossible to achieve using the tools and design patterns used to provide enterprise or consumer identity.

16 https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

Fortunately, the e-science AAI community has a long history of building globally viable solutions for digital identity, which can continue to grow and develop within the EOSC framework.

AA5 – User Environments

User environments are the digital platforms users go to in order to interact with EOSC and EOSC resources. These include portals, dashboards, landing websites and, in general, services through which the EOSC resources are accessed and made useful to researchers. User environments for all actors in the scientific lifecycle, such as researchers, service providers, developers, funders, organisations, citizens, small and medium-size enterprises (SMEs), etc., may also include other platforms to fulfil users' requirements and workflows. Such requirements usually contain services for data simulation and analysis, ranging from generic services such as Jupyter Notebook to domain-specific applications per scientific application, including cloud and computing resources. Environments will provide users and user communities with an integrated view of their research needs, thus including not only the data-related aspects (data search, transfer, ingestion), but also all the services and resources needed in order to enable reuse via processing and analytics, even if these come from different providers and infrastructures.

The aspiration of EOSC is that user services can not only be found and used, but also be processed and combined into new added-value research options. This vision of composability would allow users to take resources from different sources and combine them, in as automated a manner as possible, within the user space to generate new scientific outputs. Such composition can be facilitated by the science gateways. Distributed and federated user environments have to take into account the researchers' need to use the best possible options to address the issue at hand. The scientific tradition also includes the way scientists produce their own tools. The capability to offer a single environment from which to direct and control the full lifecycle of data and processing is the added value of EOSC for the researchers.

AA6 – Resource Provider Environments

Resource provider environments are key to EOSC in that they are the way EOSC brings in the supply side from the broader community, from generic e-infrastructure through to thematic resources coming from the ESFRI clusters and research infrastructures. They offer an interface between the resource provider, a community and EOSC, consisting of a framework of processes, tools, approved standards, APIs and other elements. Adopting the EOSC Interoperability Framework within resource provider environments will enable EOSC users to compose solutions easily by assembling resources (e.g. computing, storage, services, data sources, datasets, publications or other research products) across resource providers to overcome their heterogeneity and interoperability barriers.

AA7 – EOSC Interoperability Framework

Achieving interoperability within EOSC is essential to federate services and provide added value for users. The draft EOSC Interoperability Framework[17] (which is currently open for comment) identifies general principles and organises them into four layers: technical,

17 https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf

semantic, organisational and legal. The Framework also contains a proposal for the management of FAIR digital objects in the context of EOSC.

The initial draft has been developed by members of the FAIR[18] and Architecture[19] Working Groups of the EOSC Executive Board. The authors conducted an extensive review of related literature and interviewed key stakeholders from European Research Infrastructure Consortia (ERICs), ESFRI projects, service providers and research communities. This helped to identify problems and requirements in each aspect of interoperability in order to provide recommendations for EOSC. Legal issues will be included in the next version, based on recommendations from a commissioned study[20].

AA8 – Rules of Participation

The Rules of Participation (RoP)[21] for the European Open Science Cloud aim to provide standards for policy, processes and procedures that provide assurance of quality and trust in the services offered through EOSC. The RoP apply to resources made accessible via EOSC, including data and services, and to the use of those resources when accessed through EOSC. The RoP define a minimum set of rights, obligations, accountabilities and responsibilities governing the activities of all those participating in EOSC, such as data and service users, data and service providers, and the operators of EOSC itself.

AA9 – Landscape Monitoring

A first Landscape report is being developed by the Landscape Working Group[22] of the EOSC Executive Board, describing activities relevant to EOSC in the European Member States (MS) and Associated Countries (AC) as well as some border countries. It summarises existing policies and investments based on inputs from the MS and AC, and on the expert knowledge of the Working Group members and delegates to the EOSC Governance Board. It also includes information from open sources (validated or extended by the countries' authorities when possible), as well as information gathered through Horizon 2020 research projects. In addition, the report provides an overview of the landscape of European infrastructures, since EOSC in its core is a federated structure of data services linked to data and services provided by research infrastructures at the national and institutional level.

A clear finding of the Landscape report validation workshop was that the information gathered in the analysis must be kept up to date. A dedicated funding line must be created to allow sustainable long-term monitoring of EOSC landscape developments at national and institutional levels. Monitoring of the evolution of national infrastructures and initiatives and the development of respective national policies, supported by a set of relevant key performance indicators (KPIs), is required in order to allow informed decisions on EOSC. The KPIs must be designed, selected and approved with all the major stakeholders as they have a formative effect and influence the development of national environments. KPIs cannot replace the expertise and knowledge of an evaluation/monitoring panel, and the monitoring cannot be reduced to administrative procedures only. All proposed KPIs shall be tested

18 https://www.eoscsecretariat.eu/working-groups/fair-working-group
19 https://www.eoscsecretariat.eu/working-groups/architecture-working-group
20 https://www.eoscsecretariat.eu/funding-opportunities/open-calls/experts-study-legal-regulatory-issues-fair-principles
21 https://repository.eoscsecretariat.eu/index.php/s/QWd7tZ7xSWJsesn
22 https://www.eoscsecretariat.eu/working-groups/landscape-working-group

against the RACER criteria, i.e. they must be relevant, accepted, credible, easy to monitor and robust.

