Strategic Research and Innovation Agenda (SRIA)

of the
European Open Science Cloud (EOSC)

Version 1.3 – 1 November 2024
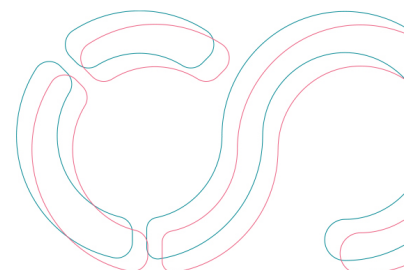
EUROPEAN PARTNERSHIP

Please note

This SRIA 1.3 was created to incorporate the MAR 2025-2027 into SRIA 1.2 and, hence, is **not** a full update of the EOSC Strategic Research and Innovation Agenda.

eosc

Table of Content

Table of Figures

Table of Tables

Foreword by Director-General Jean-Eric Paquet, Directorate-General for 'Research and Innovation'[1]

The Commission Communication on a new ERA for Research and Innovation [EC_COM_New_ERA] stresses the importance of Open Science as the most efficient and effective way of carrying out research to increase knowledge circulation, open collaborative work and the wide and rapid sharing of research outputs. The ultimate objective of Open Science is to increase scientific quality, the pace of discovery and technological development, as well as societal trust in science. The Communication highlights, specifically, the role of the European Open Science Cloud (EOSC) [EOSC] to drive a well-functioning and high-performing European R&I ecosystem by fostering the flow of research data and scientific knowledge between researchers, institutions and disciplines.

The challenges that our society faces, such as the COVID-19 pandemic and the climate emergency, are increasingly complex, cross-border and inter-sectoral in nature. The research needed to address them is often data intensive, spanning many disciplines and countries. Long gone is the era of the lonely scientist. Now is the time for collaborative science, which is open, trusted and digitally enabled. The digital outputs of research (publications, data or software, among others) must be shared widely and as early as possible between researchers, but also between communities and with society. This requires that these outputs are made FAIR (findable, accessible, interoperable and reusable) [GO_FAIR_Principles] and that the related data services are available. This is what EOSC is about: developing, deploying and evolving a trusted environment providing two million European researchers with seamless access to research data, research infrastructures, e-infrastructures and related services, enabling them to share, curate, discover, access, process and reuse research outputs of all kinds across borders and scientific disciplines.

This Strategic Research and Innovation Agenda (SRIA) provides a clear roadmap over the next seven years to achieve the EOSC vision and objectives. It results from a collective, forward-looking co-creation process to identify and prioritise complementary activities at EU, national, and institutional levels. Its content has been validated through wide consultation of EOSC stakeholders, including representatives of EU Member States and other countries associated to the EU R&I programme, research-performing and research-funding organisations, research infrastructures and e-infrastructures, research libraries and research associations.

I want to warmly congratulate the members of the EOSC Executive Board and the many experts who worked hard to drive the SRIA development. Thanks to their remarkable professionalism and dedication, their efforts have been successful.

I look forward to a fully operational EOSC that delivers the 'Web of FAIR Data and Services for Science'. This will be essential to affect the digital transition in the European Research Area, enabling data-intensive interdisciplinary research and innovation, and accelerating the process of scientific discovery and technological development.

---

[1] Mr. Paquet served as Director-General of DG RTD from 1 April 2018 until 1 September 2022.
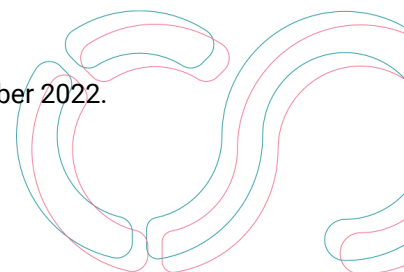
Preface by Karel Luyben, EOSC Association President (2020-2025)

This Strategic Research and Innovation Agenda (SRIA) is the latest in a series of increasingly significant steps towards making the European Open Science Cloud (EOSC) a reality. A collaborative effort, the development of the SRIA is an example of the multi-stakeholder commitment, cooperation and consensus that are essential to realising EOSC's vision and potential.

EOSC was conceived as the solution to the problem of how to manage and exploit the unprecedented volume of data arising from digital technologies, in the interest of European science and scientists. By providing an open and trusted environment for accessing and managing a wide range of publicly funded research data and related services and complementary commercial services, EOSC will transform how researchers access and share data throughout the research lifecycle, helping European scientists reap the full benefits of data-driven science and giving Europe a global lead in both research data management and scientific progress. It will federate existing data- and e-infrastructures, currently dispersed across disciplines and the EU Member States, around a federating core, with frameworks, principles and rules – many of them based on already established initiatives – to ensure the data are findable, accessible, interoperable and reusable (FAIR), while respecting data sovereignty, security and data protection requirements and regulations. Its role is to facilitate and optimise the sharing of existing data, digital objects, services and resources, not to reinvent; it is driven by the needs of science, scientists and the wider stakeholder community, not by technological advancement for its own sake.

To date, multiple projects funded by Horizon 2020 calls have helped to lay foundations for EOSC, engaging a wide range of stakeholders and communities and delivering relevant outputs including use cases, demonstrations, data service tools, policy documents and, most visibly, the EOSC Portal. Progress towards enabling the creation of the EOSC ecosystem has thus been significant – yet fragmented. Beginning with the transition governance structure and culminating in the creation of the EOSC Association, steps have been taken to overcome that fragmentation in the future, to provide a framework for consolidation, coordination and collaboration, and to bring the community together at European, national, regional and institutional levels.

As EOSC enters the next implementation phase under Horizon Europe, the commitment, cooperation and consensus of all stakeholders are key, from researchers through service providers and infrastructure operators to funders and policy makers. The EOSC Association – with, at the time of writing, 142 members (21 of them nationally mandated organisations) and 49 observers representing the full spectrum of stakeholders – is well placed to deliver that commitment, cooperation and consensus, while its Co-programmed European Partnership with the European Commission provides the optimum vehicle for realising the vision outlined in this SRIA.

The challenges involved are immense, and are acknowledged throughout the SRIA, not least in the critical success factors and 14 action areas it enumerates. But the societal benefits EOSC can deliver are equally immense – nothing less than furthering the ability of researchers to solve such global problems as the coronavirus pandemic and climate change. In keeping with that global scale, and with the global nature of the ecosystem within which EOSC operates, the SRIA recognises the importance of collaboration with the rest of the world, of integrating with the Open Science initiatives in other regions. Digital technologies have

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu   - 7 –
Reg. number: 0755 723 931 | VAT number: BE0755 723 931
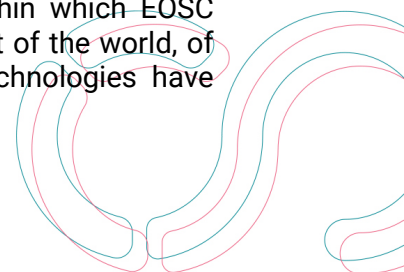
brought all parts of the world closer, while the effect on the environment of anthropogenic activities similarly impacts all parts of the world; we are in this together.

In delivering a pan-European infrastructure for Open Science, EOSC has the potential to be as much of a game-changer for the sharing and exploitation of data as the World Wide Web has become. For despite its name its reach is global, not just confined to Europe, and even though 'science' refers to all branches of knowledge and areas of study and research, its scope extends beyond science to the public and private sector, and to society at large.

I firmly believe that all this is achievable. In my role as Chair of the EOSC Executive Board, I have had the great privilege of working with a community that is as vibrant as it is diverse. Looking back on what we have achieved in Europe by working together makes me proud to be part of this EOSC community and confident of our ability to meet the challenges and grasp the opportunities ahead.

I should like to thank all the contributors to this SRIA: the EC, the EOSC Executive Board, EOSC Governance Board, EOSC Secretariat and especially the Working Groups and Task Forces.

In my role as President of the EOSC Association, I enjoy working with the EOSC community to help realise the vision and objectives of EOSC.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 8 –
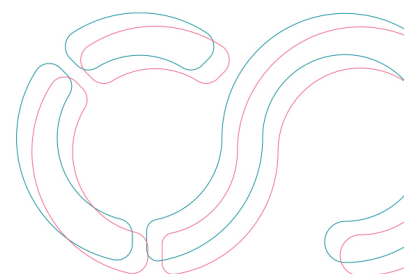
Executive summary

The overall purpose of this Strategic Research and Innovation Agenda (SRIA) is to define the general framework for future strategic research, development and innovation activities in relation to the European Open Science Cloud (EOSC). This framework will be developed and further defined in the context of the EOSC Partnership proposed under the Horizon Europe programme.

Its intended audience comprises the individuals and organisations interested or involved in EOSC, or impacted by it, both now and within the timeframe of Horizon Europe, including research-performing organisations, research-funding organisations, service providers, governmental organisations, companies/businesses and citizens, as well as the European Commission.

The SRIA begins by describing the developments leading to conditions favourable for creating EOSC. It outlines the history of the digitisation of research (Section 1), and the EC/EU policy context for Open Science and Open Data, together with the development and governance structures of the EOSC initiative (Section 2). The SRIA goes on to explain the three overarching objectives that are driving EOSC (Section 3), the guiding principles that are shaping it (Section 4), and the challenges and prerequisites to implementing the EOSC ecosystem (Sections 5 and 6). It discusses the anticipated benefits of EOSC, and the critical success factors (Section 7), before presenting a roadmap, priorities and key performance indicators for the near future, in this SRIA v1.2 for Stage 1 (period 2021–2022), Stage 2 (period 2023–2024) and Stage 3 (period 2025–2027) (Section 8). The SRIA ends by drawing together the main points and conclusions (Section 9). Summaries of documents related to the SRIA are provided in Appendix A.

The digital age, the most recent stage in an evolving continuum of ways in which technology has supported science, presents an opportunity to improve the conduct of research in multiple directions, including with regard to openness, speed of access to scientific results, reproducibility and multi-disciplinarity. This should result in better science, increased trust in science, and an improved ability to meet global challenges. However, this potential will only be realised if research infrastructures evolve to allow scientists to exploit, in an easy-to-use and integrated environment, the (vast amounts of) relevant data being produced.

EOSC will help deliver Europe's contribution to the realisation of scientists', and science's, potential in the digital age, enhancing Europe's leadership position in exploiting digital capabilities at the service of science. It is an integral part of, and supports, the European Commission's strategy for realising and revitalising the European Research Area (ERA). In particular, it helps deliver the policy priorities of Open Innovation, Open Science and Open to the World and the goal of findable, accessible, interoperable and reusable (FAIR) data. It also contributes to the six priorities driving the EC's work programme for 2019 to 2024, and to the specific objectives of the Horizon Europe framework programme.

Three overarching objectives form the driving force in building EOSC and developing the EOSC ecosystem. These are laid out in a matrix, known as the EOSC Objectives Tree (Figure 0.1.1), which also identifies the main problems, barriers and benefits.
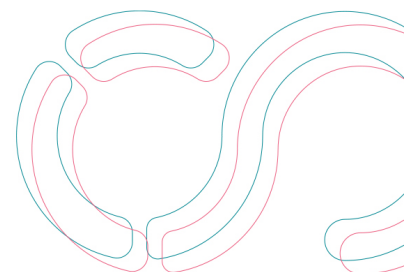
European Open Science Cloud Objectives Tree
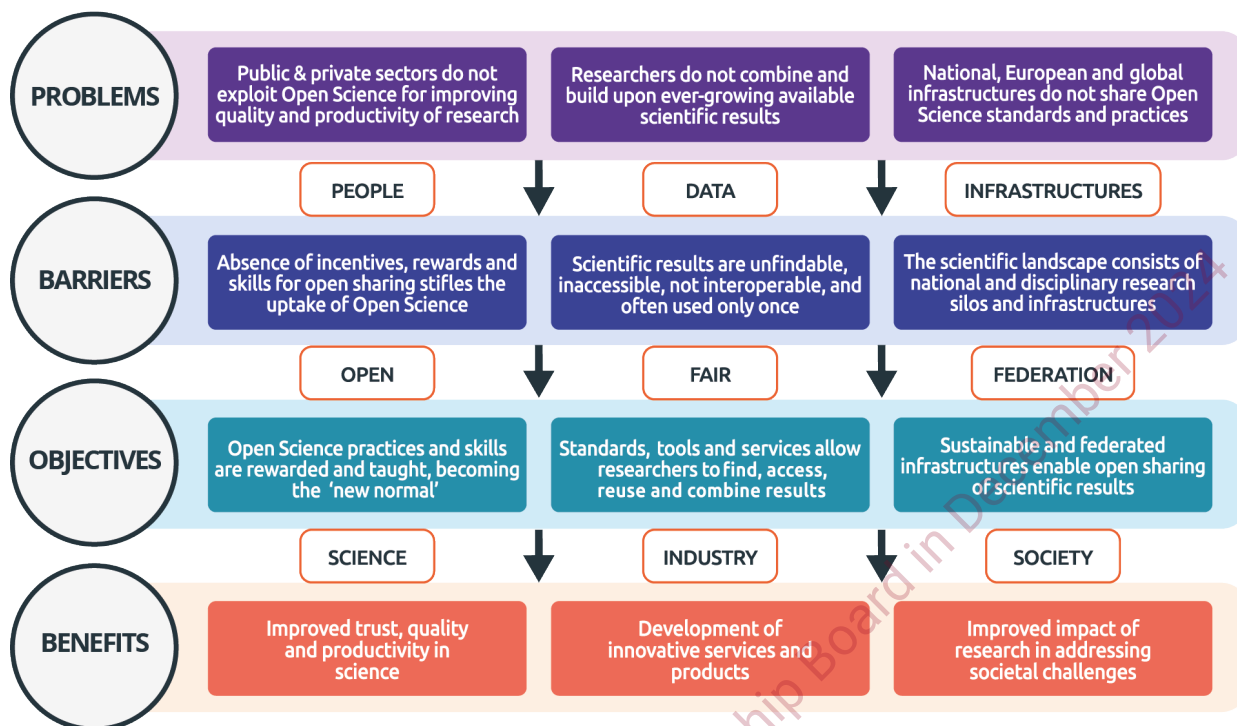
Figure 0.1.1: European Open Science Cloud Objectives Tree

In addition, a set of guiding principles has been agreed which will help position EOSC within Horizon Europe and shape its development:

- **Research-community centred**: EOSC will place research at the centre of the initiative and will thus prioritise engagement with research communities to understand their requirements, helping them and ensuring academic sovereignty of research data.
- **Multi-stakeholderism**: EOSC will succeed if and only if it follows a multi-stakeholder approach;
- **Openness**: EOSC will ensure that research artefacts are 'as open as possible, as closed as necessary'[2];
- **FAIR principles**: EOSC will assemble research artefacts that are findable, accessible, interoperable and reusable;
- **Federation of infrastructures**: EOSC will federate existing and upcoming research infrastructures;
- **Machine-actionable**: EOSC will strike the right balance between machines and people in delivering the services that will serve the needs of European scientists.

Building on the guiding principles, a number of recommendations for research communities and policy makers have been identified, to help them progress towards an Open Science ecosystem that is based on, incentivises and facilitates Open Science principles and practices in performing and sharing science. Research communities should:

- Normalise their Open Science processes (standards);
- Regulate them (policies);
- Facilitate their implementation (guidelines and frameworks, e.g. information models that describe flows and elements);

---

[2] For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".
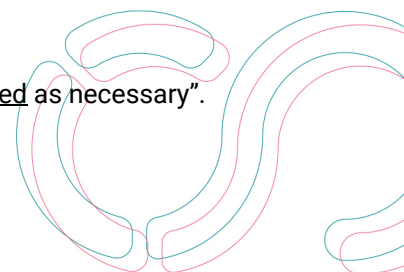
- Make sure their thematic services embed Open Science aspects by design (roadmaps).

Together with those identified by the EOSC FAIR Working Group, these recommendations have provided a basis for determining the action areas that will help deploy the EOSC ecosystem and be part of the EOSC programme over the Horizon Europe period. For each action area – 14 in total – the status has been assessed, gaps identified and priorities proposed; the priorities, in turn, have helped determine the roadmap and key performance indicators for the near future. The seven areas relating to the primarily technical challenges and prerequisites to implementing the EOSC ecosystem – referred to as the implementation challenges – are as follows:

- **Identifiers**. The persistence of the identity of digital objects and stability of references to those objects are essential to sustaining a trusted distributed research ecosystem that supports verifiable and reusable research.
- **Metadata and ontologies**. An overarching, coordinated approach is required, forming the basis for interoperability.
- **FAIR metrics and certification**. 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 progressive, and take the specific research context and needs into account.
- **Authentication and authorisation infrastructure (AAI)**. The purpose of AAI in EOSC is to allow identified scientists to (re)use identifiable documents, data and software, and exploit identified services, while enabling high-trust collaborations to be established and maintained with little or no friction to the end user. The goal is to build a foundation for AAI that will ensure long-term availability of the aspects of digital identity that are unique to scientific collaborations.
- **User environments**. Throughout the distributed, federated and clustered architecture of the EOSC ecosystem, the user environments must meet the users' requirements and expectations, particularly with regard to discovery and composability of resources.
- **Resource provider environments**. As a federation built out of many independent organisations and resource providers – a system of systems – EOSC should be inclusive rather than selective. The added value of EOSC exists only when as many as possible of the resource providers serving the scientific community can enter and offer resources.
- **EOSC Interoperability Framework**. Achieving a good level of technical, semantic, organisational and legal interoperability within EOSC is essential to federate services and provide added value for users, across disciplines, countries and sectors.

The seven action areas relating to the social, financial, legal, educational, cultural challenges and prerequisites to implementing the EOSC ecosystem – referred to as the boundary conditions – are as follows:

- **Rules of Participation**. A process of change in the research environment is required to adopt Open Science practices, make digital research objects FAIR and federate research data infrastructures. The Rules of Participation (RoP) provide transparent and consistent terms for involvement in EOSC, helping to build the trust and confidence required to support and deliver this process of change.
- **Landscape monitoring**. Sustainable long-term monitoring of EOSC landscape developments – the infrastructures, initiatives, investments, and policies at national and institutional levels – is required to allow informed decisions on EOSC.
- **Funding models**. Viable funding models are an essential element of ensuring an operational, scalable and sustainable EOSC ecosystem.
- **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

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 11 –
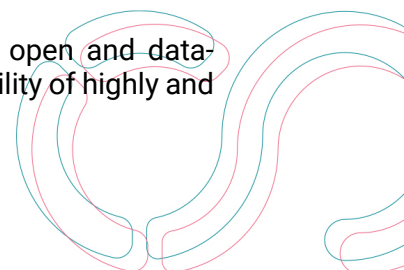
appropriately skilled people. The vision of a strong EOSC ecosystem that exploits digital technologies and has data and software at its core necessitates a comprehensive skills and education strategy.

- **Rewards and recognition**. A culture change needs to be realised in the way scientists are appraised, by looking at their broader contribution to education, research, impact and leadership. A responsible rewards and recognition system is a catalyst to foster good research practice and quality in terms of content, openness, scientific integrity and contribution to society.
- **Communication**. EOSC's diversity of stakeholders requires a communication policy that meets the different needs of each group, providing clarity on the why, how and what of EOSC, and sending out its messages in a consistent way.
- **Widening to public and private sectors and going global**. To successfully extend the EOSC ecosystem beyond the core research community, EOSC must demonstrate value and impact that is relevant and meaningful to the diverse groups belonging to the broader public sector and to the private sector. In parallel, there is clearly a global dimension to EOSC, a common vision that enables Europe to enhance scientific collaboration with other parts of the world and drive a cultural change towards Open Science, which EOSC must respect and exploit to maximise its potential impacts.

EOSC aims to be a major European vehicle for helping transform individual research efforts into collective efforts to face global challenges such as the climate crisis, the extinction of species, global poverty and social inequality. Referring again to the Objectives Tree (Figure 0.1.1), the anticipated benefits of EOSC in the areas of Science, Industry and Society are as follows:

- **Improved trust, quality and productivity in science**, by:
  o **Encouraging collaboration and openness**, potentially changing the way research is performed and enabling new mechanisms for communication and evaluation of research.
  o **Trusted frameworks for data availability and security**, unlocking the full value of research data.
  o **Infrastructure planning**, reducing the disparities in Open Science readiness in different countries and improving the quality of the integrated research landscape.
  o **Broadening discoverability** within and across domains, encouraging multi-disciplinary research.
  o **Making new connections**, maximising the potential of machine searching through standardised metadata and persistent identifiers.
  o **Addressing global challenges** by facilitating cross-disciplinary research through the federation and interoperability of data and research infrastructures.
  o **Enhancing reproducibility** by opening up research processes and outputs and providing a shared, sustained and stable infrastructure for research.

- **Development of innovative services and products**, such as:
  o **Opportunities to improve support for researchers**, widening opportunities for collaboration by enabling pan-European access to data and services, facilitating rapid delivery and reuse of results at each stage of the research lifecycle, and fostering the development of value-added services.
  o **Opportunities to improve support for the private and public sector** by extending the additional functionalities, services and open data EOSC provides beyond the research community, bringing more actors and investments into the research and innovation process.
  o Opportunities to improve European leadership and collaboration in a global setting by positioning EOSC as the European component of a Global Open Research Commons.

- **Improved impact of research in addressing societal challenges**, by:
  - o Increased awareness of and trust in scientific **research in society**, and a higher level of interdisciplinarity and scientific evidence in decision making, planning and strategy at societal level, increasing the return on public investment in science.
  - o **Supporting international collaboration** by ensuring that relevant data are rapidly discoverable, accessible and reusable and value-added services and tools are available.
  - o **Lifting science beyond the human scale** by facilitating the deployment of smart algorithms, machine learning and AI services onto the Web of FAIR Data.

A number of critical success factors must be in place for those benefits to be realised. They include:

- Researchers performing publicly funded research make relevant results available, as openly as possible;
- Professional data stewards are available in research-performing organisations in Europe to support Open Science;
- Researchers are incentivised to perform Open Science;
- The scope of EOSC is widened to serve the public and private sectors;
- Research data produced by publicly funded research in Europe is FAIR by design;
- The EOSC Interoperability Framework supports a wide range of FAIR digital objects including data, software and other research artefacts;
- EOSC is operational and provides a stable infrastructure, supporting researchers addressing societal challenges;
- EOSC is populated with a valuable corpus of interoperable data;
- EOSC is a valuable and valued resource to a wide range of users from the research and education, public and private sectors (including for-profit).