AA10 – Business Models

Viable business models are an essential element of ensuring an operational, scalable and sustainable EOSC federation after 2020. The Sustainability Working Group[23] has taken an iterative approach to identifying business models for EOSC as they are closely coupled with the governance structures and legal entity. The Working Group has documented its progress in a series of reports, beginning with a 'strawman' report[24] in September 2019 on which community feedback was gathered, leading to a 'tinman' report[25], which was completed in December 2019. Analysis of the feedback received on the tinman report prompted the commissioning of a series of targeted studies, starting with the EOSC-Core operational costs.

This study involves the identification of the opportunities presented by and nature of the EOSC ecosystem, use cases and revenue models. Scenarios are being developed in collaboration with stakeholders, related projects and experts to understand cost structures. The first deliverable included a preliminary ecosystem model for EOSC, while the intermediate deliverable expanded the model, building on the initial interviews with service providers and users. This work has highlighted some difficulties in identifying the costs associated with EOSC services because the accounting sytems of the current projects and sources consulted are frequently not organised in a manner that allows them to associate costs to individual services. It is recommended that the next round of projects to be funded via the INFRAEOSC-03-2020 and INFRAEOSC-07-2020 funding calls address this issue in the accounting of services' operational costs. The final deliverable, to be produced before the end of the summer, will include a review of costing models, insights and conclusions on the models, and a cost-model spreadsheet, allowing the Sustainability Working Group to explore scaling scenarios.

The results of this study and others, which will explore business models for the full Minimum Viable EOSC, will be used to develop a third document, referred to as the Iron Lady report, to be published in October 2020.

AA11 – Skills and Training

In order to leverage the potential of EOSC for open and data-intensive research, a key challenge for Europe is to ensure the availability of highly and appropriately skilled people with an excellent knowledge of standards and best practices for delivering, using, sharing and analysing open and FAIR data as well as applications and tools.

Aligned with the new Digital Skills Europe objective[26], EOSC sets out to develop a large talent pool equipped with adequate digital skills embracing a wide range of data-related profiles, but also anticipating skills needed to manage software, code and workflows relating to artifical intelligence and high-performance computing. EOSC is expected to facilitate and

23 https://www.eoscsecretariat.eu/working-groups/sustainability-working-group

24 https://www.eoscsecretariat.eu/file/swg-solutionsforasustainableeeosc00pdf. Please note that you need to log in to the EOSC Liaison Platform in order to access this file. A link and instructions to apply for access to the Liaison Platform are provided on an intermediate page for those unable to log in.

25 https://www.eoscsecretariat.eu/file/solutionsforasustainableeosc-tinmandraft02dec19pdf. As above, please note that you need to log in to the EOSC Liaison Platform in order to access this file.

26 https://ec.europa.eu/digital-single-market/en/policies/digital-skills

constitute an open research labour force of data scientists who have expertise in analytics, statistics, machine learning, data mining and data management; data stewards who have strong domain knowledge and the ability to apply this know-how within organisations to create value; research engineers who are able to develop and embed analytical tools in data-intensive workflows; Open Science specialists and data stewards who provide their expertise to publish research in an open and FAIR way, while taking care of legal / intellectual property rights (IPR) issues and ethical issues.

To realise this vision of a strong digital open research ecosystem with data and software at its core, EOSC needs to launch a concerted effort in education and training to develop and upskill its human resources. Key mechanisms should consider: transparency and recognition of skills and qualifications; equitable and balanced digital research labour market; cross-sector (research–industry–public sector) mobility and employability; digital and Open Science culture; and methodologies for design skills and training strategies that proceed at the same speed as technological developments.

AA12 – Rewards and Recognition

Present-day rewards and recognition (R&R) systems are shaped by government-mandated national and institutional policies and regulations, but they are also stimulated by the competitive environment in which academics and institutions compete for funding and other resources. Many R&R systems currently used by research-performing and research-funding organisations tend to incentivise and reward a narrow range of academic activities – e.g. publishing in journals and attracting external research funding – and rely on a limited and often problematic set of evaluation tools (e.g. simplistic publication metrics such as the journal impact factor and the H-index). This leads to unequal appreciation of the various fields of science and hinders knowledge utilisation and the uptake of Open Science practices[27]. To address this, high-level principles to guide research evaluation were presented in 2015 in the Leiden Manifesto[28], and earlier this year the Declaration on Research Assessment (DORA) initiative summarised five design principles[29] to help institutions experiment with and develop better research assessment practices. The DORA initiative also maintains a curated list of good practice examples of implementation from institutions showing leadership in this area[30].

Generalising, a culture change needs to be realised in order to increase the quality of education, research, impact and leadership. More than a technical issue (e.g. 'better indicators'), a responsible R&R system is also a social issue: a catalyst to foster good research practice and quality in terms of content, openness, scientific integrity and contribution to society. Future evaluation of scientists should have a better balance in valuing achievements in education (if appropiate); research; influence (on science and/or society and/or economy and/or teaching); organisation and leadership. A diversity of career paths should be made possible in order to reward good performance in different areas. Evaluation at different levels

27 Cohen et al. (2019) 'Towards Responsible Research Career Assessment' http://doi.org/10.5281/zenodo.3560479 and Saenen, B., Morais, R., Gaillard, V., and Borrell-Damián, L. (2019) 'Research Assessment in the Transition to Open Science', 2019 EUA Open Science and Access Survey Results https://eua.eu/resources/publications/888:research-assessment-in-the-transition-to-open-science.html.

28 Leiden Manifesto (2015) http://www.leidenmanifesto.org.

29 DORA (2020) 'Rethinking research assessment – Idea for action' https://sfdora.org/wp-content/uploads/2020/05/DORA_IdeasForAction.pdf.

30 DORA (2020) 'Good Practices – Research Institutes' https://sfdora.org/good-practices/research-institutes/.

(institution, department, individual) and for different career stages (from early-stage researchers to full professors) should be based on appropiate criteria. All aspects of Open Science should be stimulated more – for example, the sharing of scientific results with society – and the accessibility of research results should be promoted and rewarded.

AA13 – Communication

EOSC will serve many stakeholders. Nine different groups are distinguished, which for communication purposes can be grouped into three main categories:

1. **Research Service Providers**
   - e-infrastructures, such as PRACE, GÉANT, OpenAIRE, EUDAT, EGI, also referred to as delivering horizontal services.
   - Research infrastructures, such as ESFRIs, also referred to as delivering vertical or thematic services.
   - Data initiatives, such as RDA, offering global platforms for sharing expertise.
   - National research- and (e-)infrastructures.
   - Cloud providers, including commercial parties, offering services to research.
   - Cloud community.
2. **Research Performers**
   - Research communities.
   - Research-performing organisations (e.g. universities, technology institutes, research infrastructures and research-enabling organisations).
3. **Research Funders**
   - Research funders.
   - Policy makers.

Each stakeholder group may have different expectations and perceptions of EOSC. Even if these interpretations differ, they can still be consistent. For example, funders may focus on governance issues, whereas researchers and providers focus on functionalities. Despite these differences, all communications should provide clarity on the why, how and what of EOSC in a consistent way, and at the same time address these different stakeholders.

AA14 – Widening to the Public and Private Sectors

In order to successfully extend the EOSC knowledge ecosystem beyond the core research community, EOSC must demonstrate value and impact that is relevant and meaningful to the diverse groups belonging to broader public and private sectors.

There is potential to widen the circles of EOSC knowledge stakeholders in phases through existing strategic alliances and by means of progressive expansion of knowledge across all categories of stakeholders, starting from the inner circles of EU consortia, public-private partnerships (PPPs), to sector-specific and citizen bodies, and further on to citizen engagement groups.

Several challenges remain in order to achieve the successful widening of the EOSC knowledge ecosystem beyond the core research community. There is a need for a consensus on the management of diverse multi-stakeholder IPR, on the principles of openness and attribution to be used by the wider community, and on the difference between ‘open’ and ‘commons’ systems.

3 Implementation

In Section 2, the fourteen Action Areas (AA) have been presented. For each of these areas an analysis is currently being carried out by the EOSC Executive Board and its Working Groups to determine the status of the area, the gaps that will need to be addressed and the priorities on which action will need to concentrate at different levels (EU, national and institutional) in order to address the gaps and bring the area to the status needed to support the three main EOSC Strategic Objectives.

Since many of the Action Areas contribute to two or three of the Strategic Objectives, they were listed in two groups: Implementation challenges and Boundary conditions.

This section first presents a preliminary set of priorities. Subsequently a non-exhaustive list of outputs/deliverables that are envisioned throughout the different stages of development is given, together with a set of key performance indicators that will allow the future EOSC Association to monitor the implementation of the strategy.