Some of these prerequisites have been translated into activities as part of the first iteration of a roadmap for delivering a fit-for-purpose, sustainable EOSC, in furtherance of the jointly defined objectives of the Co-programmed European Partnership on EOSC. The three implementation stages are aligned with Horizon Europe funding periods, while the proposed activities are aligned with the three overarching objectives of EOSC, dovetail with existing deliverables of the Horizon 2020 work programme, and, in their prioritisation and timing, reflect community feedback from the open consultation process. The work plan outlined in SRIA 1.0 focused on the first implementation stage, 2021–2022, which aimed to develop added value from a federation of infrastructures by providing the core functions of the Minimum Viable EOSC (MVE) that should enable EOSC operations (the EOSC-Core). For each strategic objective a set of priority activities is defined, together with the most appropriate levels of implementation (European, national, institutional) and expected outcomes. Key performance indicators (KPIs) are also defined, with tentative target levels of accomplishment and dates, and measurement method (direct, survey). The work plan for Stage 1 will set the foundations for the subsequent roadmaps for Stage 2 (period 2023–2024) and Stage 3 period 2025–2027). Ultimately, EOSC will deliver a research environment that promotes Open Science and increases trust and reproducibility in research outcomes. The overall impact is a pan-European research landscape that offers significantly improved discovery, access, interoperability, and exploitation of research outputs for researchers and research and innovation stakeholders.

In conclusion, EOSC is well placed to deliver Europe's deployment of Open Science within the Horizon Europe work programme and beyond. At this moment (1 November 2023) the EOSC Association, with 171 members (including 28 nationally mandated organisations) and 87 observers, involves research and innovation stakeholders both across and outside of the EU. The Association provides a sustainable means, recognised by the EC, to serve and strengthen the EOSC community, coordinate the identification of needs for the development of EOSC,

promote alignment of EOSC contributions at all levels and support Open Science development in Europe.

EOSC Association AISBL
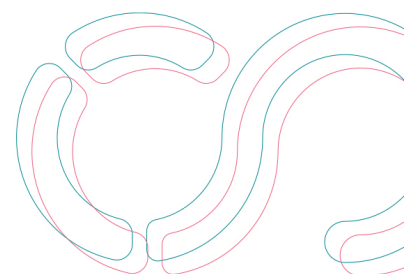
How to read this document

This Strategic Research and Innovation Agenda (SRIA) contains a wide range of information – both historical and forward-looking, high level and detailed, aspirational and practical – and can therefore be read either in its entirety or in part, depending on the interests and/or needs of the reader. Suggestions for approaching the document are summarised in Table 0.1. The sections that relate most closely to future research, development and innovation activities in respect of the European Open Science Cloud (EOSC) are **highlighted**.

If you are interested in . . .	Then please see . . .
A **standalone summary** of the document	Executive summary
The **history of the digitisation of research** – the science-supporting technological context from which EOSC has evolved	Section 1 New ways of science
The **recent EC/EU/ERA policy context for Open Science and Open Data** and the **development of the EOSC initiative and its governing bodies**	Section 2 EOSC in the making
The **strategic objectives** that are driving EOSC	**Section 3 Strategic objectives of the European Open Science Cloud**
The **guiding principles and recommendations** that are shaping EOSC	**Section 4 Guiding principles**
The primarily **technical challenges and prerequisites to implementing the EOSC ecosystem**	**Section 5 Implementation challenges**
The **social, financial, legal, educational, cultural challenges and prerequisites to implementing the EOSC ecosystem**	**Section 6 Boundary conditions**
The anticipated **benefits of EOSC and critical success factors**	Section 7 Expected impacts
The **priorities for EOSC in the first (period 2021-2022), the second Stage (period 2023-2024) and the third stage (2025-2027) of Horizon Europe and key performance indicators**	**Section 8 Roadmap**
A **recap of the main points and conclusions**	Section 9 Conclusions
Summaries of **related documents** that, together with this SRIA, define the structure, aims and work of EOSC	Appendix A

Table 0.1: How to read this document
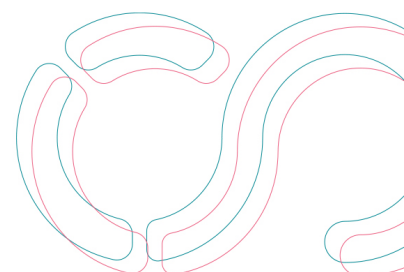
A note on Terminology

Readers are reminded that throughout the document, 'data' is used as an encompassing term referring to all digital research artefacts, including datasets, metadata, publications, intermediate results, workflows, notebooks and software code. Similarly, 'science' refers to all branches of knowledge and areas of study and research, *including* arts and humanities subjects rather than in contradistinction to them, while 'scientist' refers to all researchers, academics and practitioners in all domains.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 16 –
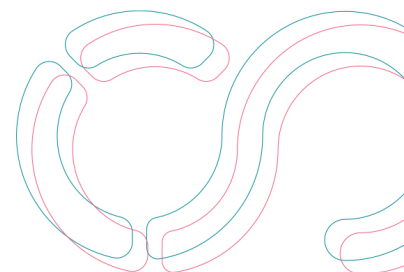
1 New ways of science

The current, digital age is the most recent stage in an evolving continuum of ways in which technology has supported and enhanced science. This section outlines the history of the digitisation of research, establishing the technological context from which the European Open Science Cloud (EOSC) has evolved. It includes lessons to be learned, and developments from which EOSC will benefit and to which it will contribute.

1.1. The opportunity for change

1.1.1. Research in the digital age

In the digital age, the world has become **instrumented, interconnected and intelligent**. **Instrumented** refers to the fact that digital information is now collected everywhere on the planet using small devices as well as large equipment. **Interconnected** refers to the fact that digital information produced anywhere on the planet can be made available anywhere else. **Intelligent** refers to the fact that people and machines can then process this information for the benefit of society at large.

In a world that is instrumented, interconnected and intelligent, human activities can be improved by discovering, retrieving, analysing, assembling and computing information in order to extract the knowledge necessary to address challenges at all scales.

Among all human activities, **research** plays an enabling role by producing scientific results that can be exploited by society to address global as well as local problems. Scientific results include publications, data, software and any research artefacts or intermediary results produced during the research lifecycle.

The digital age allows the ways research is conducted to change in multiple directions, resulting in better science, increased trust in science, and the ability to meet global challenges.

- Scientists will be able to do better research by getting early (sometimes real-time) access to scientific results, optimising their own work. Disciplines organised around large shared equipment already provide examples of the benefits of sharing information across the globe.
- In a world that is becoming more and more complex, the availability of multiple information sources will allow trust in scientific results to be strengthened by facilitating reproduction of scientific experiments and comparison of outcomes. Trust in science has to become the foundation of the new societal paradigm if Europe wants to maintain and develop its way of life.
- Scientists will be able to engage in multi-disciplinary initiatives to address the key global challenges of the twenty-first century such as climate change, health, food and biodiversity or building energy-efficient vehicles and smarter cities. More generally, all efforts dedicated to achieving the 17 Sustainable Development Goals of the United Nations [UN_SDG] would benefit from access to a wide set of information coming from very different origins.

However, while an instrumented, interconnected and intelligent world has unprecedented potential to solve the key challenges of the time, this potential will only be realised if research infrastructures are evolving to allow scientists to make the best use of the available information.

The European Open Science Cloud (EOSC) will deliver Europe's contribution to enabling scientists to realise their potential in the digital age.
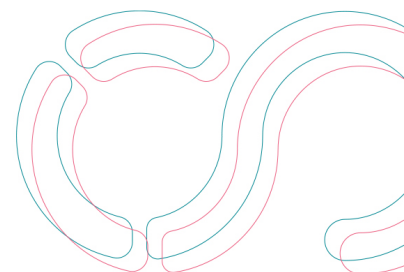
1.1.2. European leadership

When the Horizon Europe programme began in 2021 [Horizon_Europe], Europe has been well placed to lead the world in exploiting digital capabilities at the service of science. After three years of preparation, the initial phase of EOSC was launched in November 2018. More than 35 research and innovation projects have developed foundational technologies and initial services on which Europe can now build. These efforts have also allowed the establishment of a Europe-wide community that is ready to engage further. While other regions in the world have launched their own efforts, none of them have done it at the scale on which Europe has invested.

Pursuing the effort to get EOSC operational as part of the Horizon Europe programme will enhance Europe's leadership position. Through coordination and concentration European research and innovation investments will be more efficient, will be able to address key global challenges and will strengthen the trust in science that society needs to build a common future.

1.2. The request for change

1.2.1. From Gutenberg to Berners-Lee

The current way of sharing research was built upon the emergence of the printing process. During the seventeenth century, the first research journals were conceived by academies of sciences. *The Philosophical Transactions of the Royal Society* was launched in 1665 and received, over the years, articles from scientists such as Newton, Faraday and Darwin. Since then, the publications process has developed in volume but the principles of their use have remained largely identical. Articles are published in journals. Journals are acquired by libraries. Scientists visit their libraries to access the knowledge delivered by their predecessors and colleagues.

The digital age has the potential to revolutionise communication between scientists. While peer-reviewed publications remain the 'official' way to deliver conclusions (potentially using the internet for faster dissemination and transitioning to an open access business model), many other types of information can be made available, increasing the bandwidth of knowledge sharing. Data, software, intermediate results, workflows and notebooks are often stored in digital form. It is up to the scientists and/or the organisations they work for and/or the organisations that fund their research to decide whether this information should be shared, and how widely. Early and open accessibility of such digital assets form a large part of the transition towards what is now called Open Science.

Many researchers will recognise that Open Science is improving science as a whole. However, to date only a subset is convinced that the opportunities it affords to them individually are greater than the drawbacks, while many leaders within the research community argue that data, software and other research artefacts should be kept closed as they are assets that research teams ought to keep for themselves if they want to stay competitive. To change this, establishing a new paradigm for rewards and recognition is essential: it can no longer be based on publications alone.

1.2.2. Lindau Declaration

Once every year, around 30 to 40 Nobel laureates convene in Lindau, Germany, to meet the next generation of leading scientists: 600 undergraduates, PhD students, and post-doc researchers from all over the world. The Lindau Nobel Laureate Meetings foster exchanges among scientists from different generations, cultures and disciplines.

Elizabeth Blackburn is a 2009 Nobel Laureate in Physiology or Medicine for her work in molecular biology. During the 68th Lindau Meeting in 2018, she introduced ten goals for science which subsequently became the core of a 2020 Lindau Declaration [Lindau_Dec]:

- Adopt an ethical code;
- Cooperate globally on global problems;
- Share knowledge;
- Publish results Open Access;
- Publish data in repositories;
- Work transparently and truthfully;
- Change reward system;
- Support talent worldwide;
- Communicate to society;
- Engage in education.

Since its original proposal, the Declaration has been open for debate, changes and amendments. The appeal aims to get widespread support for a new approach to global, sustainable, cooperative Open Science. "Around the world, the Open Science movement continues to gain support from a growing number of researchers and members of the public – a scientific meta topic also represented in the Lindau Meeting programme in 2021."[3]

This exemplary initiative illustrates the current status of Open Science. Thought leaders have understood the potential of the digital age, the impacts on the ways to do research and the benefits for society at large. The request for change now comes from the pioneering research community at its most talented level. The fact that developing such a Declaration is needed also shows that strong initiatives have to be taken in order to fulfil the potential and overcome the caution or conservatism of other members of the research communities.

1.3. Open Science

Assembling different contributions, Wikipedia defines Open Science as 'the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of an inquiring society, amateur or professional'. It continues: 'Open science is transparent and accessible knowledge that is shared and developed through collaborative networks. It encompasses practices such as publishing open research, campaigning for open access, encouraging scientists to practice open-notebook science, and generally making it easier to publish and communicate scientific knowledge.' [Wikipedia_OS] The resulting research activity flows are shown in Figure 1.1.

Another definition of Open Science is provided by the FOSTER portal: 'Open Science is about **extending the principles of openness to the whole research cycle** […], fostering sharing and collaboration as early as possible thus entailing a systemic change to the way science and research is done.' [FOSTER_OS]

---

[3] Quote from the Annual Report 2021 of the Lindau Nobel Laureate Meetings, available at: https://www.lindau-nobel.org/wp-content/uploads/2021/11/2021_Annual-Report_Web.pdf
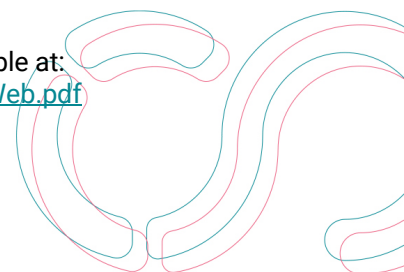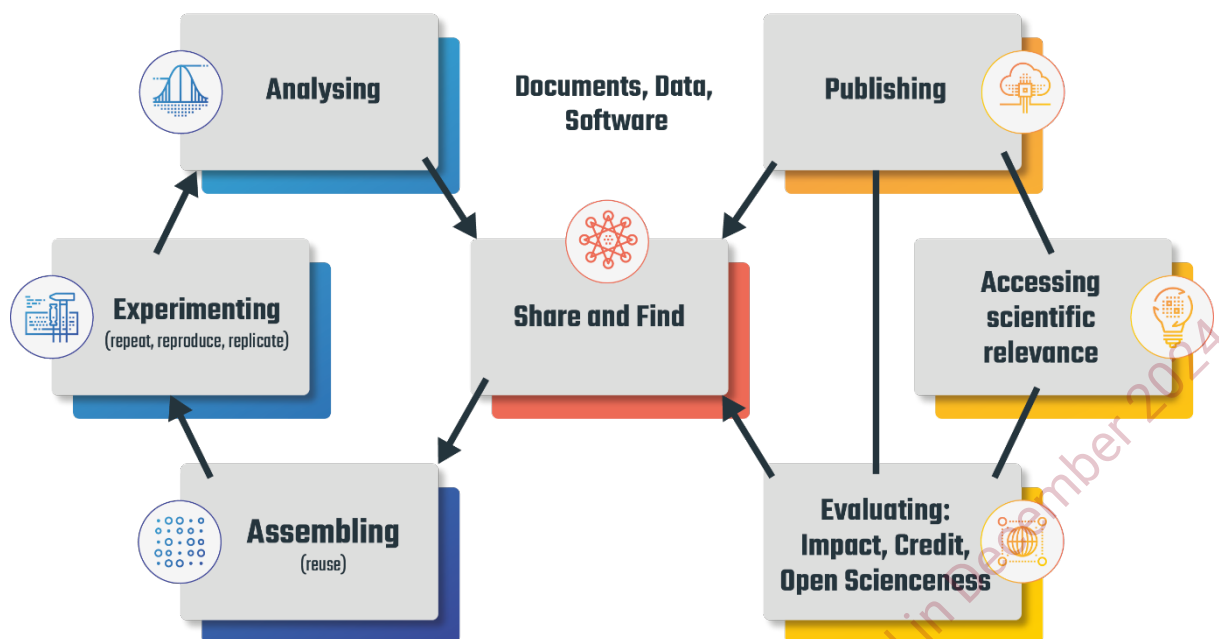
Figure 1.1: Research activity flows

1.3.1. Brief historical context

Until the emergence of academies and journals, science was mostly an individual endeavour supported by patrons. Results were kept secret as much as possible in order for the patrons to be able to benefit from the research results.

In the seventeenth century, both the creation of academies where scientists could cooperate and exchange knowledge, and the deployment of printing capabilities, which produced academic journals, allowed a move towards a more open way of science.

Nowadays, academic journals have taken a key role in the research lifecycle by allowing the transfer of knowledge, but also as a basis for research assessment through citation mechanisms.

When World War II ended, the global scientific community had the opportunity to look at the future with new eyes and with new goals in sight. In the United States, Vannevar Bush delivered the report 'Science the Endless Frontier', at the request of President Roosevelt.

This report led to the creation of the National Science Foundation (NSF). Public investment in research was recognised as a priority. Since then, public-funded research has developed around the whole world. In the same period, Europe organised cooperation by establishing research organisations such as CERN, for example, which was created in 1954.

In the 1980s, with the final objective of defining and implementing an overall development, research and demonstration strategy at Community level, the European Commission established the **First Framework Programme** covering three years from 1984 to 1987. The total budget dedicated to the programme was €3.75 billion. The programme focused on specific scientific and technical objectives, such as 'improvement of the management of energy resources'; 'promotion of industrial competitiveness'; 'improvement of living and working conditions'; 'promotion of the agricultural competitiveness'; 'improvement of raw materials management'; 'stepping up development aid' and 'improving the effectiveness of the Community's scientific and technical potential.' [EC_FP1]

Successive framework programmes came with increasing budgets. With increased funding, over time, countries became engaged in policy decisions regarding the use and impact of

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 20 –

research activities. The way knowledge, specifically that created with the support of public funding, would be shared became a key societal and political topic.

The debate was fuelled by prior research. For example, the Mertonian paradigm, introduced by Robert Merton in his book *The Sociology of Science* in 1942, was based upon four 'norms':

- **Communism**. All scientists should have common ownership of scientific goods (intellectual property), to promote collective collaboration; secrecy is the opposite of this norm.
- **Universalism**: Scientific validity is independent of the socio-political status/personal attributes of its participants.
- **Disinterestedness**. Scientific institutions act for the benefit of a common scientific enterprise, rather than for the personal gain of individuals within them.
- **Organised scepticism**. Scientific claims should be exposed to critical scrutiny before being accepted, both in methodology and institutional codes of conduct.

At the turn of the twenty-first century, the digital age created new avenues for knowledge sharing. These new opportunities have been recognised by research communities across the world. More international collaborations were launched, leveraging the interconnections made possible by the internet. Open Science emerged from the meeting of the needs (sharing knowledge) with the means (digital technologies).
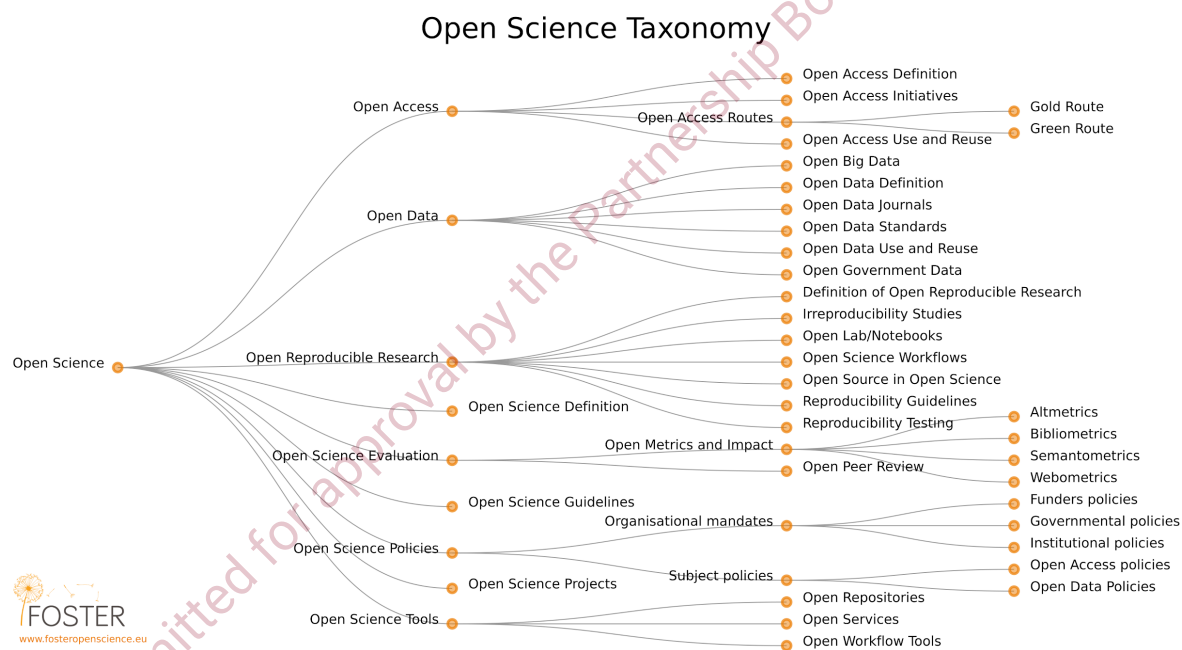
Figure 1.2: Open Science taxonomy (from the FOSTER project)

The European Commission identified early the potential of digital technologies in changing the way research is conducted. In 2017, the FOSTER project was funded to study the practical implementation of Open Science in Horizon 2020 and beyond [FOSTER]. The project developed the FOSTER portal as a platform that brings together the best resources addressed to those who need to know more about Open Science, or need to develop strategies and skills for implementing Open Science practices in their daily work. The resources are structured around the Open Science taxonomy shown in Figure 1.2 above; the role of Open Science in the research lifecycle is summarised in Figure 1.3 below.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 21 –
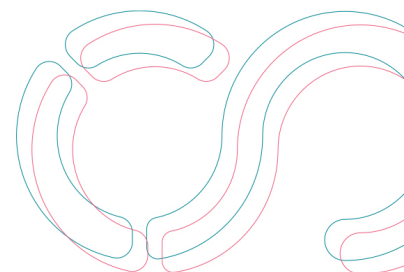
Figure 1.3: Research lifecycle and Open Science (from the FOSTER project)

1.3.2. Open Science facets: documents, data and software

The FOSTER project also identified the Open Science 'facets' that could be shared by scientists within and between research communities (Figure 1.4).
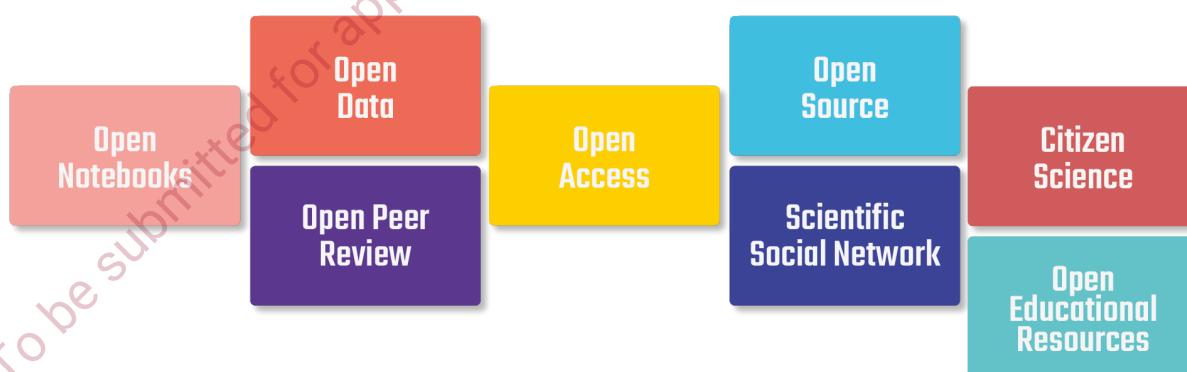
Figure 1.4: Open Science facets (from the FOSTER project)

Those 'facets' are different in nature and therefore sharing them requires specific approaches. The complementarity and differences between documents, data and software are well known by the computer science community. The sharing processes for each of these are explored in the following sections.
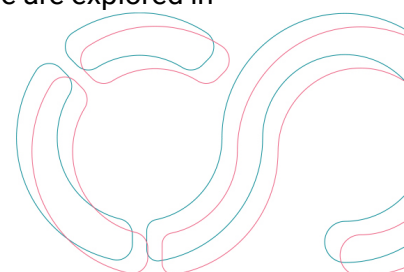
1.3.2.1. Documents

Publications, notebooks and educational materials are documents written in natural languages. They are designed to be read by people, while machines may leverage their content through document processing. Publications were (and still are) the basis for information exchange between scientists. The first instance of the World Wide Web to be deployed, in the early 90s, was a Web of documents. Therefore, technology met user needs and digital publications became the norm. This soon created friction with regard to the intellectual property rights, which up to this point were mostly owned by publishing corporations. On 14 February, 2002, the Budapest Open Access Initiative produced its original declaration which started as follows: 'An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds.' [Budapest_OAI] The Open Access movement launched the debate with publishers which is now focused on legal or contractual issues about 'ownership' of the content.