The future final set of priorities will lead to activities that will be implemented by the different stakeholders at different levels. While some of the priorities will need to be translated into actions funded at EU level under Horizon Europe or possible other EU funding programmes, other activities will be designed to foster commitments at national and regional as well as institutional level. It is important to highlight that the evaluation of proposals from the INFRAEOSC-03-2020 and INFRAEOSC-07-2020 calls is ongoing. Although these calls are from the Horizon 2020 programme, it is expected that the projects resulting from these proposals will carry out some of the activities proposed in the following tables.

NOTE: *The numbering of the Priorities is not related to their level of importance or the timeframe for their implementation. They are numbered so that they can be referred to in the accompanying questionnaire.*

3.1. Priorities

Implementation challenges
AA1 – Identifiers ***P1: Implement the EOSC Persistent Identifier (PID) Policy and develop additional infrastructure required to support the publication, curation and tracking of research outputs.*** • Prioritise identifiers for: instruments, services, organisations and software, although there is a need for particular domains to provide their own community standards. • Develop a 'meta resolver' that can deal with any type of relevant identifier. • Define specifications (schemata) for PID records / kernel information to support machine-actionable PIDs. • Produce type definitions for the most common data formats or building blocks. • Provide standardised interfaces and protocols for exchanging information on PIDs to support the creation and use of a PID Graph. • Develop tools to support the certification of PID infrastructures against the EOSC PID Policy.

AA2 – Metadata and Ontologies

P2: Offer a common dataset search to enhance discovery via EOSC.

- Provide or embrace/stimulate existing registries of metadata schemata and ontologies, defining clear protocols for federation/harvesting, crosswalks, and tools for metadata management.
- Develop EOSC guidelines for minimum metadata (e.g., DCAT, DDI 4 Core, DataCite core schema) to allow data discovery and metadata exchange across federated repositories and scientific communities.
- Develop services that build on metadata registries and can facilitate the diffusion of metadata schemata across communities, sharing and community maintenance of crosswalks, measurement of metadata resources uptake across communities, validation of data sources against metadata schemata, etc.

P3: Support communities to develop metadata standards and controlled vocabularies to enable all stakeholders to engage equally in EOSC.

- Support research communities (independently of the current status of their semantic artefact adoption) so as to generate clear and precise definitions for the terms they use, as well as for their metadata and data schemata (and incorporate those they are already using) and their documentation. Provide support to make these definitions publicly available and referenceable by persistent identifiers for machine actionability.
- Dedicate urgent, additional resources (financial, but also skills and training) specifically to communities with less developed or no community standards, to mitigate the risk of EOSC becoming inaccessible to the majority of researchers within academic institutions.
- Support the maintenance of repositories of semantic artefacts, and governance frameworks for such repositories, depending on common practices and the stages of semantic resource development and usage of different communities.

AA3 – FAIR Metrics and Certification

P4: Implement metrics to assess FAIR digital objects and iterate based on testing.

- Support communities to clarify their requirements with respect to FAIR.
- Test the FAIR data maturity model in a wide range of communities, in a neutral forum and seeking international agreement, to fine-tune and customise the recommendations, identify adverse consequences and apply corrections.
- Evaluate and compare assessment tools (manual, automated) and identify their biases and applicability in different contexts.

P5: Support services to demonstrate they enable FAIR via certification or the definition of assessment frameworks.

- Support communities to align certification schemes with FAIR.
- Support data repositories and services to progress towards certification.
- Support data and service providers to progress in the FAIRness of their holdings.
- Support the definition of FAIR for software and of the assessment framework for key elements of the FAIR ecosystem, in particular PID services.
- Define and implement governance of the principles, assessment frameworks and metrics, adapted to each specific case.

AA4 – Authentication and Authorisation Infrastructure

P6: Establish and implement a common framework for managing user identity and access in a highly distributed ecosystem.

- Ensure long-term attribute availability, assurance, freshness and provenance.
- Scale the current proxy (BPA) architecture and supporting infrastructure.
- Address near- and long-term user experience challenges.
- Provide solutions for identity beyond the research and education community in support of public sector and private sector services.
- Enable identity for the individual scientists regardless of institutional affiliation, collaborations and communities while supporting long-term aspects of research.
- Develop future trust fabrics and authorisation models in support of dynamic and ad hoc (on-demand) collaborations.

AA5 – User Environments

P7: Ensure a feedback mechanism to engage with users and further develop the EOSC environment to meet their needs.