Since then, multiple examples of open access initiatives have flourished around the world. Launched in the United States in 1991 and currently managed by the University of Cornell, arXiv is an open access repository of electronic preprints (known as e-prints) approved for posting after moderation, but not full peer review. It consists of scientific papers in the fields of mathematics, physics, astronomy, electrical engineering, computer science, quantitative biology, statistics, mathematical finance and economics, which can be accessed online. In many fields of mathematics and physics, almost all scientific papers are self-archived on the arXiv repository before publication in a peer-reviewed journal. Following the arXiv model, similar archives have been established in many different disciplines. Recently, preprints have become popular in life sciences and have turned out to be essential in the scientific communication related to COVID-19.

In France, Hyper Articles on Line (HAL) is an open archive where authors can deposit scholarly documents from all academic fields. French scientists are encouraged to deposit their publications here. New research assessment practices are developed by considering publications if and when they are openly available in HAL.

1.3.2.2. Data

Data are heterogeneous in nature and their volume explosion requires the systematic use of machines. Infrastructures have been built, and continue to expand, to store and preserve data for future reuse. Machines are used for many purposes. Raw data have to be processed to generate useful data. Large datasets need to be processed to visualise useful information. Analysing large datasets and extracting information through computing such as machine-learning technologies has become common practice. Building models and assessing their value through computer simulation has also become common practice and requires new computing architectures as models become more and more complex.

The importance of data management for science has been recognised for a long time. The pervasive availability of digital information is now being viewed as bringing a paradigm shift in the way science is conducted. Jim Gray, who received the Turing Award in 1998 'for seminal contributions to database and transaction processing research and technical leadership in system implementation' [Wikipedia_Gray1] has introduced the concept of data-intensive science or e-Science as the 'fourth paradigm' of science (after empirical, theoretical and computational paradigms) and asserted that 'everything about science is changing because of the impact of information technology' and the data deluge [Wikipedia_Gray2].

The heterogeneity of data and their originating research communities has also resulted in a heterogeneity in the way data is made available. Unlike narrative publications, which can all be accessed through similar means, there is no single way to access research data. Only recently has there begun to be more uniformity in the way data can be accessed.

In Europe, the Zenodo project is 'built and developed by researchers, to ensure that everyone can join in Open Science. The OpenAIRE project, in the vanguard of the open access and open data movements was commissioned by the EC to support their nascent Open Data policy by providing a catch-all repository for EC-funded research. CERN, an OpenAIRE partner and pioneer in open source, open access and open data, provided this capability and Zenodo was launched in May 2013. In support of its research programme CERN has developed tools for Big Data management and extended Digital Library capabilities for Open Data. Through Zenodo these Big Science tools could be effectively shared with the long-tail of research.' [Zenodo]

1.3.2.3. Software

Software source code uses programming languages that are designed to be used by both machines and people. The role of software has become essential as research activities often depend on specific or generic software. Infrastructures for storing and preserving research software (and, if necessary, the environment in which it is executed, e.g. virtual machines) in both source or executable forms are more recent, while the need for reliable service is more and more required.

In order to be usable by scientists, research software archives need to comply with specific requirements. They have to keep multiple versions in order for scientists to be able to use the version that will ensure reproducibility. Research software uses generic components such as operating systems, compilers, scientific libraries, etc. Therefore, in order to allow reproducibility, these generic components also need to be kept. As a consequence, archiving of research software has to be part of general-purpose software archives.

Software Heritage [SW_Heritage] is an initiative launched by Inria, the French National Institute for Research in Digital Science and Technology, in 2015. Its goal is to archive, preserve and make available the code of all open source software available. Archiving research software will have to consider leveraging initiatives such as Software Heritage in order to deliver the value needed by scientists to reproduce scientific experiments in a trustworthy manner.

Other opportunities to share software are coming from cloud-based infrastructures where computing services are made available to scientists over the internet.

In order to deliver the potential that Open Science promises, a new generation of infrastructures is needed to make documents, data and software available to scientists in an easy-to-use and integrated environment. This new generation of infrastructures will comply with a range of guiding principles described in detail in Section 4.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 24 –
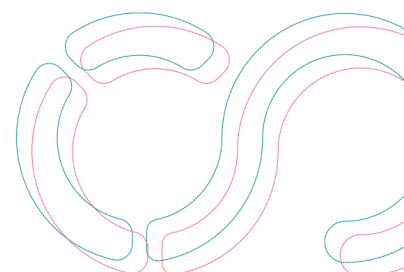
Use Case: IPOL – a research journal on reproducible research algorithms

Since Donoho, Buckheit and others warned in 1995 about the credibility crisis in reproducible research [Donoho_1995], the problem of reproducibility and reliability of research results has been confirmed by thousands of scientists from different fields. While in some disciplines (biology, for example) it might be difficult to get back to exactly the same conditions when replicating an experiment, in computational sciences there is no excuse not to address this issue. This recommendation is made strongly in the 'Scholarly Infrastructures for Research Software' report issued by the EOSC Architecture Working Group [WG_Arch_SIRS] and models such as IPOL are proposed as a way forward.

Image Processing on Line (IPOL) [IPOL], a peer-reviewed research journal in signal processing (mainly image and video analysis) which emphasises the role of mathematics as a source for algorithm design and the reproducibility of the research, was founded in 2009 as an attempt to provide an answer to the problem of reproducibility and reliability of research results. At first it was focused exclusively on image-processing algorithms, but soon it expanded to more general signal-processing topics, such as video and audio, and very recently (2020) started to address machine-learning applications.

In a classic publication, the text of the article itself is the only result of the research, and most of the time the source code or any additional data is not given or is simply considered as supplementary material. In the case of IPOL, a single publication is made up of three items which are all under the same DOI:

- The PDF of the article (as in a classic journal, under a free-documentation licence);
- The source code of the method (under an open source licence);
- Any data needed to reproduce the results presented in the paper.

In some applications the article and the source code are not sufficient to reproduce the results in the article and, therefore, data must be part of the publication. For example, a method that presents a neural network architecture will require the description of the method in the article (say, explanations on the choice of the architecture), the source code (which includes any details), and also the weights of the network after being trained.

Source code and data are fundamental to ensure reproducibility and therefore the reviewers chosen by IPOL editors must check carefully that the sources match exactly the pseudo-codes given in the article, and that any data needed to replicate the results are available together with an explanation of how they was produced exactly.

Given the importance of source code and data in IPOL, in 2020 the journal started a fruitful collaboration with the universal archive Software Heritage [SW_Heritage_IPOL] to ensure that sources and data are permanently archived and properly referenced.

Each IPOL publication also comes with an online demonstration where users can test the algorithms quickly with their own data. Every experiment (input, output and parameters) performed with user-uploaded data is added to the archive. After years of activity, the IPOL archives are rich and very useful to identify the interests of the scientific community and industrial applications.

Moreover, the definition of execution environments that includes the exact dependencies used in the experiment along with the source code (e.g. in the form of software containers) enables a more accurate and convenient environment to reproduce the experiments. For example, in IPOL all Python submissions are executed in a virtual environment on which the exact version of the packages is declared. Annotating both source code and the execution environments will be key to fostering reusability of research.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 25 –

1.3.3. Open Science adoption: progress and resistance

Nowadays, Open Science practices have become a reality, aiming at culturally and technologically upgrading the research lifecycle, to accelerate research and make it more efficient and sustainable, and maximising its overall impact on and trustworthiness for society as a whole.

Research communities, research funders, research organisations and policy makers are contributing in their own capacities to make science 'as open as possible, as closed as necessary'[4]. Research communities define discipline-specific Open Science *policies*, and the related implementation *guidelines* and *incentives*, in respect of researchers' needs, practices, services and legal constraints. They also define *Open Science implementation roadmaps*, to plan a progressive removal of the barriers.

Research funders, research organisations and policy makers are supporting and funding the e-infrastructures, the research infrastructures, the researchers and the projects necessary to support the Open Science paradigm. Accordingly, they themselves define policies, guidelines, roadmaps and mandates, to foster and advocate Open Science, which should facilitate the implementation of research community roadmaps while taking into account their policies and guidelines. (For example, for life sciences the European Research Council (ERC) recommends Europe PMC – an Open Science platform dedicated to life sciences – for the deposit of preprints and open access versions of scientific articles to facilitate their diffusion, and recommends ELIXIR Deposition Databases for Biomolecular Data). The resulting Open Science matrix is shown in Figure 1.5.
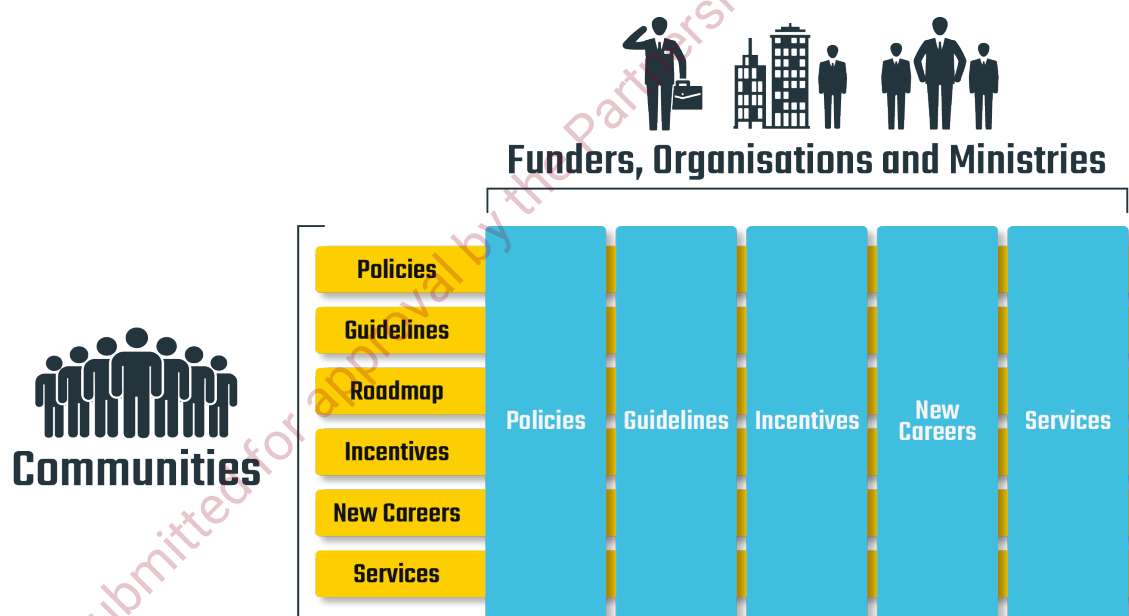
Figure 1.5: Open Science at the crossroads between communities and funders, organisations, and ministries

Such trends set the path to Open Science, expecting it to become the 'new normal' in a not-so-distant future, but also highlight a number of cultural and technological barriers that still need to be overcome.

Despite the acceleration of digital science, scientific practices often still follow traditional patterns, which focus on the dissemination of science via research literature, in some cases

---

[4] For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".
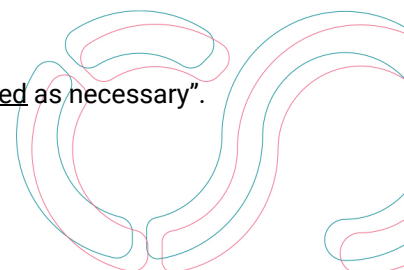
via research data, rarely via research software, and almost never via sharing of reproducible experiments.

Research funders and policy makers have been trying to improve this, but still technologies, policies, guidelines and incentives are not fully deployed. Although the current digital age enables automation of substantial parts of the research lifecycle, science is far from being as fully tracked, described, transparent and reusable as it could and should be at community level and cross-community level.

In addition, current incentives for researchers, including career prospects, are not rewarding Open Science practices, with scientific credit and research impact for researchers. When it comes to rewards, policies, institutions, funders and initiatives focus almost solely on scientific articles. In this context, open access to literature has a non-trivial cultural obstacle, as scientific credit is currently often based on easily available indicators (e.g. Impact Factor) measured via citation indexes that constitute the core business of private publishing companies, which in turn offer subscription-based, non-open access journals.

Around the world, multiple initiatives have been engaged in securing open access to publications. In Europe, for example, Plan S is an initiative for open access science publishing launched in 2018 by cOAlition S, a consortium of national research agencies and funders from twelve European countries [Plan_S]. The plan requires scientists and researchers who benefit from state-funded research organisations and institutions to publish their work in open repositories or in journals that are available to all by 2021.

> The recent COVID-19 emergency has given clear evidence of the benefit that Open Science practices can bring, but also of the large areas for improvement. The demand for a portal providing access to COVID-19-related data sources, computing and thematic services within the ELIXIR research infrastructure was rapidly satisfied, benefiting from past research infrastructure investments in the domain.

Still, reproducibility is not fully supported, programmatic access remains hard and so are monitoring and discovery of research results within and across disciplines.

Actions to address these challenges are being undertaken. These require creating and reinforcing synergies within and across research communities, research-performing organisations, funders and policy makers, to make science more efficient and a new 'tangible product' enabling new career opportunities and added value.

Indeed, this process introduces new career opportunities, for professionals specialising in the definition and implementation of Open Science policies, guidelines and roadmaps (e.g. data stewards, Open Science managers). In order to ensure a fertile, multi-disciplinary, monitorable research ecosystem, several practices must be agreed upon at cross-community level. For example, a common understanding of the research entities involved in the scientific process (e.g. standard identifiers registries for authors, organisations and services), in order to enable a common way to track and monitor science, and hence to evaluate its openness. Other examples are open access guidelines for literature, incentives towards the implementation of 'Open Science by design' services, or incentives towards publishing products (e.g. open source research software).

The classification and sharing of policies, guidelines and roadmaps, as well as skills and experiences, as made available by research communities, research funders, research-performing organisations and policy makers, would accelerate and optimise the implementation of Open Science, while facilitating convergence of intent.
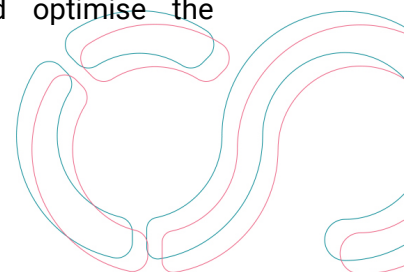
Such integration and sharing will have to take into account the fact that the initial efforts towards Open Science have been deployed at very different levels from one discipline to another and from country to country.

1.3.4. Limits to Open Science

While it will take time for Open Science to be pervasive and become the 'new normal', the current efforts being undertaken across the world and in Europe, in particular, produce constant progress. The support from many stakeholders, each in their own capacity, strengthens the practices and therefore the impacts of Open Science.

However, to describe a full vision of what Open Science will bring, it is also important to take into account limits that will have to be respected in order for Open Science to deliver value to society while mitigating undesirable consequences. There are four essential limits to Open Science:

- Privacy;
- Security;
- Property;
- Sovereignty.

1.3.4.1. Privacy

Privacy rights for individuals are legally recognised in Europe. The principles of the General Data Protection Regulation (GDPR) state that 'personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes (purpose limitation)'. The regulation goes on to state other principles such as data minimisation, accuracy, storage limitation, integrity and confidentiality. Exceptions are made to facilitate research, but these must always be balanced with the risks for the individual.

EOSC will take into account this regulation. As a consequence, privacy will require limits to be placed on the dissemination of datasets that contain personal information in a transparent manner.

1.3.4.2. Security

In the Open Science context, security refers to the protection of data integrity and is, of course, necessary for Open Science to deliver its value to scientists. Throughout its evolution the internet has considered security as an essential requirement for the infrastructure and EOSC will benefit from the corresponding efforts, products and services.

Federating security policies implemented by the research infrastructures may, in specific cases, limit openness. EOSC will have to strike the right balance between trustworthy security policies and open access to research artefacts.

There are other cases where security will have to be taken into account. These relate to the content of the information itself, which may require special attention. It will be the role of the infrastructure governance to decide whether or not information may be made available openly and to whom.

1.3.4.3. Property

Data acquisition can be a costly process. The entity performing that process has rights and responsibilities with regard to defining the use of the data. Depending upon the funding mechanism, the 'owner' may decide to limit access to the data. Open Science infrastructures will have to provide a way for stakeholders to exercise their rights, possibly limiting open access to the data. In any case, the stakeholder policy will have to be shared openly.
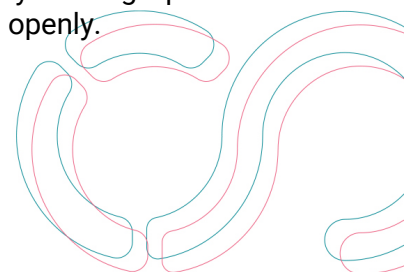
1.3.4.4. Sovereignty

Member States may decide that data management has to follow specific rules. Member States should be able to exercise their full right and power over documents, data and software, limiting full openness. Striking the right balance between ensuring sovereignty and ease of use will require special attention.

As a consequence of those limits, an infrastructure for Open Science has to offer capabilities for identity and rights management. Being a federation of research infrastructures, those capabilities will have to be powerful enough to offer individuals, organisations or governments a way of exercising the required control while keeping knowledge 'as open as possible, as closed as necessary'[5].

1.4. Next-generation infrastructure

1.4.1. Learning lessons from the recent past

In order to position EOSC in its context, it is important to briefly review the evolution of **digital services** made available to scientists over the years, together with the **key technologies** (networking, hardware and software) that were developed and deployed to allow those services to flourish, as well as the **funding models** and the **policy decisions** that exploited these new capabilities.

This review is summarised in Table 1.1 and allows EOSC to be positioned in time and technological space, leading to the definition of the **'raison d'être' of the initiative**.

1.4.1.1. 1970 The internet, mainframes and leased lines

The world recently celebrated 50 years of the internet. In September 1969, a few characters were exchanged between four mainframe computers installed in different locations in the west of the United States. The first message was intended to send the word *login* from one computer to another (and it failed). It was the first step towards delivering the **remote login** service: allowing an end user to use remote computers such as the mainframes or supercomputers of the day.

In the years following this historical moment, the internet allowed the development of many other services, predecessors to the ones routinely used today: **email**, **file transfer**, **chat**. For a long while those services were only deployed within research communities. It is important to note that this deployment was rapidly global. Connections, gateways and routers were assembled to build the **first generation** of the global internet infrastructure and to allow scientists to improve the way they collaborated. Funding mostly came from **research-funding organisations**. The development of internet **Standards** grew as a bottom-up effort, driven by the Internet Engineering Task Force (IETF), exploiting the communication capabilities provided by the internet (e.g. mailing lists, file transfers, news) to assemble hundreds of computing and telecommunications scientists and design the internet architecture

---

[5] For research artefacts, the interpretation of "as <u>closed</u> as necessary" is "as <u>restricted</u> as necessary".
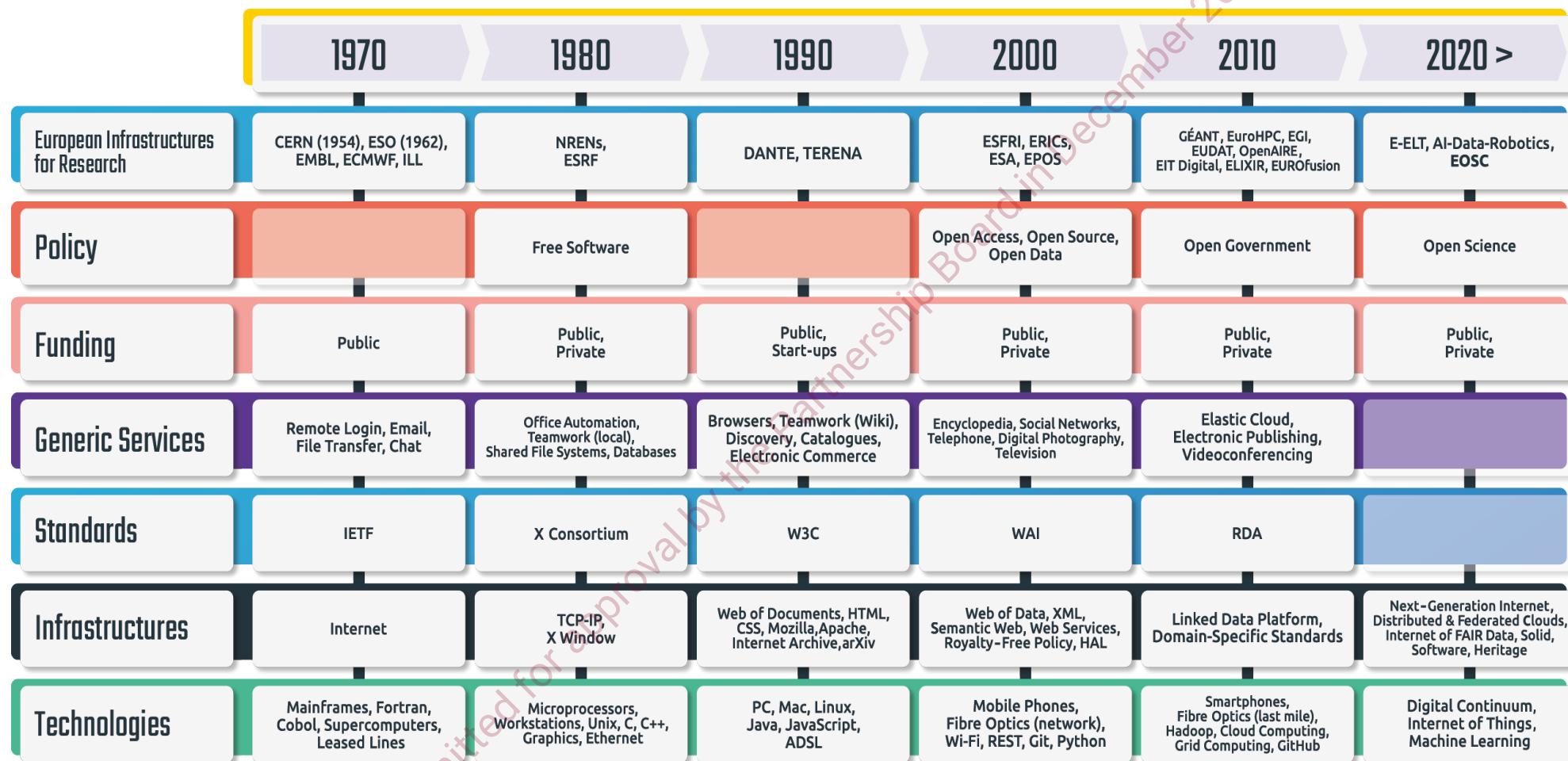
eosc

Table 1.1: EOSC in its technological context

	1970	1980	1990	2000	2010	2020 >
European Infrastructures for Research	CERN (1954), ESO (1962), EMBL, ECMWF, ILL	NRENs, ESRF	DANTE, TERENA	ESFRI, ERICs, ESA, EPOS	GÉANT, EuroHPC, EGI, EUDAT, OpenAIRE, EIT Digital, ELIXIR, EUROfusion	E-ELT, AI-Data-Robotics, EOSC
Policy		Free Software		Open Access, Open Source, Open Data	Open Government	Open Science
Funding	Public	Public, Private	Public, Start-ups	Public, Private	Public, Private	Public, Private
Generic Services	Remote Login, Email, File Transfer, Chat	Office Automation, Teamwork (local), Shared File Systems, Databases	Browsers, Teamwork (Wiki), Discovery, Catalogues, Electronic Commerce	Encyclopedia, Social Networks, Telephone, Digital Photography, Television	Elastic Cloud, Electronic Publishing, Videoconferencing	
Standards	IETF	X Consortium	W3C	WAI	RDA	
Infrastructures	Internet	TCP-IP, X Window	Web of Documents, HTML, CSS, Mozilla, Apache, Internet Archive, arXiv	Web of Data, XML, Semantic Web, Web Services, Royalty-Free Policy, HAL	Linked Data Platform, Domain-Specific Standards	Next-Generation Internet, Distributed & Federated Clouds, Internet of FAIR Data, Solid, Software, Heritage
Technologies	Mainframes, Fortran, Cobol, Supercomputers, Leased Lines	Microprocessors, Workstations, Unix, C, C++, Graphics, Ethernet	PC, Mac, Linux, Java, JavaScript, ADSL	Mobile Phones, Fibre Optics (network), Wi-Fi, REST, Git, Python	Smartphones, Fibre Optics (last mile), Hadoop, Cloud Computing, Grid Computing, GitHub	Digital Continuum, Internet of Things, Machine Learning

1.4.1.2. *1980 Unix, the personal computer, ethernet*

A few years later, progress in microprocessor technologies allowed the design of personal computers, which rapidly became the main tool for scientists to do their research. **Scientific workstations** were born, providing researchers with the best technologies of the day in terms of computing, graphics and networking. Equipped with office automation and computer-aided design software, these devices were connected to **local area networks**, changing the way scientists would collaborate within a team or a laboratory. Unix, C, C++, TCP-IP and X Window System were the software standards that allowed interoperability between these devices. Scientists were able to share their results **within their teams and/or their laboratories** by sharing databases, for example. Distributed file systems allowed the development of new ways of collaboration. It is important to note that many of those software standards were **developed within computer science laboratories both private and public** (Unix, C and C++ at Bell Labs, TCP-IP at UC Berkeley, X Window System at MIT).

1.4.1.3. *1990 The World Wide Web: the internet becomes pervasive*

Twenty years after the birth of the internet, the Web was invented at CERN, the European Organisation for Nuclear Research. In March 1989, Tim Berners-Lee wrote a memo entitled 'Information Management: A proposal'. The project was approved and Berners-Lee developed the World Wide Web using a NeXT machine, the most advanced workstation of the moment. It is interesting to note that **user research organisations** were also at the origin of the deployment of the Web. For example, in the United States, the first Web server was installed at Stanford Linear Accelerator Center (SLAC). Also, while many Web browsers blossomed around the world, Mosaic from the **National Center for Supercomputing Applications** was made available on PC, Mac and Unix machines and became a huge success. Scientists were able to share documents, graphics and images thanks to the worldwide deployment of the Web. Essential generic services such as discovery or service catalogues were developed within **computer science departments** (Stanford University). **Open source** efforts delivered key software components of the infrastructure such as Apache, the Web server, wiki, the collaborative tool that was to be used widely by research communities. Services that were not planned in the original design of the internet (e.g. real-time signals) and that used to require specific networks (telephone, television) moved to the internet and offered new opportunities for innovation. As a whole, the success of the Web fuelled the **massive deployment of the internet infrastructure with private and public funding**. In order to 'lead the Web to its full potential', Tim Berners-Lee moved from CERN to MIT and launched the World Wide Web Consortium (W3C), following the lessons learned from the X Window Consortium. By design W3C would have multiple hosts and Inria, the French National Institute for Research in Digital Science and Technology, in Europe and Keio University in Japan became the European and Asian hosts of W3C.

1.4.1.4. *2000 Documents and data*

While the first version of the World Wide Web allowed the creation of a **Web of documents**, using its original components URL, HTTP and HTML, very soon the request for a **Web of data** led W3C to develop **XML and a family of related standards**. This effort brought together different communities with participants coming from academia, public and private organisations. It also became clear that beyond the description of syntaxes, in order to make full use of data, semantics needed to be formalised. Tim Berners-Lee launched the development of the **Semantic Web** within W3C. The growing impact of the Web on society was recognised by W3C. The **Web Accessibility Initiative** (WAI) became part of the strategic priorities of the consortium in order to design a Web that could be used by people with disabilities. At the same time, in order to strike the right balance between public and private investments, W3C developed a **royalty-free policy** for using W3C standards.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 31 –
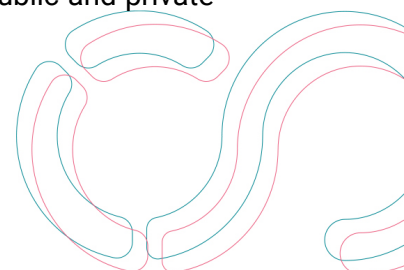
During the same period, the deployment of the internet was able to benefit from a wide range of new networking technologies, from **fibre optics** (within the core of the network) to **Wi-Fi** (at the edge). The last-mile challenge addressed by **ADSL** technology was also about to be covered, by the deployment of **mobile infrastructures**. The design of the internet allowed the use of all these technologies in order to build the resilient infrastructure it is today.

The vision of Tim Berners-Lee became true when 'thousands of flowers bloomed' on the Web [Forbes_TB-L], ranging from an **open encyclopaedia** to the **emergence of social networks**. The internet infrastructure was able to carry **telephone** and **television** signals. Digital photography was about to become widespread.

1.4.1.5. 2010 Smartphones, cloud computing, linked data platform

The momentum of the internet only accelerated further when microelectronics technologies allowed the functions of a telephone, a computer and, soon enough, a television to be embedded in a handheld device. The **smartphone** was born, filling the pockets of millions of people around the globe, including researchers.

At the same time, the decreasing cost of computing and storage resources and the improvements in bandwidth of the internet allowed the launch of the **cloud computing** paradigm. Progress in computing architecture during the early 2000s allowed the **Grid Computing** route to be added to the general evolution of supercomputing. Scientific problems could be addressed with a wide range of architecture possibilities.

Building upon the outcome of the Semantic Web efforts, W3C launched the **Linked Data Platform** initiative with the goal of creating the architecture components that will allow data to be 'linked' and lay the ground for the **internet of FAIR** (findable, accessible, interoperable and reusable) **data**.

Many people in the research community recognised that the time had come to leverage the progress of the internet infrastructure, leading to the launch of the **Research Data Alliance** (RDA) in 2013 to 'build the social and technical bridges to enable the open sharing and reuse of research data'.

1.4.1.6. 2020 Lessons learned

The emergence, over the years, of so many new digital products and services followed similar paths:

- New user needs served by breakthrough technologies;
- Next-generation services deployed on existing infrastructures;
- Use of novel services pioneered by research communities and then deployed for the general public;
- Initial efforts supported by public funds and then embraced and further developed by industry;
- Innovation fuelled by private funds;
- Pervasive deployment delivered by open and proprietary offerings.

Over the last 50 years, exceptional developments have allowed scientists to use machines that improve the exchange of documents, data, software and related information between people.

Looking into the future, further improvements in digital technologies will create new opportunities. Machines will be assembled into complex systems and put at the service of research teams composed of experts from any discipline working from anywhere in the world.

While the potential offered by current (and future) devices is unique in human history, the current limits come from the programmability of those complex systems to develop friendly
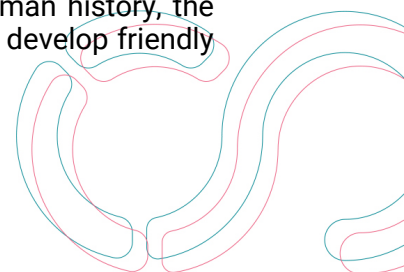
user-oriented services and the capacity to find, access and reuse data in an interoperable framework.

The Horizon Europe programme will address many of these challenges through various partnerships and the EOSC Partnership will form links with those inside Europe.

1.4.2. Networking: the next-generation internet (NGI)

The NGI is an ambitious research and innovation programme with an EC investment of more than €250m for the initial phase between 2018 and 2020, and is an important part of the upcoming Horizon Europe programme (2021–2027). Focus has been on advanced technology applied to evolve the internet into an 'Internet for Humans'. The initiative addresses the challenges of privacy and trust, search and discovery, by promoting decentralised architectures, blockchain, the Internet of Things (IoT), social media and interactive technologies, as well as technologies supporting multilingualism and accessibility. Also, the whole new area of next-generation Internet of Things research will be covered under the NGI programme.

EOSC will benefit from the evolution of the internet towards an 'Internet for Humans'. EOSC will be able to face the challenges of privacy, security, property and sovereignty by leveraging the results of the NGI initiative. EOSC will also make use directly of the new IoT technologies and infrastructures, as this is one of the sources of the large amount of data that can be used for research inside the EOSC ecosystem.

1.4.3. Hardware: the computing continuum

In their paper 'Harnessing the Computing Continuum for Programming Our World', Beckman, Beck, Dongarra et al. describe the challenges facing scientists in mastering systems composed of elements as different as smart sensors at the one end and supercomputers at the other [Beckman_2019]. The **Computing Continuum** is described in Figure 1.6.

EDGE			FOG			HPC/Cloud
Size	Nano	Micro	Milli	Server	Fog	Campus	Facility
Example	Adafruit Trinket	Particle.io Boron	Array of Things	Linux Box	Co-located Blades	1000-node cluster	Datacenter & Exascale
Memory	0.5 KB	256 KB	8 GB	32 GB	256 GB	32 TB	16 PB
Network	BLE	WiFi/LTE	WiFi/LTE	1 GigE	10 GigE	40 GigE	N*100 GigE
Cost	$5	$30	$600	$3K	$50K	$2M	$1000M

Count = 10^9
Size = 10^1

Count = 10^1
Size = 10^9

Count x Complexity = ~Constant   Stateful vs. Stateless

Source: Beckman, Beck, Dongarra, Ferrier, Reed, and Taylor / University of Utah

Figure 1.6: The Computing Continuum (from the ETP4HPC SRA)

In its Strategic Research Agenda (SRA) published in March 2020, the Institutionalised Partnership EuroHPC extends the concept and introduces a new paradigm called the 'Digital Continuum':

*'The rapid proliferation of digital data generators, the unprecedented growth in the volume and diversity of the data they generate, and the intense evolution of the methods for analysing and using that data are radically reshaping the landscape of scientific computing. The most critical problems involve logistics of wide-area, multistage workflows that move back and forth across the **computing continuum**, between the multitude of distributed sensors, instruments and other devices at the network's edge and the centralised resources of commercial clouds and HPC centres.'* [ETP4HPC_SRA]

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu   - 33 –
Reg. number: 0755 723 931 | VAT number: BE0755 723 931
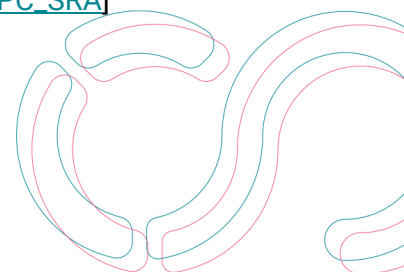
The ETP4HPC SRA has been designed to strengthen and develop further the European position with respect to the 'Digital Continuum' during the Horizon Europe programme. EOSC will ensure close collaboration in order to contribute to the Digital Continuum.

1.4.4. Software: visualise, analyse, predict

Key elements of the research lifecycle involve observation, explanation and prediction. If and when large datasets are available, scientists need to use machines to support their work. Observation requires machines to help in visualisation; explanation requires machines to analyse data and derive models; prediction requires machines to check hypotheses. The larger the datasets, the harder becomes visualisation, analysis and prediction. Scientists need to use advanced software in order to improve their insights. Scientists also use machines and software to check hypotheses, simulate phenomena and strengthen their ideas and models.

During the Horizon Europe programme, the AI (Artificial Intelligence), Data & Robotics partnership will help position Europe in the global development of AI technologies. The summary of the partnership proposal states:

'Access to relevant and high-quality data is widely recognised to be one of the crucial elements in building an AI economy in Europe. Building on the great efforts to make industrial and public sector data more accessible during Horizon 2020, the access to data will have to scale up in Horizon Europe, address a broader set of sectors and drastically increase the quantity of high-quality datasets available.' [EP_AID&R]

EOSC will make available the high-quality scientific datasets to be consumed by machine-driven AI applications at the service of science.

1.4.5. Data: findable, accessible, interoperable and reusable

The first Web client, developed by Tim Berners-Lee, was both a browser and an editor. A user could therefore not only read but could also create content. The Web was conceived originally as a collaborative space. However, when the first popular browser, Mosaic, came along, in 1993, it included images but the editing capability was taken out. It was considered too difficult a problem.

The Web was also originally designed to be a space for data as well as documents. The Linked Data Platform, recently developed, is an important step towards giving data first-class citizen status on the Web.

These two principles, of **read-write capability** and **managed data accessibility**, were part of the original vision for the Web. They are still not available as Web features. They may be present at the application level. For example, wikis or social networks offer the write capability within their own environment.

Since 2015, Tim Berners-Lee has been working on SOcial LInked Data (SOLID) in order to offer those two capabilities for the whole Web. Those capabilities are essential in order for EOSC to achieve its full potential.

Referring to Neil Armstrong's famous sentence when landing on the moon, Berners-Lee defines what he thinks is 'A small step for the Web':

'I have always believed the Web is for everyone. […] This is why I have, over recent years, been working with a few people at MIT and elsewhere to develop SOLID, an open source project to restore the power and agency of individuals on the web. […] SOLID is a platform built using the existing web. It gives every user a choice about where data is stored, which specific people and groups can access select elements, and which applications you use […] SOLID is guided by the principle of "personal empowerment through data" […] I'm incredibly optimistic for this next era of the web […] The future is still much bigger than the past.' [TB-L_Step] *'The issue with writing data, as Wikipedia and others have learned,*

*is that you need a degree of **control over who can write what**. The writer needs to have permissions describing what individuals can do to the data. And to have permissions **you need to have a system for identity** – a way of uniquely confirming that an individual is who they purport to be. Hence, based on existing Web standards and the result of decades of work, **SOLID has read-write functionality, incorporating permissions and identity, along with data manageability and real-time updates.** It realises the Web as originally envisioned and provides a platform for the next generation of truly empowering and innovative applications.'*

While the success and the deployment of SOLID is yet to be proven, the issues that SOLID addresses are at the core of what EOSC needs in order for scientists to find, access and reuse interoperable research results.

1.4.6. Machines for scientists: EOSC foundations

'Machines need direction from human minds, and human minds need inspiration from human leaders'. Arno Penzias, Nobel Prize-winning physicist, reminds us that it is up to us to build the environment and the infrastructures that will facilitate the exchange and composition of ideas, allowing scientists to cooperate globally and help solve the scientific and societal challenges of our time. In order for scientists to share the universe of scientific networked-accessible information, the essential foundations are:

- Persistent identifiers: a mechanism for naming and locating documents, data and software in a persistent manner;
- Metadata and ontologies: a mechanism for discovery of and access to documents, data and software in a structured manner;
- Internet identity: an authentication and authorisation infrastructure (AAI).

The first mission of EOSC will be to provide those mechanisms and that infrastructure to enable machines to get direction from human minds for the benefit of all.

EOSC will allow identified scientists to store, share, discover and access identifiable documents, data and software.

EOSC will allow identified scientists to (re)use identifiable documents, data and software, exploit identified services, reproduce experiments and address the problems of our time.

Section 1 of this SRIA has presented an overview of the development of scientific practice, as influenced by changing ideas and evolving technologies. Section 2 provides further context, focusing on the recent, current and future status of and strategy towards science and data in Europe, and the evolution of EOSC.
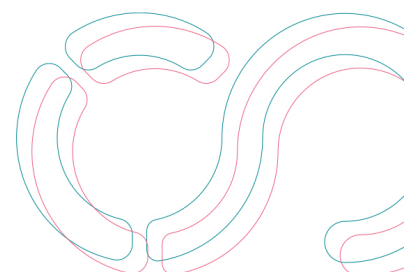
2 EOSC in the making

The European Open Science Cloud (EOSC) initiative is the tangible outcome of a number of key European and global policy milestones and position statements regarding Open Science. EOSC is an integral part of, and supports, the European Commission's strategy for realising the European Research Area (ERA), in particular the policy priorities of Open Innovation, Open Science and Open to the World and the goal of findable, accessible, interoperable and reusable (FAIR) data. This section outlines the EC policy context for and stages of EOSC's development, including its governance structure and activities during the transition period 2019–2020, and the landscape of national infrastructures and international initiatives, as well as the role of the governance of EOSC in the form of the EOSC Association.

2.1. European Research Area

The European Research Area (ERA) was launched by the European Commission in 2000 with the aim of better organising and integrating Europe's research and innovation systems and enhancing cooperation between the EU, the Member States, their regions and their stakeholders. It also aimed for the free circulation of researchers, scientific knowledge and technology throughout the EU and focused on stimulating cross-border cooperation and on improving and coordinating the research and innovation policies and programmes of the Member States.

In May 2016, the Commission published 'Open Innovation, Open Science and Open to the World – a vision for Europe' [EC_Open_Vision] as a key policy priority for realising the ERA, with the following goals in mind:

- **Open Innovation** will help Europe capitalise socially and economically on research and innovation results by bringing more actors and investments into the research and innovation process;
- **Open Science** will help Europe benefit from digitisation and support new ways of doing research and innovation, as well as opening up access to research data and results via digital technologies and collaborative tools;
- **Open to the World** will make Europe a leading voice in global debates and tackle societal challenges by engaging more in science diplomacy and global scientific collaboration.

It is as part of this strategy for Open Science that the European Commission adopted the European Cloud Initiative – Building a competitive data and knowledge economy in Europe [EC_Cloud] and launched the initiative of creating the European Open Science Cloud (EOSC). Both initiatives were designed to give a strong push in Europe towards Open Science and findable, accessible, interoperable and reusable (FAIR) research data management and to ensure that European researchers and professionals reap the full benefits of data-driven science. Building EOSC basically equates to designing a virtual commons where science producers and science consumers come together for more insights, new ideas and more innovation.

In April 2020, the European Commission introduced the ERAvsCorona Action Plan as part of the EU response to the coronavirus pandemic [EC_ERAvsCoronaAP]. Building on the overall objectives and the tools of the European Research Area, the action plan is a working document developed jointly by the Commission and national governments. It covers short-term actions based on close coordination, cooperation, data sharing and shared funding efforts.

On 30 September 2020, the European Commission adopted a Communication on a new European Research Area for Research and Innovation [EC_COM_New_ERA]. This initiative will improve Europe's research and innovation landscape, accelerate the EU's transition towards

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu   - 36 –
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

climate neutrality and digital leadership, support its recovery from the societal and economic impact of the coronavirus crisis, and strengthen its resilience against future crises.

> 'We live in times when scientific activities require faster and effective collaborations. We need to strengthen the European Research Area. An area embracing all of Europe, because knowledge has no territorial boundaries, because scientific knowledge grows with collaborations, because knowledge is trusted if there is open scrutiny of its quality. It has also more chances to achieve peaks of excellence and support an innovative and risk-taking industry to shape a resilient, green and digital future.'
>
> Mariya Gabriel (Commissioner for Innovation, Research, Culture, Education & Youth) on 30 September 2020

The Communication highlights the need to further promote researchers' mobility, skills and career development opportunities within the EU, gender equality, as well as better access to publicly funded peer-reviewed science. The Communication defines four strategic objectives:

1. **Prioritise investments and reforms in research and innovation** towards the green and digital transition, to support Europe's recovery and increase competitiveness.
2. **Improve access to excellent facilities and infrastructures** for researchers across the EU.
3. **Transfer results to the economy** to boost business investments and market uptake of research output, as well as foster EU competitiveness and leadership in the global technological setting.
4. **Strengthen mobility of researchers and free flow of knowledge and technology**, through greater cooperation among Member States, to ensure that everyone benefits from research and its results.