Provide the following functionalities:

- Advanced discoverability of portals.
- Meta catalogues to aggregate information from the resource catalogues of the service providers.
- The licences, terms of use, access policies and user authentication and authorisation methods set by service providers.
- Legal and organisational framework and its implementation in the distributed architecture.
- Portals, other richer digital platforms and required supporting components such as distributed data, computing and storage providing necessary capabilities and capacity.
- Interoperability with portals, thematic and regional community services and resources.
- Science gateways for composability of resources from different sources to generate new scientific outputs.

AA6 – Resource Provider Environments

P8: Implement procedures to ensure services that meet requirements can be federated into EOSC easily and efficiently.

- Facilitate onboarding of resources, scaling out EOSC resources through automation, standardisation and validation for EOSC compliance via compliant service catalogues.
- Enable easy, consistent and transparent access to resources by integrating resources with EOSC AAI, aligning terms and conditions and complying with the Rules of Participation.
- Support the composability of resources and help overcome legal, organisational, semantic and technical barriers to allow sustainable resource provisioning, by adapting resources to the EOSC Interoperability Framework.
- Establish a community of practice among the resource providers to accelerate the uptake and adoption of interoperability frameworks and guidelines coming out of EOSC.

AA7 – EOSC Interoperability Framework

P9: Promote the use of open specifications, where available, to ensure technical interoperability when establishing EOSC services.

- Ensure service-level agreements for all EOSC resource providers are easy to understand by users from different communities.
- Provide search tools for coarse-grained and fine-grained datasets (and other research objects). There will be a range of general-purpose and domain-specific/specialised search tools, exploiting general-purpose and domain-specific metadata.

P10: Agree and implement a common set of rules to ensure data and services within EOSC support interoperability.

- Define metadata schemata for service-level agreements.
- Promote standardised, machine-actionable licences to ensure reusability of content.
- Require content providers to publish to agreed formats and/or vocabularies for a specific community.
- Require service providers to ingest or output data according to standardised data formats and/or vocabularies, make available certain APIs or confrom to agreed levels of quality.

Boundary conditions

AA8 – Rules of Participation

P11: Define the cooperation framework enabling RDIs to work together more fully and effectively.

- Define the rules and procedures supporting compliance with the EOSC catalogues, including description templates, onboarding and updating processes, and removal procedures.
- Implement compliance and certification procedures arising from solutions to the implementation challenges such as those related to certification, FAIR, identifiers, interoperability, metadata, training, etc.
- Ensure adoption of the EOSC AAI framework for managing user identity and access.
- Enable monitoring and reporting of service levels through adoption of the EOSC Interoperability Framework.

P12: Evolve EOSC by recognising enhanced standards for policy, processes and procedures to provide increasing levels of assurance of quality and trust in the services offered through EOSC.

- Establish and operate a mechanism to maintain the RoP and monitor compliance with them.
- Develop, recognise and adopt EOSC standards supporting Open Science.
- Support global cooperation between Open Science initiatives.
- Enable access for European researchers to the most effective tools to support their research from public and private sectors.

AA9 – Landscape Monitoring

P13: Ensure continuous monitoring of the existing readiness of countries to contribute to EOSC.

- Monitor standardised national Open Science and FAIR data strategies, including the description of these policies.
- Check the existence of a central/national contact point for Open Science.
- Monitor national policies on open access publishing and open access to publications, and the financial incentives and support schemes.
- Monitor national policies on data and services, and whether their open access to data includes financial incentives and support schemes.
- Monitor national policies on open learning, including financial incentives and support schemes.
- Monitor the national, regional, or sector-level research evaluation schemes of universities and other research-performing organisations, and check whether they include Open Science principles and open access schemes.

P14: Suggest priorities for action based on the monitoring.

- Stimulate progression of the institutional structure(s) at national level that are accountable for defining and implementing EOSC-related policies and strategies, including their hierarchical structure.
- Stimulate EOSC-dedicated funding streams and criteria in national funding mechanisms or programmes.
- Stimulate dedicated funding streams or other measures (programmes, grant schemes, project support, financial and other incentives) that target the promotion and/or implementation of Open Science principles at institutional level.
- Stimulate funding investments and operational costs of infrastructure(s) at national level contributing to EOSC.

AA10 – Business Models

P15: Perform cost assessments.

- Assess cost estimates associated with the EOSC-Core services.
- Assess cost estimates associated with the full Minimum Viable EOSC (MVE).

P16: Ensure sustainable financing for EOSC.

- Develop financing schemes for EOSC.
- Develop monitoring schemes for the in-kind contribution of members.
- Balance the costs of ESFRI and ERIC infrastructures.

AA11 – Skills and Training

P17: Develop Open Science training and professionalise associated roles.

- Develop and support the next generation of data/EOSC professionals through the development and recognition of digital career profiles, the development of quality assurance mechanisms for career progression, and the establishment of networks of experts at cross-disciplinary and cross-national levels.
- Educate and reward researchers at all levels via curricula that meet the demands of open and data-intensive science, and the establishment of advanced learning environments that engage them in a participatory process with society.

- Build capacities to sustain learning corpora for digital skills and tools so that EOSC becomes a trusted and long-lasting knowledge hub for performing open and data-intensive science.
- Develop an EOSC leadership programme to foster the right policy environment that supports digital skills and training at institutional and national level, anticipating demands resulting from technological advancements.

AA12 – Rewards and Recognition

P18: Create a Europe-wide framework for rewards and recognition that includes Open Science.

- Produce a country-level document outlining a contextual, inclusive approach to research evaluation, taking Next Generation metrics into account.
- Discuss the approach within and between (all) the institutions in the country.
- Discuss these developments between the countries and learn from each other.
- Produce, as an outcome of the discussions, guidelines for adapting Rewards and Recognition systems to align with the priorities outlined above.

AA13 – Communication

P19: Inform stakeholders about the developments of EOSC.

- Perform an in-depth stakeholder analysis.
- Set up a Strategic Communication Plan.
- Develop and deploy communication channels.
- Develop stakeholder messaging that is impactful (addressing the why) and functional (addressing the how and the what).
- Set a value statement and carry out an impact analysis.

AA14 – Widening to the Public and Private Sectors

P20: Widen EOSC stakeholder engagement in a strategic and timely manner.

- Incentivise engagement of citizen scientists with EOSC.
- Incentivise mechanisms for value creation by app developer communities.
- Stimulate industrial collaboration projects and the inclusion of SMEs and developers in the design and implementation of specific EOSC software applications and components.
- Align with complementary initiatives such as the Industry Commons[31], grounded in principles of FAIR data.
- Stimulate the formation of cross-disciplinary communities to act as multipliers for the EOSC users.
- Stimulate and reinforce national top-down initiatives for the promotion of research, with bottom-up approaches by diverse citizen scientist and developer communities.
- Promote Open Science success stories as a way to support the widening of EOSC.
- Secure support of Open Science by national governments and funding organisations.

31 Industry Commons is an initiative by the Directorate-General for Research and Innovation funded by a series of Nanotechnologies, Advanced Materials, Biotechnology, and Advanced Manufacturing and Processing (NMBP) calls to establish standards for cross-domain industry collaboration.

3.2. Potential outputs

Stage 1 (2021–2023)

- Common framework for managing user identity and access established.
- FAIR further defined for implementation within the different disciplines.
- Interoperability Frameworks implemented for FAIR sharing within disciplines and for interdisciplinary research to support FAIR digital objects.
- Mechanism established for discovery of and access to data and services across the federated EOSC ecosystem.
- Metrics developed for FAIR digital objects to measure and increase FAIRness.
- Assessment frameworks provided to certify that repository services enable FAIR.
- Efficient and smooth onboarding process for service providers defined.
- Stepwise European alignment of key elements (such as national policies and standards) facilitated, towards ensuring that all digital research objects are open and FAIR by design.
- Clear, consistent and integrated EOSC messaging offered on the benefits and stories, as well as on features and information across channels and stakeholders.
- Training and deployment of professional data stewards supported, together with European curriculum frameworks.
- Recognition and reward of Open Science, data skills and data stewardship among members of EOSC encouraged.
- Online dashboard provided, to present the evolving landscape of policies, infrastructures and initiatives relevant to EOSC.
- Monitoring systems offered that will gather usage data, evidence and best practices valuable to academics, industry, the public sector and the policy makers.

Stage 2 (2024–2025)

- Service providers participating in the EOSC-Exchange conform to predefined Rules of Participation.
- Components of the EOSC architecture assembled (AAI, APIs for access by machines, service management, minimum metadata framework, open metrics, PID services, helpdesk, portal, etc.).
- Standard mechanism adopted for naming and locating data and services.
- Monitoring system established for the uptake, availability and reliability of EOSC-Core services and access to EOSC-Exchange services.
- Feedback mechanism implemented to engage with users and further develop EOSC-Core and EOSC-Exchange tools and services.
- Members of the EOSC Association recognise and reward adoption of FAIR, strongly incentivising data reuse.
- Semantic and other technologies implemented, to facilitate automated processing.
- Governance framework established for managing semantic artefacts.
- Models developed for the long-term sustainability of the EOSC-Core and EOSC-Exchange.
- EOSC ecosystem expanded (e.g. data, software, services, tools) to serve entities from the private sector and public authorities so that they can access and reuse results of publicly funded research, innovating new data services.