Fourteen actions have been defined and will be instrumental in realising the European Research Area [EC_ERA_Actions].[6]

> **Action 9**: Launch, via the Horizon Europe programme, a platform of peer-reviewed open access publishing; analyse authors' rights to enable sharing of publicly funded peer-reviewed articles without restriction; ensure a European Open Science Cloud that is offering findable, accessible, interoperable and reusable research data and services (Web of FAIR Data and Services); and incentivise Open Science practices by improving the research assessment system.

Once developed, the EOSC ecosystem should be a central element supporting a revitalised European Research Area, which aims to strengthen the foundations, quality and impact of the research and innovation system in the EU and in Member States.

In this new phase of the ERA, connectivity for the creation, circulation, diffusion and uptake of knowledge will be essential both to consolidate an ERA fit for the digital age and to develop a single EU market for data across sectors.

2.2. Priorities of the new Commission

In her statements to the European Parliament in July and November 2019, Commission President Ursula von der Leyen outlined the political priorities that would shape the Commission's work programme for the years 2019 to 2024 [Von_der_Leyen_Agenda]. These priorities include:

---

[6] In the meantime, the new ERA Policy Agenda, annexed to the Council conclusions on the ERA governance, sets out 20 concrete ERA actions for the period 2022-2024 to contribute to the priority areas defined in the Pact for Research and Innovation. More information at: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-area_en

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 37 –

- A European Green Deal;
- An economy that works for people;
- A Europe fit for the digital age;
- Protecting our European way of life;
- A stronger Europe in the world;
- A new push for European democracy.

The EOSC Partnership Proposal [EOSC_PP] already provides some preliminary insights as to how EOSC can contribute to the achievement of these priorities. It addresses common political priorities of the EU and its Member States such as making Europe fit for the digital age, interlinking data spaces across a more efficient European Research Area, mainstreaming Open Science and enabling European innovation to become more data-driven. Research outputs that are FAIR by design, combined with top-class digital infrastructures and artificial intelligence solutions, will ensure a true European capacity to tackle the Sustainable Development Goals (SDGs), to reach the EU's ambition for the Green Deal and to implement other national or sectoral policies. EOSC will ensure that European research and innovation (R&I) contributes in full to knowledge creation, to meeting global challenges and to taking part in European economic prosperity.

2.3. The European strategy for data

On 19 February 2020, the European Commission released 'A European strategy for data' [EC_Data_Strategy], one of the pillars of an overall digital strategy focusing on the need to put people first in developing technology, as well as on the need to defend and promote European values and rights in how technology is designed, made and deployed in the real economy.

The European strategy for data aims at creating a single market for data that will ensure Europe's global competitiveness and data sovereignty. Common European data spaces will ensure that more data becomes available for use in the economy and society, while keeping companies and individuals who generate the data in control.

Data is an essential resource for economic growth, competitiveness, innovation, job creation and societal progress in general. Businesses will have more data available to innovate. This will be done by launching practical, fair and clear rules on data access and use, which comply with European values and rules such as personal data protection.

To ensure the EU's leadership in the global data economy, the European strategy for data intends to:

- Adopt legislative measures on data governance, access and reuse, for example for business-to-government data sharing for the public interest;
- Make data more widely available by opening up high-value publicly held datasets across the EU and allowing their reuse for free;
- Invest €2 billion in a European High Impact Project to develop data-processing infrastructures, data-sharing tools, architectures and governance mechanisms for thriving data sharing and to federate energy-efficient and trustworthy cloud infrastructures and related services;
- Enable access to secure, fair and competitive cloud services by facilitating the set-up of a procurement marketplace for data-processing services and creating clarity about the applicable regulatory framework of rules on cloud;
- Empower users to stay in control of their data and invest in capacity building for small and medium-sized enterprises and digital skills;
- Foster the rollout of common European data spaces in crucial sectors such as industrial manufacturing, Green Deal, mobility or health.

The European strategy for data states notably that 'Data is at the centre of this [digital] transformation and more is to come. Data-driven innovation will bring enormous benefits for

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu   - 38 –
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

citizens, for example through improved personalised medicine, new mobility and through its contribution to the European Green Deal. In a society where individuals will generate ever-increasing amounts of data, the way in which the data are collected and used must place the interests of the individual first, in accordance with European values, fundamental rights and rules. Citizens will trust and embrace data-driven innovations only if they are confident that any personal data sharing in the EU will be subject to full compliance with the EU's strict data protection rules. At the same time, the increasing volume of non-personal industrial data and public data in Europe, combined with technological change in how the data is stored and processed, will constitute a potential source of growth and innovation that should be tapped.'

The EOSC ecosystem can be seen as part of the developments relevant for making 'Europe fit for the digital age'. The work conducted within EOSC to enable interoperability across research domains and data discovery to support multi-disciplinary reuse is critical to supporting collaboration with the data spaces envisaged by the European strategy for data. Research infrastructures already play a key role in EOSC. Engaging further with the research communities will be key to developing an EOSC for and by the researchers. Strong links with research domains will naturally foster opportunities for collaboration with the data spaces.

2.3.1. Europe-wide common data spaces

The European strategy for data defines nine initial common European data spaces that should be developed, building on the ongoing experience with the research community gained through the European Open Science Cloud. These data spaces are:

- An *industrial* (manufacturing) data space, to support the competitiveness and performance of the EU's industry;
- A *Green Deal* data space, to use the major potential of data in support of the Green Deal priority actions on issues such as climate change, circular economy, zero-pollution, biodiversity, deforestation and compliance assurance;
- A *mobility* data space, to position Europe at the forefront of the development of an intelligent transport system;
- A *health* data space, essential for advances in preventing, detecting and curing diseases as well as for informed, evidence-based decisions to improve the healthcare systems;
- A *financial* data space, to stimulate innovation, market transparency, sustainable finance, as well as access to finance for European businesses and a more integrated market;
- An *energy* data space, to promote a stronger availability and cross-sector sharing of data, in a customer-centric, secure and trustworthy manner;
- An *agriculture* data space, to enhance the sustainability performance and competitiveness of the agricultural sector through the processing and analysis of production and other data;
- Data spaces for *public administrations*, to improve transparency and accountability of public spending and spending quality, fighting corruption, both at EU and national level, and to address law enforcement needs and support services of public interest;
- A *skills* data space, to reduce the skills mismatches between the education and training system and labour market needs.

These European data spaces will give businesses in the EU the possibility to build on the scale of the single market. Common European rules and efficient enforcement mechanisms should ensure that:

- Data can flow within the EU and across sectors;

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 39 –

- European rules and values, in particular personal data protection, consumer protection legislation and competition law, are fully respected;
- The rules for access to and use of data are fair, practical and clear, and there are clear and trustworthy data governance mechanisms in place;
- There is an open, but assertive approach to international data flows, based on European values.

Future actions will focus on:

1. Data spaces in key industrial and societal sectors: pooling and sharing of data in sectors identified as priorities (including, but not limited to, health, climate, environmental, manufacturing, agriculture, energy, financial and mobility data). The large-scale actions may include the creation of data platforms enabling secure and compliant sharing and reuse of sensitive, confidential, proprietary and personal data, as well as large-scale experimentation based on AI. Where relevant, the latter will take place in connection with the large testing and experimentation facilities mentioned below.
2. High-value datasets from the public sector: pooling, preparing and making available high-value datasets. This should lead to the availability of free and easy-to-use EU-wide datasets in areas such as geospatial and earth observation/environment and will include large-scale experimentation and AI use cases.
3. Developing incubators for aggregating demand for data assets and to bring together data providers, integrators, brokers, data users and service providers, especially small and medium-sized enterprises (SMEs). These will operate in coordination with the Digital Innovation Hubs network.

Many new business models emerge from the combination of data sources. Examples include just-in-time delivery of goods and the personalised treatment of diseases. Therefore, more access to data almost always means an acceleration of implementation and an increased accuracy in service delivery. The functioning of these European data spaces will depend on the capacity of the EU to invest in next-generation technologies and infrastructures as well as in digital competences such as data literacy. This in turn will increase Europe's technological sovereignty in key enabling technologies and infrastructures for the data economy. The infrastructures should support the creation of European data pools enabling Big Data analytics and machine learning, in a manner compliant with data protection legislation and competition law, allowing the emergence of data-driven ecosystems. These pools may be organised in a centralised or a distributed way.[7] The organisations contributing data would get a return in the form of increased access to data of other contributors, analytical results from the data pool, services such as predictive maintenance services, or licence fees.

The European strategy for data recognises EOSC as the nucleus for a science, research and innovation data space, which will progressively be articulated with the nine new sectoral data spaces foreseen by the strategy. These new data spaces will build on the ongoing EOSC experience gained with the research community. Therefore, there is huge opportunity to exploit EOSC as a flagship example of synergies between EU policies given its role in the renewed ERA, the European data strategy and, more widely, the European data economy.

2.4. From Horizon 2020 to Horizon Europe

Horizon 2020 calls helped to lay the foundations of EOSC from 2017 onwards. They have allowed the engagement of a wide range of research institutions across countries and communities and parallel research investigations to be run on a wide range of questions

---

[7] In the latter case the data are not moved to a central place in order to analyse them together with other data assets. The analytical tools come to the data, not the other way around. This makes it easier to keep the data secure and to ensure control over who accesses what data for what purposes.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 40 –

related to EOSC. The EC-grant approach has delivered a rich series of results such as use cases, demonstrations, data service tools and policy documents of direct relevance to EOSC. It is worth noting that major areas of work are still in progress and results are still becoming available in 2022.

Horizon Europe, the current EU Framework Programme for Research and Innovation, was launched on 2 February 2021 [Horizon_Europe]. Horizon Europe aims to:

- Develop, promote and advance scientific excellence, support the creation and diffusion of high-quality new fundamental and applied knowledge, skills, technologies and solutions, training and mobility of researchers, attract talent at all levels and contribute to full engagement of the EU's talent pool in actions supported under the Programme;
- Generate knowledge, strengthen the impact of research and innovation in developing, supporting and implementing EU policies and support the access to and uptake of innovative solutions in European industry, notably in SMEs, and society to address global challenges, including climate change and the Sustainable Development Goals;
- Foster all forms of innovation, facilitate technological development, demonstration and knowledge and technology transfer, strengthen deployment and exploitation of innovative solutions;
- Optimise the Programme's delivery for strengthening and increasing the impact and attractiveness of the European Research Area, to foster the excellence-based participations from all Member States, including low R&I performing Member States, in Horizon Europe and to facilitate collaborative links in European research and innovation.

Horizon Europe has brought a number of new features compared with Horizon 2020, such as, for instance, a mission-oriented strategy and an increased citizen involvement as a means to create more impact through the whole programme. Another of these new features and a key component of Horizon Europe will be Open Science. With its new framework programme, the European Commission aims to continue acting as a frontrunner in Open Access and Open Science. In Horizon Europe, the European Commission proposes notably that:

- Research data will be open by default, with exceptions in the cases justified in the Model Grant Agreement, thus following the principle 'as open as possible, as closed as necessary'[8];
- The development and implementation of a Data Management Plan (DMP) will become mandatory, even if not making research data open;
- Emphasis will be placed on supporting as much as possible the proliferation of research data that are as far as possible findable, accessible, interoperable and reusable (FAIR);
- Use of trusted repositories and infrastructures connected to EOSC will be encouraged and possibly required in some work programmes depending on the state of deployment of the EOSC-Core functions.

As concluded in the EOSC co-Programmed Partnership, EOSC can play a fundamental role both in contributing to achieve Horizon Europe's specific objectives, but also in supporting the implementation of the programme's Open Science features and in bringing evidence on Horizon Europe research outputs and underpinning the measuring of progress and evaluation of the difference the framework programme makes. At the same time, its domain-agnostic objectives to federate infrastructures and develop a web of FAIR digital objects brings new potential to contribute to the Horizon Europe missions, partnerships and clusters.

---

[8] For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".

The assessment of the first phase of EOSC implementation by the Commission and the Member States highlights the need to move from the essentially EC-grant-based approach of Horizon 2020 to a wide and inclusive partnership in Horizon Europe, with clear directionality and additionality characteristics at EU, national, community and institutional levels.

EOSC addresses the European/global challenge of excellent research in a context of data-intensive science. This challenge can be fully addressed neither by the EU alone nor by any Member State or Associated Country in isolation. Developing wider synergies between multiple EOSC stakeholders will be essential to realise the EOSC ambition.

The need for a continued R&I agenda supporting EOSC development in the period 2021–2027 has been acknowledged in several meetings of the EOSC governing bodies. A partnership approach to EOSC R&I was recognised as the best means to achieve the ambitions and policy objectives set for EOSC.

All thematic partnerships in Horizon Europe can derive benefit from a successful development of EOSC as this will equip them with minimal, rigorous standards and protocols and maximum freedom of implementation to share and reuse data and other digital objects across relevant domains of research. As demonstrated in the case of the COVID-19 crisis, rapid open sharing of data greatly accelerates research and discovery, allowing an effective response to society's need. Interaction with many of these partnerships, in particular, challenge-driven ones, will thus increase the potential to bring value to researchers in Europe and beyond and to underpin and consolidate a European Research Area that is fit for the digital age.
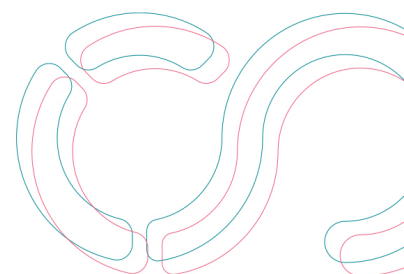
Use Case: COVID-19 – understanding the SARS-CoV-2 virus, its structure and societal impact

The ongoing COVID-19 pandemic has a profound impact on Europe's citizens. Faced with an unknown, dangerous and rapidly spreading pathogen, a massive research response was mounted to understand the novel virus and to develop clinical diagnoses and prognoses, treatments, preventive measures and vaccines. The impact of the pandemic has been felt in all aspects of society, and understanding these effects is of major importance for future pandemic preparedness. Whereas the focus has been on biological, clinical and epidemiological data, it is evident that research will enter a second phase with emphasis on social behaviour, economic measures and the effectiveness of vaccination programmes. Another aspect is how to improve information sharing and filter out fake news. For example, the effects of social media on information access and society response remain poorly understood, but disinformation – targeted and incidental – has been rife throughout the pandemic.
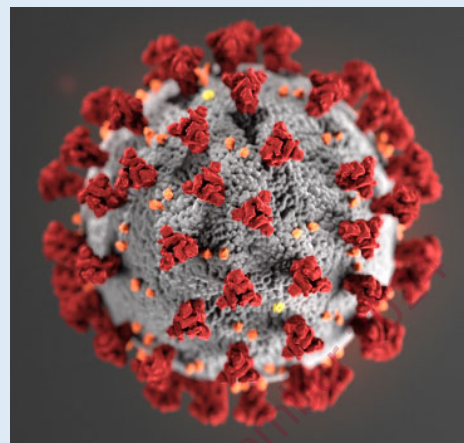
There is a need to combine and connect biological, genetic, epidemiological, health, social, political and economic data to monitor COVID-19. European researchers contributing to the response to the COVID-19 pandemic thus need to be able to store, share, access, analyse and process research data and other research digital objects across disciplines and national borders and to collaborate with global partners. Therefore, instead of setting up new data silos, the European COVID-19 Data Portal [C19_Portal] should be complemented with social and life sciences data to promote an integrated understanding of the European outbreak response and future preparedness, as well as to demonstrate the value of FAIR data to society and public engagement. The integration and management of health data is an enormous challenge that will have an unprecedented impact on personalised health.

Viral and human infectious disease data on SARS-CoV-2 and COVID-19 from national centres as well as clinical data should be linked across research fields, e.g. with contextual economic, social, cultural and migration data from the social sciences and humanities and in compliance with the security and protection of sensitive data collection and analysis. To bring together these data, the question of how to make the metadata of data objects interoperable needs to be addressed. Within the science cluster projects (ENVRI-FAIR, EOSC-Life, ESCAPE, PaNOSC, SSHOC), there are different metadata schemas in use based on strong and well-established domain standards (e.g. ECRIN metadata schema for clinical research, MIABIS for biobanks, DICOM for images, DDI for social sciences). The goal is to assess the metadata schemas and domain catalogues and to develop a strategy for how to map and relate the different schemas to support the search for data objects related to COVID-19.

Within EOSC, the science clusters will review models for automatic raw and intermediate data preservation and sharing, including software and analysis workflows; identify and share best practice; and link the catalogues with the emerging EOSC services and the EOSC Interoperability Framework. This work will align with, and enrich, the data currently being exposed within the European COVID-19 Data Portal. All science clusters will bring their experience and expertise to organise data from many different sources (including addressing multilingualism issues) and different types and modalities, so that they can be easily gathered, analysed and modelled holistically, and not remain fragmented as at present.

These combined analyses are still rare, due to lack of appropriate platforms, infrastructures and FAIR data. On one side, it is important to ensure the fulfilment of the ethical and legal boundaries ('as open as possible, as closed as necessary'[9]), although mechanisms for pseudonymisation and traceability do exist, enabling data sharing. On the other side, streaming and imaging data will require vast amounts of storage, network and computing capacity, and the same goes for multilevel social network analyses. The growing need for data storage and the large requirement of network bandwidth for sharing and analysis is a fundamental issue in advanced research across all domains which EOSC aims to address.

2.5. A history of EOSC

Open Science has been a policy priority of the European Commission since 2016 [EC_Open_Vision]. Together with Open Innovation, which will involve public and private sector actors in research to create new tools and services, and Open to the World, which will ensure involvement and open collaboration with non-European stakeholders, Open Science will open up the whole research process through digital technology. Open Science is a transformative driver that will shape the research and innovation policies for a renewed European Research Area. To further develop and implement the policies for Open Science, the European Open Science Policy Platform (OSPP) [EC_OSPP] was established as an advisory group consisting of stakeholders from the research community. The OSPP issued its final recommendations in 2020.

To enable the development and uptake of Open Science in Europe, the EC has proposed the creation of a European Open Science Cloud (EOSC). EOSC will essentially involve the federation of existing research data infrastructures and the realisation of a Web of FAIR Data and Related Services for Science, making research data interoperable and machine-actionable following the FAIR guiding principles [FAIR_Principles]. This web of data will allow researchers to find, exploit and combine linked datasets, providing a basis for artificial intelligence (AI) tools, leading to new discoveries and research paradigms. EOSC will initially focus on traditional research data but will also include research publications and research code. EOSC will encourage FAIR datasets to be made fully open, and will follow the principle of 'as open as possible, as closed as necessary'[10]. This is typically important for biomedical, military, sensitive, private and commercial datasets which may not be opened immediately or fully or indeed ever released.

In an initial phase of development from 2017 to 2020, the EC made a financial investment of approximately €320 million to begin building the foundations of EOSC through project calls in Work Programmes in Horizon 2020. This investment was targeted to develop a new pan-European access mechanism to public e-infrastructures, to coordinate related national activities, to connect European research infrastructures (RIs) to EOSC, to set up and begin the implementation of the FAIR guiding principles, and to start a FAIR-compliant certification scheme for research data infrastructures. These projects have involved the community of stakeholders of EOSC and have been steadily developing the broader EOSC ecosystem.[11] To help steer the initial development of EOSC, the EC appointed two high-level Expert Groups, which delivered recommendations on a vision for EOSC in 2016 [EC_EG1_EOSC] and on how to practically implement EOSC in 2018 [EC_EG2_EOSC], and an Expert Group on FAIR data, which offered recommendations on how to make FAIR a reality in 2018 [EC_EG_FAIR].

The initial development phase supported more than 35 projects, laying the foundations of EOSC and showcasing its diversity and complexity. The EOSCpilot project engaged extensively with stakeholders and proposed a governance framework and policies, as well as developing interoperability pilots across scientific domains [EOSCpilot]. EOSC-hub brought together service providers to create a single contact point to discover, access and use a wide range of resources for data-driven research [EOSC-hub]. The five ongoing cluster projects will connect the European Strategy Forum on Research Infrastructures (ESFRI) projects and landmarks to EOSC in the domains of environmental sciences via ENVRI-FAIR [ENVRI-FAIR], life sciences via EOSC-Life [EOSC-Life], astronomy and particle physics via ESCAPE [ESCAPE], photon and neutron sciences via PaNOSC [PaNOSC], and social sciences and humanities via SSHOC [SSHOC]. The five regional projects aim to coordinate the efforts of national and

---

[9] For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".
[10] For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".
[11] See Annex 1 of the landscape report for a list of EOSC projects funded under Horizon 2020 [EOSC_Landscape].

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 44 –

thematic initiatives in contributing to EOSC through groupings of European countries via EOSC-Nordic [EOSC-Nordic], EOSC-Pillar [EOSC-Pillar], EOSC-Synergy [EOSC-Synergy], ExPaNDS [ExPaNDS] and NI4OS-Europe [NI4OS-Europe]. HNSciCloud established a hybrid cloud platform to support high-performance and big-data computing through commercial procurement [HNSciCloud], work that is continuing through ARCHIVER [ARCHIVER] and OCRE [OCRE].

Finally, towards the end of Horizon 2020, a number of projects were awarded, i.e. EOSC Future [EOSC_Future], EGI-ACE [EGI_ACE], DICE [DICE], RELIANCE [RELIANCE], OpenAIRE-Nexus [OpenAIRE_Nexus], and C-SCALE [C_SCALE].

In 2022, nine projects were awarded under the first Horizon Europe INFRAEOSC call. Some started their work as of 1 June 2022. As of 1 October 2022, all nine projects, i.e. AI4EOSC [AI4EOSC], EOSC Focus [EOSC_Focus], EOSC4Cancer [EOSC4CANCER], EuroScienceGateway, FAIRCORE4EOSC [FAIRCORE4EOSC], FAIR-EASE, FAIR-IMPACT [FAIR_IMPACT], RAISE, and Skills4EOSC, have started. A first version of the Vademecum[12] was developed detailing the collaboration between the projects and with the EOSC Association, to maximise the impact of the projects for the EOSC ecosystem.

From the second Horizon Europe INFRAEOSC call, six projects have been awarded and are in the process of starting, i.e. GraspOS, CRAFT-OA, Blue-Cloud 2026, AquaINFRA, RDA TIGER, and SciLake.

2.5.1. Transition period 2019-2020

The initial phase of development for EOSC is tied to the funding programme of Horizon 2020, which comes to an end in December 2020. To direct the strategic implementation of EOSC, the EC published an implementation roadmap in 2018 detailing six main action lines to realise an architecture, data, services, access and interfaces, rules and governance for EOSC [EOSC_Roadmap]. This roadmap not only served the first implementation phase of EOSC in 2018−2020 under Horizon 2020, but also prepared for the second implementation phase of EOSC under the new funding programme of Horizon Europe for 2021−2027. The roadmap envisioned a pan-European federation of research data infrastructures built around a federating core, providing access to a wide range of publicly funded services supplied at national, regional and institutional levels, and to complementary commercial services. Lessons learned in the first implementation phase have shown that while the project-based approach is very successful in involving the many stakeholders and communities in developing the EOSC ecosystem, the individuality and freedom of projects has led to a fragmented landscape of systems and stakeholders.

With the aim of bringing the community together and ensuring a smooth transition from the first to the second implementation phase of EOSC, a three-tiered transition governance structure was established to run from 2019−2020 [EOSC_Gov]. The EOSC Executive Board, consisting of eight members representing organisations and three independent experts, advised and supported the strategy, implementation, monitoring and reporting on the implementation progress [EOSC_EB]. The EOSC Governance Board, consisting of representatives of Member States, Associated Countries and the EC, oversaw and supported the activities of the Executive Board and ensured an effective implementation of EOSC [EOSC_GB]. The Stakeholder Forum, consisting of the full EOSC community of organisations, projects and initiatives, allowed the collection of input and provision of feedback on the implementation of EOSC via events, online consultations, and the interactive Liaison Platform [EOSC_SF]. The EOSC governance structure was supported by the EOSCsecretariat.eu project,

---

[12] The September 2022 version of the Vademecum, presented at the Horizon Europe EOSC-related projects coordination meeting in Brussels, on 30 September 2022, is available at: https://bit.ly/EOSCvademecum22

which not only functioned as the governance secretariat, but also managed a co-creation fund for activities and proposals from the stakeholder community to co-develop and co-implement EOSC [EOSC_Sec].

The Executive Board identified priority areas for EOSC and created six working groups (WGs) consisting of experts from the EOSC projects and stakeholder community [EOSC_WGs]. WG Architecture was defining a technical framework to enable and sustain an evolving EOSC federation of systems, including application programming interfaces (APIs), authentication and authorisation infrastructure (AAI), and persistent identifiers (PIDs) [EOSC_WG_Arch]. WG FAIR was defining requirements for developing, assessing and certifying EOSC services in order to foster cross-disciplinary interoperability through FAIR [EOSC_WG_FAIR]. WG Landscape was mapping the landscape and readiness of existing research infrastructures in Europe that could be connected to EOSC [EOSC_WG_Land]. WG Rules of Participation was designing the rules to define the rights and obligations governing transactions between EOSC users, providers and operators [EOSC_WG_RoP]. WG Skills & Training was providing a framework for a sustainable training infrastructure to support the uptake of EOSC [EOSC_WG_Skills]. Finally, WG Sustainability was providing recommendations on the implementation of a scalable and sustainable EOSC, including business models, integration of national infrastructures, and legal models for EOSC [EOSC_WG_Sustain].

The activities of the Executive Board and WGs were steered by a Strategic Implementation Plan (SIP), which defined the background, vision, priorities and main goals of the Executive Board and WGs for EOSC [EOSC_SIP], and a work plan for 2019–2020, which set out the timeline, methods and delivery of key outputs of the Executive Board and WGs [EOSC_Work_Plan]. The overarching objective of the Executive Board was to provide recommendations on mechanisms and possible forms for governing EOSC in the second phase of implementation in 2021–2027 and to hand over all outputs to the new governance structure. In contrast to the first implementation phase of individual projects independently realising EOSC, the second implementation phase should have consolidated all project outputs and ensured directionality (through a common vision and objectives) and additionality (through complementary commitments and contributions). The transition governance bodies identified a Co-programmed European Partnership as the best instrument to overcome the fragmentation and to provide a framework for collaboration and pooling of resources at European, national, regional, and institutional levels. The motivation by all Member States to establish an EOSC Partnership in Horizon Europe was clearly expressed through the EOSC Governance Board since 2019. This reflected a general interest by the Member States to target the whole research ecosystem in Europe and not only the EU tier implemented through calls. Such a partnership would strengthen ownership by the research communities, achieve scale by aggregating demand by researchers and other users, and pool existing capabilities at European, national and regional levels. An EOSC Partnership can be seen as a means to obtain commitments to realise the EOSC-Core and expand it iteratively – possibly with new partners – to realise the wider, trusted and open EOSC distributed environment.

The transition governance bodies founded a new legal entity called the EOSC Association which was envisioned to have as members all relevant stakeholders in the EOSC ecosystem and to enter into a contractual arrangement with the European Commission to direct the Partnership under Horizon Europe (further information is provided in Section 2.7).

2.5.2. National infrastructures across Europe

The Landscape Working Group established by the EOSC Executive Board set out to survey and document the landscape of infrastructures and initiatives across Europe related to the development of EOSC. The work built on existing surveys and information provided by national authorities, various stakeholder communities and the relevant Horizon 2020 projects in close collaboration with the Member States and Associated Countries. Initial inputs included the

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 46 –
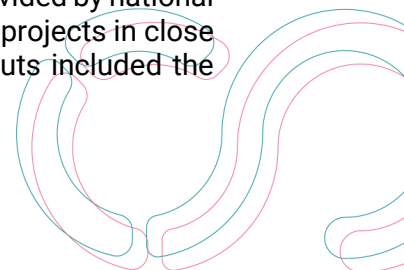
recent report of the e-Infrastructure Reflection Group, findings of the EC group of national points of reference, the surveys carried out by the OpenAIRE project, the EOSC-Pillar project, analysis of preliminary mapping of the UK's research and innovation infrastructure landscape, the experience of the ESFRI workshop on cross-disciplinary collaboration of ESFRI landmarks, other relevant documents identified by the WG members, and outcomes of the survey (country sheets) performed by the Landscape WG itself. The WG had collated inputs from e-infrastructures including data and high-performance computing (HPC) facilities, from European and national research networks, from pan-European infrastructures and ESFRI roadmap projects and clusters, and from supporting initiatives such as the Research Data Alliance (RDA). Information has so far been collated on 49 Member States and Associated Countries.

The WG had also surveyed the landscape of policy development across Europe regarding Open Science and EOSC. Most countries have research evaluation policies in place, as one would hope. The majority (61%) of the Member States and Associated Countries responding at the time had policies in place regarding open access to scholarly publications, but only 34% had a policy in place regarding FAIR data (though, encouragingly, 44% have one either in planning or under development). Few countries seemed ready to mandate that research data should automatically be made open. Relatively few countries (21%) mentioned EOSC in their policies, but 38% planned to do so in future; only three countries so far (Bulgaria, Denmark, and Romania) included mention of EOSC in their criteria for funding. More than half of the responding countries had nominated contact points for Open Science (53%) and for EOSC (42%).

Taken together, sources showed that there was a significant investment across Europe in e-infrastructures and data-oriented infrastructures. For EOSC to reach its full potential, these investments needed (and still need) to be either federated as part of EOSC or made accessible to users through EOSC. There has also been a significant, though not yet universal, adoption of policies towards Open Science and FAIR data. Although EOSC has not been very visible up until then, as part of national investment strategies, there was evidence that future policies and strategies will increasingly align around the concept. While it was proven hard to obtain definitive and quantitative data on national levels of investment, it was already clear that the bulk of these countries have significant investments in national e-infrastructures of one kind or another that could in principle either be federated as part of EOSC or made accessible to users through EOSC. The same applied to many of the data-intensive ESFRI landmark infrastructures. The scale of such investments over the past decade was already certainly in the billions of euros and hence much larger than the planned central investment in the EOSC core. This showed that EOSC will only reach its true potential through effective federation of national and research infrastructure resources.

EOSC as a sustainable collaboration system aims to link research data repositories and interconnect services and infrastructures. Research infrastructures are facilities that provide resources and services for research communities to conduct research and foster innovation. The landscape of European infrastructures was surveyed in this context, mainly focusing on the description of e-infrastructures and research infrastructures across different research areas. The survey of the landscape regarding EOSC-relevant infrastructures covered the following:

- **e- infrastructures**. e-Infrastructures address the needs of European researchers for digital services in terms of networking, computing and data management, and foster the emergence of Open Science [EC_OS] as an essential block of the ERA.

- **Networking and other services**. Each European country has a National Research and Education Network (NREN), connecting research and higher education institutions
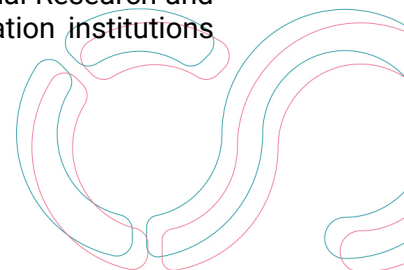
with high-performance networks, and offering a range of related services (e.g. eduroam).

- **Data infrastructures**. Data infrastructures consist of data assets supported by people, processes and technology and include the technical and human infrastructures that support management and sharing of research data.

- **Computing infrastructures**. The EuroHPC Joint Undertaking [EuroHPC_JU] is a legal and funding entity with the aim of developing a pan-European supercomputing infrastructure and supporting research and innovation activities by developing a European supercomputing ecosystem, stimulating the technology supply industry, and making supercomputing resources in many application areas available to a large number of public and private users. In several countries, EGI and EUDAT coordinate significant high-throughput compute (HTC) and data services at an international level based on a partnership model, while HPC centres generally join the PRACE partnership initiative and participate in EuroHPC. EOSC will bridge this separation and help address the question of the relation between centralised and federated e-infrastructures.

- **Thematic infrastructures**. Thematic infrastructures create a shared and collaborative research environment, known as the RI ecosystem, which has shaped big science for decades. Examples include the European Organisation for Nuclear Research (CERN), the European Southern Observatory (ESO), etc. RIs manage a large amount of data and have often triggered the development of data technologies and related policies.

- **RI clusters**. RI clusters are groups of RIs horizontally interlinked to be able to address globally important scientific and technological challenges. They have strong links with research communities and projects, manage significant data volumes and develop innovative data analytics tools, ensuring effective research data exploitation. Five ESFRI cluster projects have been launched in 2020, providing a focus for various ESFRI projects and landmarks to connect to EOSC. In general, the expectation of EOSC raised in the position papers of ESCAPE, PaNOSC, ENVRI-FAIR, EOSC-Life and SSHOC is that EOSC would enable the accessibility and reuse of research data and increase its scientific value.

The landscape-surveying exercise continued until the end of 2020. In its entirety, the activity indicated which infrastructures were to be considered to be the key elements of the future shape of EOSC. The differences among particular European states should be taken into account. This time-demanding process has made EOSC implementation gradual and dynamic. Links between the national thematic infrastructures and e-infrastructures, including data infrastructures, and national open access (OA) repositories, have been investigated. The readiness of the states depends upon acceptance of EOSC.

2.6. International dimension

EOSC operates in a global ecosystem with the clear aim, as already described above, to promote the 'Open Science, Open Innovation and Open to the World' principle in its international activities. The international dimension of EOSC is framed by the (i) regulatory framework, the Acquis Communautaire, (ii) Open Science culture, as well as (iii) the existing infrastructures and initiatives of the Economic Partnership Agreement (EPA) members.

Open Science is a new era in the evolution of science, which requires a cultural shift. It is driven by a number of organisations, both long-standing, who are in the process of adapting their methods to the new developments, and recently set up, via a bottom-up process. The transition and expansion of Open Science presents a constant increase in scale and scope for science at the local and global level, resulting in a peak in the need for investment, limited by public resources.

The current level of integration in the field of science in the European Research Area allows EU Member States and Associated Countries to share the burden of investment to achieve Open Science in the ERA. Moreover, the investment of the EU in e-infrastructure in recent years allows the opening-up of EOSC to third countries, based on shared values, principles and conscious choice. Given the different approach taken in the regions, EOSC will need to offer a tailor-made approach, taking into account local capabilities and demands.

The European Open Science Cloud is an opportunity to give fresh impetus to Science Diplomacy.

EOSC does not exist in a vacuum. Regional and national Open Research Data Commons and/or Open Science Clouds are being established concurrently. These developments enable the EU to enhance scientific cooperation with other parts of the world and drive Open Science culture based on commonly agreed values. EOSC operates in a global system which influences the world and is influenced by parallel activities from around the globe. There are major global trends which can be observed and groups such as the RDA Global Open Research Commons provide a useful forum in which to identify these and exchange lessons learned. Many international organisations such as RDA, CODATA, WDS and GO FAIR enable tighter collaboration between global initiatives, working together towards common goals for Open Science, thus driving global convergence on standards. At the same time, regional Open Science initiatives are getting more aligned and coordinated and there is a willingness for collaboration to avoid the creation of Open Science silos.

That said, EOSC has been enshrining a number of principles regarding international cooperation, with which potential partners could and should comply. While these rules and principles may be seen as a burden or an exclusionary tactic, in reality these ground rules enable a competitive, transparent Open Science ecosystem that enables quality science.

- **Data portability**. EOSC will not allow vendor lock-in at the EOSC-Exchange level and expects the same from services provided by third country partners.

- **Digital sovereignty**. Participation of third-country entities in EOSC is on a voluntary basis, but if they do participate it is expected that they will comply with relevant legislation and rules.

- **Ethics and values**. EOSC and the European research community represent certain values. EOSC recognises, however, that these might differ from those of other countries and is open to investigate whether its ethics and values should be reassessed in the context of globalisation.

- **Individual and community data autonomy**. EOSC condemns digital feudalism and supports the Global Indigenous Data Alliance (GIDA) and the CARE principles [GIDA; CARE].

- **Interoperability**. The EOSC-Core will provide an infrastructure with basic functionalities, such as persistent identifiers (PIDs) or authentication and authorisation infrastructure (AAI) services. Research outputs will have to comply with the FAIR principles and services will have to be FAIR-enabling. EOSC will use open source solutions but will of course make some technology choices regarding the fundamental functionalities. Third-country participants who wish to participate as a user or service provider will have to comply with these requirements.

- **Reciprocity**. Reciprocity is a principle enshrined in the future International Cooperation rules of the framework programme. International partners to EOSC should provide access to their National Open Science Cloud or similar, and to their service portfolio. This would enhance the free flow of (research) data and services.
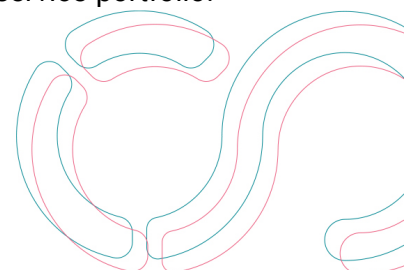
- **Security**. Third-country participants accept the cyber-security levels set by EOSC and commit to a Code of Conduct in the EOSC ecosystem.
- **Openness**. Third-country participants participate voluntarily in EOSC as users or service providers.

2.7. Strengthening the community

2.7.1. EOSC Association and EOSC Partnership

The decision towards a Co-programmed European Partnership on EOSC required the incorporation of a new legal entity able to enter into a contractual arrangement / Memorandum of Understanding (MoU) with the European Commission: **the EOSC Association**.

The **EOSC Association** was established on 29 July 2020 as a not-for-profit international association (AISBL) in Belgium, involving research and innovation stakeholders across the EU and beyond. The Association is open for membership and observership to organisations that embrace the vision and values of the Association and have a substantial and significant interest in, as well as can potentially contribute to and have an impact on, EOSC.

The EOSC Association plays a crucial role in gathering EOSC stakeholders such as research funders, policy makers, research-performing organisations and operators of research infrastructures to contribute to and monitor the future EOSC developments. The Association provides a single European voice for the purpose of advocacy and representation of all stakeholders in a collective manner. It facilitates communication, outreach and engagement with its members and observers, external service providers, research communities, stakeholder organisations and society as a whole to assure transparency, and promotes Open Science for the benefit of all.

The signature, by the EOSC Association and the European Commission, of the contractual arrangement / MoU marked the start of the Co-programmed European Partnership on EOSC under the Horizon Europe framework programme.

This **EOSC Partnership** brings together all relevant stakeholders to co-design and deploy a European Research Data Commons where data are findable, accessible, interoperable and reusable (FAIR). An open and inclusive Partnership will help ensure directionality (common vision and objectives) and complementary commitments and contributions at all levels. It helps to provide a framework to reach consensus amongst those committed to achieving results.

The EOSC Partnership aims to expand on the planned Minimum Viable EOSC (MVE) to create a growing ecosystem, bringing together relevant European initiatives around the FAIR data economy, fostering collaboration among those initiatives towards the objective of open research, attracting small and medium-sized enterprises (SMEs) and start-ups to use and benefit from the federated services and data sources, and raising awareness in society about the benefits of FAIR-data-driven innovation. The SRIA strategically defines the roadmap for implementing and further developing EOSC with the wider stakeholder community.

The EOSC Partnership develops and implements EOSC through financial, in-kind and policy commitments.
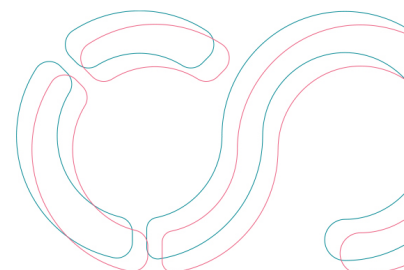
As a contractual document between the EOSC Association and the European Commission, the MoU sets out two tiers of commitments from its two signatories, reflected, for convenience, in this SRIA:[13]

- **Tier 1** is the contribution from the European Commission. Tier 1 will mainly include both financial and policy commitments. Through its work programmes, Horizon Europe mainly launches open calls for proposals for the EOSC community to develop and implement EOSC. Horizon Europe also mandates and stimulates policies supporting EOSC and Open Science, such as publishing in open access journals and making data FAIR. The EOSC Association supports the Commission by providing input to guide the calls and policy instruments for EOSC, mainly through the SRIA versions and Roadmaps, in the context of the EOSC Partnership Board.

- **Tier 2** is the contribution from the EOSC Association and its constituent entities, among which some are entities mandated by Member States or Associated Countries. Tier 2 mainly involves financial, in-kind and policy contributions from research-funding organisations, research-performing organisations, service-providing organisations, as well as other organisations that are members of the EOSC Association.

In addition, the EOSC Association makes a best effort to ensure contributions from other stakeholders in the EOSC community. These can be:

- Member States, Associated Countries and their respective national funding organisations which will directly finance research infrastructure developments in their respective countries that are compliant with EOSC standards and are easily integrated into EOSC. These financial contributions in the country contribute to building and growing the EOSC ecosystem. Member States, Associated Countries and national funders also develop policies that stimulate and support existing organisations in their countries to be as compliant as possible with EOSC and work towards sharing their research assets through EOSC. This includes policy changes that allow services to be made available across national borders, as well as aligning and supporting requirements for researchers such as data management plans, metadata standards, and making data FAIR. Such costs should be specifically earmarked for EOSC activities rather than general research activities in the country.

- Other individual organisations providing in-kind and policy contributions. Research-performing organisations ensure that their data is made FAIR and linked to EOSC via trusted repositories that are compliant with EOSC standards. They also train and support researchers in Open Science by, for example, hiring professional data stewards and mandating data management plans and FAIR data. They further promote Open Science practices and the use of EOSC among their researchers, as well as implement new systems that recognise and reward researchers for doing Open Science in career and grant evaluations. Service-providing organisations integrate their services with the EOSC-Core, abiding by the Rules of Participation for EOSC. This increases the value-added service offer of EOSC and ensures composability of services to fully address the needs of researchers. The Association supports these organisations by sharing current developments, policy recommendations and best practices.

One of the primary tasks of the Association is to continuously develop the SRIA, which shall influence future EOSC activities at institutional, national and EU level (including the EOSC-

---

[13] The SRIA Version 1.0 was initiated by the EOSC Executive Board under Horizon 2020 and has been developed collectively with the EOSC Community. This SRIA Version 1.1 is the result of integrating the MAR 2023-2024, ratified by the EOSC Community, into SRI v1.0. During 2023, SRIA Version 2.0 will be developed, in concertation with the EOSC Community and the European Commission.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 51 -

related work programmes in Horizon Europe). The SRIA is not a static document, but rather is envisioned to be a living document that will adapt to the changing EOSC ecosystem and the needs of EOSC stakeholders in future versions that will be updated by the EOSC Association with input from the European Commission.

The Association supports EOSC's mission of enabling seamless access to data through interoperable services that address the entire research data lifecycle in a number of ways. It identifies key infrastructure requirements for the representation, capture, storage, processing and appropriate sharing of diverse forms of data by engaging with stakeholders and service providers. It enables key services, including but not limited to e-infrastructures, to promote broad and secure access to data resources and data processing services, through its role in shaping the relevant parts of the Horizon Europe work programmes and in monitoring the output of funded actions. It coordinates and fosters technical environments and promotes the skills that enable the federation of existing and new scientific data infrastructures. The Association's Partnership with the European Commission and engagement with its stakeholders allows it to maintain alignment between the operations sponsored by the Association and the European Commission's Open Science strategy.

2.7.2. Governance of the EOSC Association

The Association encourages a broad spectrum of stakeholders to join EOSC, ensuring a balanced representation regarding types of infrastructural, organisational and sectoral members as well as geographic spread. This includes research data infrastructures, research-performing and research-funding organisations, researcher associations, and public and commercial service providers. Organisations based in EU Member States and countries associated with the framework programme for research can join as members, while other organisations are able to participate as observers. Organisations from widening countries and countries with limited EOSC involvement are especially encouraged to join as observers.

The Association is governed through three bodies: the General Assembly, the Board of Directors, and the Secretariat. The General Assembly is composed of the members and observers and is the supreme authority of the Association. The Board, led by the President, directs the activities of the Association by implementing the decisions adopted by the General Assembly. The Secretariat, led by the Secretary General, supports and advises the President and the Board as well as the General Assembly, and coordinates implementation of their decisions.

In addition to those bodies, an Expert Group of the EC, called the "EOSC Steering Board", representing Member States and countries associated to the Horizon Europe framework programme, was formed and is represented in the EOSC Partnership Board, together with delegates from the Commission and the Association. Additionally, the Expert Group provides policy and strategic-level advice to the Association and overall Partnership from the perspective of the Member States and Associated Countries.