Stage 3 (2026–2027)

- EOSC value and impact drive European and global policy dialogue.
- Researchers are incentivised to explore and implement FAIR-by-design practices in their laboratories / observatories / analytical resources.
- The EOSC ecosystem facilitates and provides the capacity for Open Science for the majority of researchers of the EOSC Association members.
- Essential additional functionalities implemented to meet the requirements of end users from the public and private sector.
- Alternative or new infrastructure that enables rewards and incentives for Open Science provided.

3.3. Potential key performance indicators (KPIs)

The following list of KPIs is illustrative, for the purposes of enabling discussion; a final list, together with metrics, will be defined following the consultation process.

***NOTE:** The numbering of the KPIs is not related to their level of importance, but is to enable them to be referred to in the accompanying questionnaire.*

KPI1 Researchers performing publicly funded research make relevant results available, as openly as possible.

KPI2 Professional data stewards available in research-performing organisations in Europe to support Open Science.

KPI3 Researchers are incentivised to perform Open Science.

KPI4 The scope of EOSC is widened to serve the public and private sectors.

KPI5 Research data produced by publicly funded research in Europe is FAIR by design.

KPI6 The EOSC Interoperability Framework supports a wide range of FAIR digital objects including data, software and other research artefacts.

KPI7 European research is increasingly discovered and reused across disciplines as a result of EOSC.

KPI8 EOSC is operational and provides a stable infrastructure, supporting researchers addressing societal challenges.

KPI9 EOSC is populated with a valuable corpus of interoperable data.

KPI10 EOSC is a valuable resource to a wide range of users from the public and private sectors.

4 Expected impacts

Realising the ambitions of the European Open Science Cloud will transform the broader research and innovation ecosystem across multiple dimensions (cultural, technical, organisational, educational, policy) and multiple levels (international, European, national, institutional). Ultimately, when the EOSC ecosystem is fully deployed and when new research datasets, tools and services become systematically FAIR by design, then researchers will be able to deliver research output such as data and software as they deliver publications today. Publications, data and software will be shared as openly as possible and accessed seamlessly so as to optimise reuse and research efficiency.

As presented in the Objectives Tree in Section 2 of this document, the impact these developments will have is major and will lead to the following three key benefits.

Improved trust, quality and productivity in science

By realising the objectives of the European Open Science Cloud, it is expected that a transformation in research culture will take place in which open research becomes the new normal in Europe. This will lead to widely recognised incentives for academics, industry and public services to share their data and other digital research objects, and improve data management training, literacy and data stewardship skills. This will ultimately help researchers to exploit high-quality data in new and efficient ways, whether in the form of managing, sharing, computing, analysing or storing their data. The increased support at all stages of the research data lifecycle will help researchers make maximum use of both their data and time so that they achieve the best possible results.

The opening of the research data lifecycle will also lead to a positive change in the research incentives and rewards systems and structures. As more FAIR data are made available, the possibilities of rewarding researchers, not only for publishing their articles but also for publishing their datasets and analytical tools, increase. This opening, in turn, will lead to a more meaningful monitoring and better reproducibility and validation of research results.

This will also allow researchers to produce more reliable science at the service of society at large, and minimise the impacts of false claims or unsupported statements. Trust in scientific insights will thus increase.

Development of innovative services and products

The implementation of the EOSC ecosystem will enable European research to make its digital transition while ensuring transparency, reproducibility and societal impact. By providing seamless access to increasing volumes of research data, EOSC will stimulate the uptake of different services, from both public and commercial providers, that align with the principles of EOSC. Therefore, the Web of FAIR Data will provide the ideal ground for building a wide range of new innovative and value-added services (from visualisation and analytics to long-term preservation). It will be as transformative as the World Wide Web has been to business and everyday life.

EOSC will be instrumental in stimulating the European private sector, for example, the cloud industry, that is willing to align to these principles while, at the same time, it will ensure that European researchers remain in control of their data, stored in trusted and FAIR-certified

European repositories, and that scientific knowledge will stay as open as possible, as closed as necessary.

Improved impact of research in addressing societal challenges

'*The European Open Science Cloud will make science respond better and faster to what society needs*'
Ursula von der Leyen, president of the European Commission[32]

These words of the European Commission president in the context of the COVID-19 crisis and the development of a COVID-19 data-sharing platform[33] by bringing together infrastructure, resources and data from different sources are a good example of the potential impact that EOSC will have in addressing societal challenges by enabling the improved quality and productivity of research and innovation.

Scientists are increasingly being asked to help develop solutions to the global societal challenges of the 21st century and must employ new and innovative technologies on research data to be able to deal with these complex and interconnected issues. EOSC will lead to a fundamental revolution in the way researchers, companies and administrations share and exploit research data, somewhat similar to the way the internet revolutionised the sharing and exploitation of information. Ultimately, each and every scientist will do research differently from the way it used to be performed. When the ecosystem of new tools and services is available, and as many new FAIR-by-design datasets are generated as possible, researchers will be able to deliver much more rapidly the outputs of each part of the research lifecycle, including data and software, with the same level of precision as they deliver publications today. For research teams and laboratories, publications, data and software will no longer be considered an indivisible whole but as interrelated individual digital objects, in order to optimise the use of research results.