2.7.2.1. Risk management

The results and recommendations of a targeted study, conducted by AON Hewitt on behalf of the EOSC Sustainability Working Group, were submitted to the governing bodies of the Association. They provided clear and structured guidance on how to incorporate risk management into the governance of the EOSC Association.

The study identified 48 gaps in the risk governance with respect to best practices and highlighted that EOSC operates within a **multiple factor environment** with a high degree of complexity affecting the governance structure. Risk factors include the organisational model, political influences, multinational and cross-disciplinary usage. Risk management activities for EOSC to date have been limited to individual project-based analysis and have therefore

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 52 –
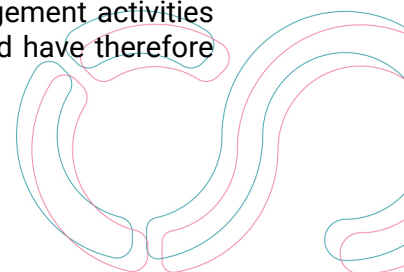
been fragmented, meaning a clear and defined risk governance structure with assigned roles and responsibilities for risk management needs to be established.

This study sets out 32 recommendations to address these gaps and ensure the effectiveness of EOSC risk governance. The main recommendations are:

- **Launch a comprehensive plan** to address these gaps and define a risk governance framework and organisation to support the structuring and development process of EOSC itself;
- **Establish a governance structure for risk management** that is clear and well formalised with appointed roles and responsibilities across the organisational structure. It will be necessary to clarify the responsibilities for the different actors involved;
- **Map the skills and competences** required to perform effective risk management at different levels of the organisational structure in order to consider all the fields of competence involved and set requirements on the composition of risk management bodies to **assure independence in decision-making**.

From an operational point of view, it is important for EOSC to **set up an infrastructure and data security team**, focused on the Minimum Viable EOSC (MVE), with responsibility to:

- Design a process that ensures the quality of the research data and data services;
- Design, update and share cyber security, business continuity and disaster recovery policy;
- Define a catalogue of potential risks (e.g. cyberattacks, business interruption, damage to data, failure of systems or applications, etc.);
- Improve technical resilience of the MVE by:
  - Performing specific business impact analysis and identifying the most relevant business interruption risk causes;
  - Establishing and updating the business continuity management plan;
  - Preparing and testing the disaster recovery plan;
  - Defining a set of guidelines concerning resilience, business continuity and disaster recovery for service providers.

The identified gaps and recommendations are considered by the EOSC Association, the EOSC contributing projects, and the EOSC Partnership overall in order to develop a comprehensive Enterprise Risk Management system (ERM). Implementing these recommendations significantly increases the value of EOSC and benefits its stakeholders by supporting its objectives and allowing a more effective use and allocation of resources. The ERM also helps to protect the assets, the corporate brand, the know-how of the key people, and optimise the operational efficiency.

2.7.2.2. Process

The EOSC Association coordinates the identification of needs for the development of EOSC and the SRIA, and provides input to all relevant stakeholders, including the EC. The EOSC Association coordinates input gathering from its members by through the Advisory Groups, inspired by the Working Groups of the EOSC Executive Board (2019 to 2020; namely Architecture, FAIR, Landscape, Rules of Participation, Skills & Training, and Sustainability) and on emerging topics. These Advisory Groups also engages representatives of organisations playing a role in the EOSC ecosystem that are not members or observers of the Association, to ensure a fully inclusive and coherent overview of the needs of all EOSC stakeholders. This approach makes it easier for organisations to learn about EOSC and consider joining the Association.

The Association directs the strategic orientation of EOSC and coordinates EOSC-related activities within its remit, including administrative, technical and communication roles. The

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 53 –
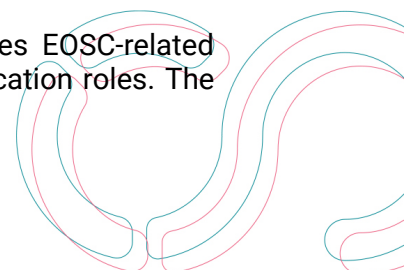
administrative role ensures management of the Association and involvement of members, observers and the Partnership Board. The technical role brings consensus and convergence in defining or contributing to the development and adoption of standards and good practices as well as monitoring the implementation of the SRIA. The communication and outreach role involves active and diverse communications and events, supporting user engagement and gathering feedback, and promoting EOSC results and success stories showcasing the added value of EOSC.

As previously indicated, the EOSC-relevant Horizon Europe work programmes are being, and will continue to be, adopted by the European Commission following relevant Horizon Europe comitology procedures. Calls for proposals have been and will be launched to implement those elements of EOSC where there is a need for pan-European collaboration and funding. In these cases, funding would be delivered mainly as grants to consortia of beneficiaries. The EOSC Association may coordinate and take part in projects (in particular coordination and support actions) in order to realise the vision and goals of the Association and the Partnership. It will not compete with the core activities of its members when participating in such projects.
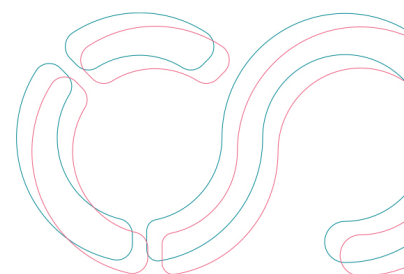
3 Strategic objectives of the European Open Science Cloud

The first two sections of this Strategic Research and Innovation Agenda have placed the European Open Science Cloud in the context of the digital age and of Europe's strategy towards Open Science and FAIR data. This section outlines the challenges that remain with regard to that legacy, and the role EOSC will take in alleviating them, specifically through the overarching objectives that must be achieved in order to realise the potential benefits for science and society.

3.1. EOSC Objectives Tree

Building the European deployment of Open Science requires addressing three main challenges relating to people (scientists and data professionals), knowledge (documents, data and software) and infrastructures:

- Convincing scientists that Open Science will allow them to do better and more rewarded research;
- Enriching publications, data and software in order to make them usable by machines and scientists;
- Federating infrastructures in order to make them all available to scientists across borders and across disciplines.

The EOSC Objectives Tree (Figure 3.1) presents these three challenges by stating the main problems, identifying the barriers, defining the objectives and highlighting the benefits.

European Open Science Cloud Objectives Tree

PROBLEMS	Public & private sectors do not exploit Open Science for improving quality and productivity of research	Researchers do not combine and build upon ever-growing available scientific results	National, European and global infrastructures do not share Open Science standards and practices
	PEOPLE	DATA	INFRASTRUCTURES
BARRIERS	Absence of incentives, rewards and skills for open sharing stifles the uptake of Open Science	Scientific results are unfindable, inaccessible, not interoperable, and often used only once	The scientific landscape consists of national and disciplinary research silos and infrastructures
	OPEN	FAIR	FEDERATION
OBJECTIVES	Open Science practices and skills are rewarded and taught, becoming the 'new normal'	Standards, tools and services allow researchers to find, access, reuse and combine results	Sustainable and federated infrastructures enable open sharing of scientific results
	SCIENCE	INDUSTRY	SOCIETY
BENEFITS	Improved trust, quality and productivity in science	Development of innovative services and products	Improved impact of research in addressing societal challenges

Figure 3.1: European Open Science Cloud Objectives Tree

The first release of the EOSC Objectives Tree was designed for the EOSC Partnership Proposal submitted to the European Commission in May 2020. It complies with the vision presented in the EC Communication 'A European strategy for data' in 2020 [EC_Data_Strategy]. New publications, data and software produced by laboratories, observatories, analytical,
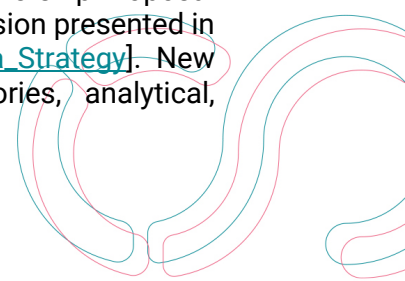
computational and scholarly work will progressively feed EOSC with quality-verified information sets ready for exploitation and reuse.

Europe has all the expertise needed to progress rapidly in the deployment of this EOSC ecosystem but it needs to bring additionality and directionality at European, national and institutional levels in order to direct future research and innovation efforts and stimulate deployment and adoption.

With the initial phase of the EOSC initiative ending in 2020, Europe now needs to strengthen and accelerate the development and implementation of EOSC, to engage more widely with multiple stakeholders, and to coordinate and synchronise the multiple relevant activities in the field that are still too fragmented among Member States' national plans and research communities.

The future of EOSC will be largely shaped by:

1. The **exponential growth** in the quantity of research artefacts: documents, data and software;
2. Science and innovation becoming **digital intensive**;
3. The evolution of **research infrastructures** towards managing digital knowledge;
4. The increased availability of **networking, computing and storage** resources;
5. The policy drive for **Open Science**.

EOSC, as a programme, will therefore be directed towards achieving the three overarching objectives defined in its Objectives Tree, each of which is discussed below.

3.2. Ensure that Open Science practices and skills are rewarded and taught, becoming the 'new normal'

A key goal of EOSC is to help move the research enterprise in Europe towards the Open Science paradigm. There is already a political will towards Open Science 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 [EC_Rec_2018/790].

EOSC will be established as a Europe-wide infrastructure for open research. The more scientists are convinced of the value of an EOSC federated infrastructure, the higher that value will be, following the network effect that led to the deployment and success of the World Wide Web using the internet.

When Open Science becomes the 'new normal', scientists will extend their requirements accordingly, and new roles and responsibilities will have to be created (e.g. data scientists, data stewards, etc.).

Scientists' rewards and recognition schemes will have to evolve also, to acknowledge that the value delivered by research is available in documents, data and software, extending the current rewards and recognition approach which is based too heavily on publications.

3.3. Enable the definition of standards, and the development of tools and services, to allow researchers to find, access, reuse and combine results

The launch of initial EOSC projects and the work of the Commission Expert Group on findable, accessible, interoperable and reusable (FAIR) data (with its report 'Turning FAIR into reality' [EC_EG_FAIR]) has allowed stakeholders to agree on the shared FAIR principles that are now at the core of EOSC [FAIR_Principles]. Making data and any other digital research artefact (such as documents, algorithms, tools and workflows) as FAIR as possible across all European research infrastructures will be a key expectation for joining EOSC.
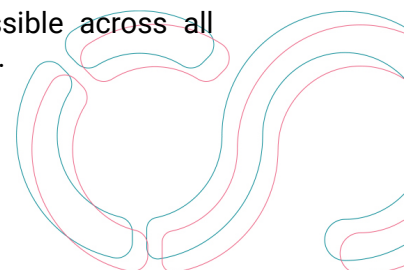
> The FAIR guiding principles for scientific data management and stewardship, by Mark Wilkinson et al. (2016):
>
> *Findable*. Data are assigned a globally unique, persistent and resolvable identifier. They are described with rich metadata which are registered or indexed in a searchable resource.
>
> *Accessible*. Metadata are retrievable by their identifier using a standardised communications protocol which is open, free and universally implementable.
>
> *Interoperable*. Data and metadata use a formal, accessible, shared and broadly applicable language for knowledge representation.
>
> *Reusable*. Data and metadata are released with a clear and accessible data usage licence. They are associated with detailed provenance and meet domain-relevant community standards.
>
> *'Importantly, it is our intent that the principles apply not only to "data" in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects – from data to analytical pipelines – benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.'*
>
> *'Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.'*

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 currently 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 research data.

Putting it in simple terms, the machine must be able to find data ('**know where it is**'), then to be able to access and identify ('**know what it is**'); in order to operate on the data the machine needs to know what can be done with this object ('**know how it can be handled**') and for reusing the digital object the machine needs to know what it is allowed to do with it ('**know which actions are allowed**'). This all needs to be well described in the metadata.

FAIR is the set of requirements that ensures that digital artefacts within EOSC can be discovered and reused. The FAIR principles articulate a set of mutual responsibilities between content creators and curators. Digital artefacts 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 artefacts to be discovered and retrieved using standard protocols, applicable for both humans and machines. Research communities need to define standards, sharing agreements and services to enable FAIR digital objects. Some, such as astronomy, life sciences and linguistics, have self-organised, but many others require support in order to narrow the gap between communities.

The strength of the FAIR principles is in defining a set of common characteristics required for all digital artefacts, irrespective of type, discipline and content. This enables machines to act

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 57 –

across a broad set of content, enabling interdisciplinary research. Many aspects of the FAIR principles, however, address 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, as noted in Recommendation 4 from the FAIR Expert Group: 'Develop interoperability frameworks for FAIR sharing within disciplines and for interdisciplinary research' [EC_EG_FAIR].

Research communities need to be encouraged to develop and maintain interoperability frameworks that define their practices for archiving, referencing and describing research artefacts of all forms. To support interdisciplinary research, these interoperability frameworks should be articulated in common ways and adopt global standards where relevant. Intelligent crosswalks, brokering mechanisms and semantic and other technologies such as artificial intelligence, should all be explored to break down silos and allow cross-disciplinary exploration, analysis and visualisation.

3.4. Establish a sustainable and federated infrastructure enabling open sharing of scientific results

This objective aims to enable the core functions of an operational EOSC ecosystem. EOSC is envisaged as a federation of infrastructures, forming a **Web of FAIR Digital Objects and Related Services for Science**. The FAIR principles and metadata standards act as guidelines for interoperability and facilitate maximum sharing and exploitation of research by the academic, private and public sector.

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**). 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.

(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. Specifically:

- A mechanism for naming and locating documents, data, software and services;
- A mechanism for discovery of and access to documents, data, software and services;
- A common framework for managing user identity and access.

The EOSC-Core will need to assemble a number of basic services and features, including:

- Repositories complying with an open charter that describe what users can expect from the service, such as descriptions of the content with rich, community-defined and FAIR metadata (including granularity levels, versioning policy), sustainability commitments, quality goals, etc.;
- Networking connectivity with commitments on upload and download capabilities;
- Authentication and authorisation rules and services for allowing access by users. These rules and services have to comply with the EOSC authentication and authorisation infrastructure (AAI) standards;
- Persistent identifiers (PID) services complying with the EOSC PID policy;
- Metadata services describing the content available in order, for example, to allow discovery by end users;
- Application programming interfaces (APIs) for access by machines. These APIs are necessary to allow the development of applications using the content. Their description must be public.
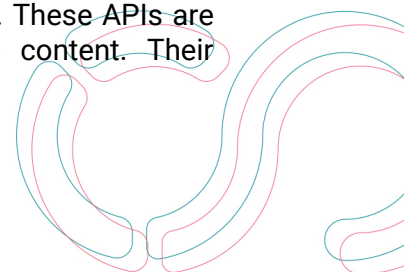
(2) The FAIR principles and metadata standards enable the **federation** of existing and planned research data infrastructures, adding a soft overlay to connect them and forming a Web of FAIR Data and Services.

As the national, European and international research data infrastructures composing EOSC and other regional infrastructures are by definition distributed, as well as supported by a wide variety of institutions (public and private) throughout the world, the envisioned EOSC can only be realised in a decentralised federated way. As described above, this requires an underlying framework based on commonly agreed, minimum standards and maximum freedom to operate with agility, whilst still ensuring global and interdisciplinary interoperability. This does not rule out multiple 'portals' in the sense of more traditional websites, where users can enter the EOSC environment, find content and related services, learn about commonly adopted approaches, formats, standards and EOSC Rules of Participation, register their resources, tools and services, etc. Currently, the projects funded under EOSC-related calls in Horizon 2020 have developed an initial EOSC Portal. In order to enable innovative value-adding services to be developed, it is essential that such access points have an API for machine access.

(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.

- **Common services**. This layer is composed of services that need to exist but may not be shared by all stakeholders. The main reason for such a layer is that certain domains or countries have already developed those services. There is no reason for them to change, while other domains or countries would benefit from using common services rather than developing their own. A good example is the archival service. All domains and countries need archival services to ensure the sustainability of their artefacts (publications, data and software). Some stakeholders have developed their own and have no reason to change. Their experiences may, however, be useful in developing common services for other stakeholders.
- **Thematic services**. This layer has no limit. It covers all the services that communities need to develop to contribute to the EOSC ecosystem. These services are delivered to researchers and all stakeholders to enhance their working environment. They are built using the relevant elements of the federating core (EOSC-Core) and may leverage common services. They will use the APIs mentioned above when necessary. Many projects are already engaged in such developments in vertical domains. The mission of EOSC is to allow those services to flourish and to support the ecosystem while stimulating the creation of new innovative services.

This section has explained the three overarching objectives that are driving EOSC and that will alleviate the challenges involved in delivering the European deployment of Open Science. The following section presents the guiding principles that are shaping EOSC and that will help position it within the Horizon Europe programme.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 59 –
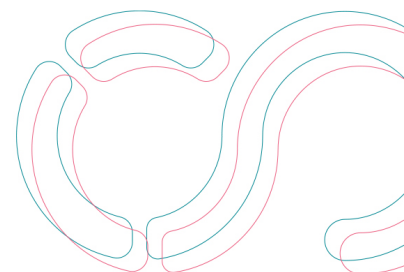
4 Guiding principles

4.1. Introduction

If EOSC is seen as the European endeavour of sharing research data, then this complements the European means to handle these data: the e-infrastructures in Europe. Data without e-infrastructures to store, compute and connect are of no use to EOSC and can only exist on paper or in the researcher's head. On the other hand, e-infrastructures without any data (only 'zeros' or 'ones') are meaningless. Data- and e-infrastructures form what can be thought of as a 'Yin-Yang' relationship. One is not possible without the other. Whether the whole of data-infrastructures and e-infrastructures should be called EOSC or whether EOSC is only a part of that is largely a question of semantics.
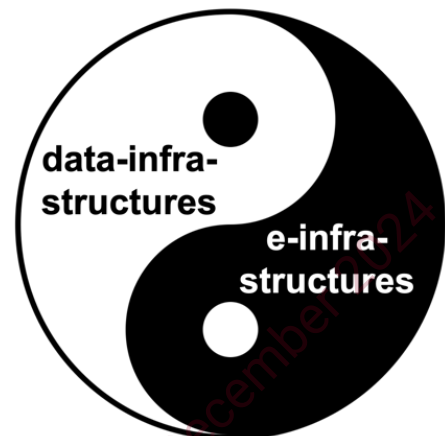
The overarching principle for developing EOSC is that research has to be at the centre of the EOSC initiative. Thus engagement with research communities is fundamental to understand their requirements and ensure that the way in which EOSC operates and the services are offered is of use and value to the researchers and respects the academic sovereignty of research data. Communities need to be consulted and encouraged to take an active role in the EOSC ecosystem. They need to represent the diversity of practice, such as research infrastructures, universities, data stewards, research software engineers, professional associations, research leaders and early-career researcher organisations. Close attention needs to be paid to the existing standards, infrastructure and support within research communities and EOSC will recognise and adapt, where possible, to enable these. The diversity in readiness levels to adopt FAIR and Open Science principles will be taken into account, including assisting those communities that are less advanced. A good relationship with research-focused stakeholders needs to be ensured. Researchers have to be present in the various EOSC governing bodies to ensure acceptance by research communities and assist in promotion and advocacy in relevant fora. Given the diversity of practice and readiness levels, there is a need to ensure this is understood, accepted and represented within EOSC. It may be hard to engage researchers directly, so EOSC often has to work through intermediaries who can represent their requirements and interests.

Within the overarching principle of the centrality of research and researchers, the way in which EOSC proposes to operate is defined by five further guiding principles.

Over the last five years, as EOSC was in the making, a number of shared principles have emerged from the work accomplished by the European Open Science community. High-Level Expert Group reports and results from first-generation pioneering projects have fuelled the debate among the EOSC community. From this debate, a set of five guiding principles has been agreed upon which will help position EOSC within the Horizon Europe programme during the next seven years. These are:

- **Multi-stakeholderism** – EOSC will succeed if and only if it follows a multi-stakeholder approach;
- **Openness** – EOSC will ensure that research artefacts are 'as open as possible, as closed as necessary'[14];
- **FAIR principles** – EOSC will assemble research artefacts that are findable, accessible, interoperable and reusable;

---

[14] For research artefacts, the interpretation of "as <u>closed</u> as necessary" is "as <u>restricted</u> as necessary".
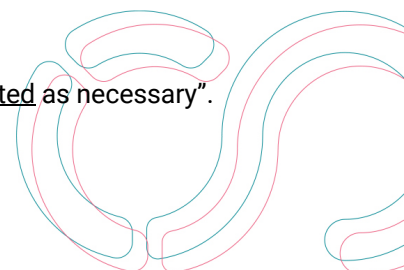
- **Federation of infrastructures** – EOSC will federate existing and upcoming research infrastructures (data- and e-infrastructures);
- **Machine-actionable** – EOSC will strike the right balance between machines and people in delivering the services that will serve the needs of European scientists.

The following sections report on:

- The roles played by the wide range of stakeholders of science (Section 4.2);
- The two key ingredients of Open Science: 'openness' and 'FAIRness' (Sections 4.3 and 4.4);
- The way to federate the efforts of research, data- and e-infrastructures to serve the needs of scientists (Section 4.5).

After recognising the role that machines will play at the service of scientists (Section 4.6), the final section (4.7) concludes with recommendations for research communities and policy makers on how to move towards the full implementation of Open Science, both culturally and technologically. Those recommendations can be seen as setting high-level requirements for the action areas that are presented in Sections 5 and 6.

4.2. Multi-stakeholderism

Today, all scientific communities generate growing numbers of research digital objects of all kinds, from raw data to publications, including workflows and software. Over the last decade or so, there have been significant investments across Europe in computer-oriented research infrastructures and e-infrastructures. The outcome is a vast quantity of infrastructure components of various scales and scopes, centralised or distributed, generic or domain-specific.

The challenge for EOSC is to federate this large variety of platforms at the subdomain, domain and interdisciplinary levels and to deliver an inclusive virtual environment to the European researchers. Most of these components have not been initially designed to work together. The challenge is not limited to linking datasets, federating infrastructures or aligning policies. It starts by linking multiple stakeholders – people and organisations – throughout the data lifecycle and across the EOSC ecosystem.

At the same time, EOSC intends to address common political priorities of the European Union and its Member States, to:

- Make Europe fit for the digital age;
- Interlink data spaces across a more efficient European Research Area;
- Make Open Science mainstream in the research community.