The impact of EOSC on the capacity of research to address current and future global challenges will be based on the following elements:

- **More multi-disciplinary research through data sharing and cross-usage of services.** For each discipline, the communication and collaboration with scientists in other disciplines will benefit from faster and seamless sharing of publications, data, software, services, tools and other digital research outputs.
- Better and faster sharing of data and results will **strengthen collaboration among researchers and disciplines** and present opportunities for new levels of integration. For the cooperation between teams to address multi-disciplinary challenges, the use of data and software across research silos will allow the exploration of new avenues to an extent that has never been possible before.
- **Increased added value of the services** in the EOSC ecosystem. A European-scale environment for computational, storage, analysis and other data-related services and tools will facilitate multi-disciplinary cooperation, leading to discoveries and solutions in key areas such as environment and health.

32 https://ec.europa.eu/commission/presscorner/detail/en/ip_20_680
33 https://www.covid19dataportal.org/

ANNEX – Glossary

- **Data.** An encompassing term used in the EOSC context for all digital outputs of research including datasets, metadata, publications and software code.
- **Data Infrastructure.** An (inter)national or institutional infrastructure that stores, handles and provides a level of access to (possibly FAIR and open) research data.
- **e-Infrastructure.** An (inter)national or institutional infrastructure that enables research through technical hardware and digital services (such as storing, computing or connecting) for sharing and exploiting research data.
- **EOSC Association**. International Non-Profit Association (AISBL) currently being founded in Brussels to represent those (eligible) stakeholders wishing to formalise their role in EOSC. The Association intends to sign a Memorandum of Understanding (MoU) with the European Commission and thus form a European Partnership.
- **EOSC-Core.** The basic architecture, standards and services that form the technical backbone of EOSC and are necessary to operate a Web of FAIR Data and Services.
- **EOSC-Exchange.** The value-added services that will build upon the EOSC-Core and offer its users additional functionality to perform Open Science and share and exploit FAIR (and open) data.
- **EOSC Ecosystem.** The encompassing set of federated (e-)infrastructures, research infrastructures, stakeholder organisations and projects that contribute to and/or use EOSC.
- **EOSC Governing Bodies.** The current interim EOSC governance structure, comprising the Governance Board and Executive Board, whose mandate will end 31 December 2020, and thereafter the governance structure of the EOSC Association.
- **EOSC Partnership.** The co-programmed European Partnership between the EOSC Association and the European Commission that will consolidate the outputs of EOSC projects from Horizon 2020 and further develop EOSC through structured funding in Horizon Europe and in-kind contributions from the member countries and stakeholders.
- **EOSC Partnership Proposal.** The published proposal for a co-programmed EOSC Partnership[34].
- **European Open Science Cloud (EOSC).** The generic term for the envisioned federation of research (data) infrastructures that will enable the Web of FAIR Data and Services and help researchers to perform Open Science and open up and exploit their data, publications and code.

34 https://ec.europa.eu/info/files/european-open-science-cloud-eosc_en

- **FAIR (Principles).** The set of guidelines for making research (meta)data findable, accessible, interoperable and reusable that ensures standardised machine actionability[35].
- **Horizon Europe.** The European Commission's ninth funding framework programme for research and innovation, which will run from 2021–2027.
- **Minimum Viable EOSC (MVE).** The EOSC-Core plus selected services from the EOSC-Exchange that provide researchers with the minimum level of functionality required to share and exploit FAIR (and open) data.
- **Open Data.** Data in an open format that can be freely used, reused and shared by anyone for any purpose.
- **Research Infrastructure.** An (inter)national or institutional infrastructure that enables research communities to perform research.
- **Strategic Research and Innovation Agenda (SRIA).** A set of recommendations from EOSC stakeholders, edited by the EOSC Association, which provides general guidelines in discussions between the EOSC Association and the European Commission in the context of the EOSC Partnership to help develop the work programmes for EOSC in Horizon Europe.
- **Web of FAIR Data and (Related) Services (for Science).** The network of connected FAIR (and possibly open) datasets and the services that researchers need to exploit these datasets for their research that are brought together and offered through EOSC.

35 https://www.nature.com/articles/sdata201618

EUROPEAN OPEN SCIENCE CLOUD

eoscsecretariat.eu

@EoscSecretariat

/company/eoscsecretariat

EOSCSecretariat.eu received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019, Grant Agreement **No. 831644.**