FAIR-by-design research outputs, combined with top-class digital infrastructures and artificial intelligence solutions, will ensure a true European research capacity to tackle the Sustainable Development Goals (SDGs), to reach the EU's ambition for the Green Deal and to implement other national or sectoral policies. Furthermore, its domain-agnostic objectives to federate infrastructures and develop a Web of FAIR Data and Related Services bring new potential to contribute to the Horizon Europe missions and clusters.

However, ensuring impact on these policy targets requires engaging further with a wide diversity and large number of stakeholders, across borders and disciplines, who are involved in the generation, storage, curation and processing of research artefacts, as well as in research policies, funding, skills and education.

The EOSC Partnership aims to embrace such a multi-stakeholder approach, to provide a framework of collaboration and to pool and align resources at European, national, regional and institutional levels. Starting from the current Strategic Research and Innovation Agenda (SRIA) document, a central task for this Partnership will be to develop, update and monitor a

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu   - 61 –
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

holistic SRIA supporting the EOSC vision. Developing and implementing such a SRIA requires the involvement of a wide range of stakeholders, including:

- Member States;
- Research-performing organisations;
- Research infrastructures and e-infrastructures (e.g. related to storage, computing and communications);
- Research libraries;
- Research associations;
- Research-funding organisations;
- Etc.

A more complete description of the EOSC stakeholders is provided by the EOSC Landscape report [EOSC_Landscape].

The approach to the implementation of EOSC during its initial phase has largely focused on EU-level activities[15] carried out through Horizon 2020 (H2020) projects. The consortia involved have brought together institutions of all sorts, from all over Europe and beyond, that have been developing and testing solutions along the six action lines described in the EOSC roadmap (data, services, architecture, access, rules and governance) [EOSC_Roadmap]. This EU grant-based approach has been successful in involving hundreds of European stakeholders across borders and communities, and has confirmed that the EOSC mission cannot be accomplished in a centralised manner but rather with a multi-stakeholder approach. In order to bring forward an initial operational EOSC capacity during the next phase of implementation, the EOSC Partnership will ensure directionality (common vision and objectives) and additionality (complementary commitments and contributions at EU, national and institutional levels).

The willingness of the EU Member States and Associated Countries to embrace this coordinated multi-stakeholder approach was expressed in December 2019 by the EOSC Governance Board and confirmed throughout the development of the proposal for an EOSC Partnership. This reflects a broad interest on the part of the EU Member States and Associated Countries in making EOSC evolve from a call-based approach to an all-encompassing ecosystem where the different stakeholders make the necessary commitments to contribute, and deliver outcomes, on the most suitable level of intervention (EU, national, institutional). Doing so, this multi-stakeholder approach should strengthen ownership by the research communities, achieve scale by aggregating demand from researchers and other users, and pool existing capacities and expertise at all levels.

All relevant research and innovation stakeholders, including scientific communities, research institutions, learned societies, community fora, national and international infrastructures (generic or thematic), funders (public or private) and industry actors (including data, software and journal publishers) are ultimately welcome to join the EOSC Partnership, if they agree to the Rules of Participation and adhere to the guiding principles, and to take part in the development of the present strategy and have their voice heard.

In summary, developing wider synergies between multiple EOSC stakeholders and ensuring systematic and structural collaboration between the EOSC stakeholders will be essential to realise the EOSC ambition. This has resulted in the proposal to create an EOSC Partnership [EOSC_PP] and is reflected in the strategy put forward by this SRIA, in which there is not a 'one solution fits all' to address all the gaps and priorities to achieve the EOSC objectives, but rather a coherent compendium of activities and deliverables that will take into account the following:

- The most suitable level of intervention (EU, national, institutional);

---

[15] Total EU investment of about €320 million in the period 2017–2020.
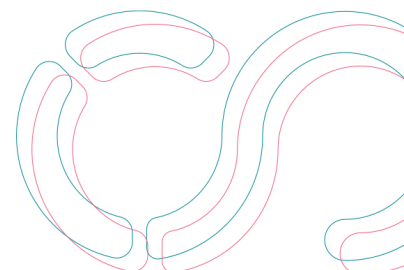
- The main targeted categories of actors (research-performing organisations (RPOs), research-funding organisations (RFOs), service providers, policy makers, regulatory agencies, research infrastructure (RI) operators, e-infrastructures, libraries, industry, etc.);
- The most suitable programme(s) (Horizon Europe, Digital Europe, Connecting Europe Facility (CEF), structural funds, recovery plan, plus non-EU programmes);
- The most adapted set of instruments, also considering the full range of research data infrastructure (RDI) activities from academic research to innovation (coordination, research projects, demonstration projects, strategic workshops, etc.);
- The range of outputs: new knowledge, prototype solutions, guidelines, standards, services, infrastructure, training material, curricula, coordination, etc.;
- The expected impacts, including support of the Commission priorities and relevant EU policies.

4.3. Openness: 'as open as possible, as closed as necessary'[16]

The rise of the digital age allows the ways research is conducted to change in multiple directions, with three main benefits:

- Delivery of better research results;
- Improved trust in research results;
- Development of multi-disciplinary programmes to address new societal and global challenges.

However, these improvements will only materialise if scientists evolve their practices and look ahead to share knowledge in ways that take advantage of the new capabilities offered by the digital revolution. At the core of these practices is openness. Scientists need to embrace the new approach, where knowledge is shared at all stages of the research lifecycle, as opposed to the old way, where results are shared primarily through publications made available when the work has achieved a sufficient maturity level.

4.3.1. Open access

The open access movement was born when scientists started to use digital technologies to share publications when they were still in preprint form (i.e. ready to be shared but not yet peer-reviewed). The time has come when this initial step can be followed by sharing not only publications but also all other research outputs such as data, software, workflows, etc.

Open access has been fully endorsed by the European Commission in FP7 and H2020 programmes, first regarding publications and then extending the mandate to research data, and it is set to stay as a best practice in knowledge sharing and communication.

Since 2012, when the EC Recommendation on access to and preservation of scientific information became available [EC_Rec_C(2012)4890], many Member States started discussing the need for establishing guidance and mandates on open access at the national level.

In addition, institutions developed their own open access policies. It is now very common, for example, for research funders to require open access to research outputs for funded projects, to monitor research impact and the return of investments in research.

Similarly, the Plan S initiative was launched in 2018 by a group of research funders [Plan_S]. It requires that scientific publications resulting from research funded by public grants must be published in compliant open access journals or platforms by 2021. This initiative is boosting the activities around open access publishing, in addition to fostering the discussion

---

[16] For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".

around roadmaps to establish open access and open science national plans and strategies in most EU Member States.

Despite the momentum behind open access, though, a monitoring mechanism to check policies' alignment and compliance with EC directives is still missing. Moreover, guidance is needed on issues such as security, privacy, property and sovereignty to ensure compliance between national and EC directives.

The coronavirus pandemic showed, even more emphatically, the need for appropriate licensing practices to mitigate exclusive rights in copyright law [LIBER_Copyright].

Last but not least, some cultural and technological barriers still exist. On the researchers' side, a strong bias still exists around open access publishing, which is often considered not comparable to traditional publishing. At the technological level, systems interoperability, enabling metadata exchange and improving dissemination and accessibility of research outputs, has improved recently, but there is still the need for further efforts to make the open access ecosystem a reality.

While it is clear for most stakeholders that Open Science practices will improve research results (by allowing scientists to benefit from each other's ongoing efforts), the work needed to realise the two other benefits – trust in science and multi-disciplinary developments – is still underestimated.

4.3.2. Trust in science through improved reproducibility

As the world has become more complex, as human knowledge has expanded in more and more disciplines, the role of science has increased while becoming more and more difficult to follow for any individual. It is therefore essential for everybody to be able to trust research results in order for science to deliver its benefits for society.

In order to build that trust in science, research has to be reproducible. Reproducibility is the ability for an experiment or calculation to be duplicated by other researchers working independently.

The reproducibility of science has been recognised as essential since the seventeenth century and the emergence of the scientific method. However, in the digital age, achieving reproducibility has become more difficult since computers have become part of the research lifecycle. Reproducibility of science requires reproducibility of software, as well as the availability of data and any other relevant information in machine-understandable form. Achieving reproducibility of science in the twenty-first century requires openness of software in all dimensions, not only source code but also knowledge of the computing environment.

Use Case: The interdisciplinary European coastline expedition (TREC) as a pilot for managing multi-modal data across institutions, disciplines and countries

Marine ecosystems are fundamental to life on our planet. Coastal ecosystems are particularly rich in diversity and abundance of both pelagic and benthic organisms. However, human activities are endangering coastal ecosystems on a massive scale. Climate change, in combination with habitat destruction, chemical and organic pollution along the coast and through river intake, and overfishing, represent major threats to coastal ecosystems.

In order to characterise marine and terrestrial ecosystems along the European coastline, a European coastline expedition (TREC) is envisaged to start in 2022. It will also extend to freshwater as well as important littoral and terrestrial coastal ecosystems such as tidal flats, rocky coasts and marshland. Organised as a highly cooperative and large-scale multinational initiative, several key European research infrastructures such as the European Marine Biology Resource Centre [EMBRC], the TARA Ocean Foundation [TARA], with its research vessel, as well as national infrastructures such as the French Ifremer [Ifremer] will be involved as

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 64 –

partners in this project. TREC will also cooperate with eLTER [eLTER] and multiple local partners who are invited to join the expedition through co-funded plug-in projects.

The TREC expedition will sample multi-modal datasets along the entire European coastline, involving and combining multiple kinds of -omics and imaging data as well as environmental metadata. The multi-disciplinary nature of the project, representing different scientific communities, research landscapes and cultures, will require huge data integration and storage efforts. Extended longitudinal sampling by local cooperation partners will produce expanding datasets over the years. The storage and management of multi-modal data across projects, laboratories, institutes, scientific disciplines and countries will be a huge challenge to the project. The sharing, joint analysis and diverse reuse of acquired data are main pillars of the TREC research strategy but will need adequate IT infrastructure.

This project responds to the challenges and threats that climate change in combination with habitat destruction, pollution, and overfishing pose to coastal ecosystems. A use case such as the TREC project will demonstrate the relevance of EOSC. In such a multi-disciplinary project of societal relevance, success will crucially depend on efficient sharing of data and processing pipelines at large scale, data standardisation, data access across domains, storage and service facilities, and accessibility to involved researchers and institutions, as well as other scientists.

Access to, linking and integration of data and metadata from different disciplines on the same ecosystem should be facilitated and accelerated to ensure interoperability of national and regional efforts. EOSC could be of help here, e.g. in the long-term cataloguing of data resources within EOSC. The expected outcomes of this use case will also be of value for activities around the European Green Deal, as cross-linking this data with socio-economic responses to changes in that ecosystem and other social science and humanities data promote an integrated understanding of the impact climate change and anthropogenic interventions have on European coastal ecosystems and the people living in and off them.

4.3.3. Facing global challenges through multi-disciplinary programmes

In a world that has become instrumented, interconnected and intelligent, it is possible to launch multi-disciplinary initiatives where scientists from different domains collaborate. In order to benefit from research artefacts coming from different disciplines, machines are used to allow computations optimised with diverse sources. It is therefore essential that research artefacts are both open and machine-understandable. This requires not only the sharing of data and software but also the sharing of metadata that describe the research artefacts. Openness needs to extend to 'information about information'. In many disciplines, efforts have been applied to design and archive ontologies that are becoming standards. In order to conduct multi-disciplinary projects, it becomes critical to develop crosswalks between metadata standards that will allow the matching of data representation designed for different domains. Openness of crosswalks themselves is therefore necessary to conduct multi-disciplinary initiatives and is a key concept in the EOSC Interoperability Framework.

4.4. FAIR guiding principles: making science transparent and reproducible

The FAIR principles were born with research data. Today, applying FAIR principles has to be extended to the whole research lifecycle, to ensure transparency, assessment, attribution and reproducibility. For this to happen, all outcomes of science, such as data, software and other digital outputs, have to be FAIR.

4.4.1. Web of FAIR Data and Related Services for science

EOSC is conceived as a Web of FAIR Data and Related Services for science. This is intended to highlight the interconnectedness of people, services and content. For research data to have context and meaning, its provenance, quality and usage need to be shared. Who created the data? For what purpose? How has it been processed? Can it be trusted? Detailed metadata

are required to enable discovery and reuse. The term 'Web of FAIR Data' is applied in its broadest sense, not just to data, but also to code, publications and other digital outputs. Services and stakeholders also need to be identifiable and well-described, with open metadata and persistent identifiers (PIDs) to allow cross-linking.

The FAIR ecosystem proposed in the 'Turning FAIR into reality' report [EC_EG_FAIR] highlighted the importance of registries for various components, in particular for policies, data management plans, identifiers, standards and repositories. Sustainable funding for core infrastructure is required to support the principles of FAIR and openness. Sociological aspects also play a key role in the evolution of the culture and practices necessary to implement and benefit fully from the Web of FAIR Data and Related Services, in particular incentives and rewards to increase adoption of FAIR across communities, and the building of the necessary skills and specialised workforce.

EOSC will be a federation of existing resources. It will of course give access to new data, but it will primarily be a federation of existing thematic data repositories and services, interfaced with existing data-sharing frameworks. To be adopted by the data providers and research communities, it should fit with their needs: resources should be able to interface with EOSC with minimal overhead, and the data and functionalities already available should remain, which implies that the EOSC environment needs to have different points of access depending on the end user.

4.4.2. Diversity of FAIR practices

Inclusiveness is therefore a critical element of success. FAIR is a journey, and research communities and data providers should be incentivised and supported to progress in this journey. The priorities for future work in implementing the EOSC FAIR framework should take into account the diversity of community FAIR practices and their different stages of preparedness. FAIR is a powerful concept, and its usefulness is demonstrated by the enormous impact of the principles on research policy globally. What these principles mean in practice, however, is still being defined, and recommendations for implementation have to be carefully tested in a wide diversity of contexts so that adverse consequences can be identified and corrected. Requirements need to be monitored and regularly updated.

Key strands of work were identified in the 'Turning FAIR into reality' report which set priorities for the implementation of a Web of FAIR Data that should be pursued on an EU level under Horizon Europe and in national and institutional funding cycles. These have been further validated by the EOSC FAIR Working Group and activities it has undertaken to assess FAIR practices across research communities and propose the EOSC Interoperability Framework. These include:

- Support for the development of community standards;
- Development of crosswalks between community standards;
- Adoption of semantic technologies and common standards for interoperability;
- Sustainable investment in registries of standards, identifiers and repositories;
- Certification of repositories to engender trust and enable FAIR;
- Tools to implement metrics and assess FAIR;
- FAIR skills and data stewardship competencies.

4.4.3. Community standards

Community standards are central to FAIR. There must be agreed formats for data, common vocabularies, metadata standards and accepted procedures for how, when and where data will be shared. Research communities need to be supported to come together to define these practices and standards. Some have already done so, but many lack the resource to do so as this work is often undervalued and not rewarded. If there is no investment in the definition of standards where these are currently lacking, then some communities will be unable to fully
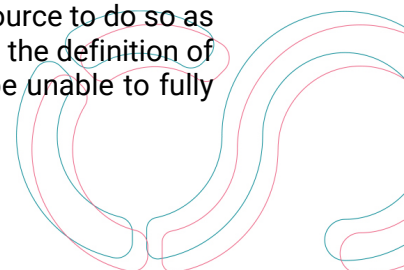
engage in the Web of FAIR Data. Levelling the playing field to enable broader cross-disciplinary research is a priority.

One aim of the FAIR principles is machine actionability. This is also, of course, a key aim of EOSC, but one that will not be fully implemented for all the resources from the start. Many disciplines, even among those that have been sharing data before the FAIR principles were defined, are not ready for interoperability. Others are less at ease with other aspects of the guiding principles. It is important to keep in mind that, as mentioned in Section 4.4.2, FAIR is a journey and that EOSC capacities will build up progressively. The 'machines in support of people' principle in the SRIA turns into a longer-term objective.

Cross-disciplinary usage of data and services is supported by the adoption of FAIR principles. In order to enable cross-disciplinary use within EOSC, the governance structures must first facilitate strong uptake of standards built on solid 'disciplinary pillars' which ensure data and metadata quality. Once these are in place, the initiative can then broker between existing thematic frameworks, enabling interoperability while also allowing the capabilities developed by the communities to be retained to fulfil their own needs. The development of use cases and implementation of the EOSC Interoperability Framework are priorities for the next phase of work.

4.4.4. Research artefacts sustainability

Not all data can be kept, all the more so if it has to be made FAIR. Appraisal is one of the archivist's tasks, and criteria have to be defined to support decisions on keeping vs. discarding data. The reproducibility of research results, the potential interest in and benefit of reusing the data, the data uniqueness (for instance, observations of natural phenomena over time), and the capacity to produce better data with current capacities, are among the aspects to be taken into account. Communities should be involved in the definition of the criteria and the decision-making process. Cross-disciplinary usage of data should also be taken into account, in particular by defining use cases to specify which data in particular have broader relevance and to avoid keeping all data by default 'just in case'.

Moreover, long-term open data archives and preservation services are required to enable a sustainable EOSC and the sustainable access to data. Data preservation not only refers to the long-term storage of data, but also includes ensuring the preservation and maintenance of data, as well as its context, understandability, interpretability, authenticity and integrity. The availability of long-term data preservation services represents an important added-value for EOSC but responsibility for the curation and management of datasets must remain with the communities. The interim findings of the FAIR Forever study,[17] conducted by DPC on behalf of the Sustainability Working Group, noted that digital preservation is not explicit in the context of EOSC and the roles, responsibilities and accountability for digital preservation are currently not clearly defined. The extent to which institutions have been given or taken explicit responsibility for preservation is unclear, assuming even that they have the capability to deliver. The concept of data stewardship at present, although it may imply preservation, is more often seen as an ambassadorial role, between the researcher and other institutional departments and staff such as the computing services, institutional repositories, libraries or archives. Clearer roles and responsibilities are needed, including the assessment of capability as well as functions, salaries and funding streams for preservation.

4.4.5. FAIR metrics and certification

The governance structure for implementing the Web of FAIR Data needs to work in close partnership with user communities. The usage of automated tools to test FAIR compliance is highly desirable for scalability, but these tools may have biases, and thorough tests and

---

[17] The study is ongoing; a report of its findings will be publicly available once the work is complete.

comparisons have to be performed, again in a variety of contexts, before they can be used for pass-or-fail evaluation. The FAIR metrics themselves also have to be subject to evaluation and iteration.

Repositories and other services enable FAIR by assigning persistent identifiers and supporting discovery and reuse. These services need to be robust and trustworthy, and existing frameworks for certification are being revised with FAIR criteria in mind. Support for services to self-certify is needed to strengthen the ecosystem and ensure the Web of FAIR Data and Related Services for science can be relied upon.

4.5. Federation of infrastructures

EOSC is expected to serve approximately 2 million researchers in Europe, many of them working at more than 800 European universities, and progressively to expand its user base to include the wider public sector and the private sector.

An EOSC that offers added value to researchers was taken as a starting point with its scope as described in the Strategic Implementation Plan [EOSC_SIP]:

'EOSC should be a federation of existing and planned research data infrastructures, adding a soft *overlay to connect them and making them operate as one seamless European research data infrastructure.'*
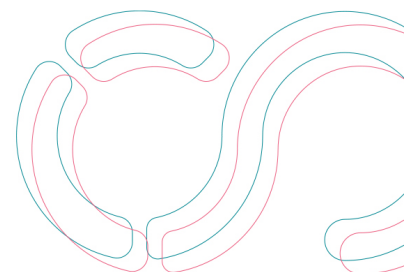
The gradual expansion to the public and private sectors will create solutions and technologies that will benefit all areas of society, e.g. science, economy and education [EC_Cloud].

Building on existing research data infrastructures, EOSC will grow through a series of iterations, as described in the FAIR Lady report [WG_Sustain_FAIRLady]. Each iteration will add more functionality and services for a wider user base and satisfy a broader range of use cases, gradually adding extra value to the end users. These added values include primary scientific advantages, such as enhanced data and service connections, a better ability to address interdisciplinary and societal challenges, and improved e-infrastructure services and tools for RIs and their data consumers.

On top of that, each iteration will offer political, social and cultural advantages, for example, advanced and improved political decision-making capabilities, increased societal awareness and gradual change in culture towards Open Science. Furthermore, persistent qualifying factors, such as transparency, high-quality data, research acknowledgement/credit and training, are important factors in each iteration.

4.5.1. First iteration – Minimum Viable EOSC

The FAIR Lady report describes the objective of the first iteration to bootstrap EOSC by establishing a *Minimum Viable EOSC (MVE)*. It will enable the federation of existing and planned research data infrastructures for the benefit of publicly funded researchers, to access openly available FAIR data and services. The *MVE* will include the *EOSC-Core* and *EOSC-Exchange*, described below, that work with the FAIR datasets to be federated via EOSC.
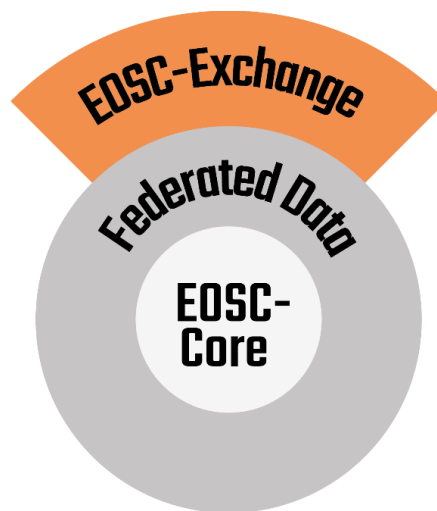
Figure 4.1: Schematic representation of key elements of the Minimum Viable EOSC

4.5.2. EOSC-Core

The *EOSC-Core* provides the minimum functionality that is required to enable Open Science practices across domains and countries. It supports FAIR data principles by providing the means to discover, share, access and reuse data and services. These elements address key technical, cultural and policy decisions of EOSC and they must be maintained over the long term. Specifically:

- A mechanism for naming and locating data and services;
- A mechanism for discovery of and access to data and services;
- A common framework for managing user identity and access.

While the *EOSC-Core* does provide frameworks to discover, share, access and reuse resources, it is the services federated via the EOSC-Core that actually transfer, store, process or preserve research data.

The initial implementation of the *EOSC-Core* will be based on the widely used production-quality components that have been jointly defined by the Architecture [EOSC_WG_Arch], FAIR [EOSC_WG_FAIR] and Sustainability [EOSC_WG_Sustain] Working Groups. These are already deployed by the EOSC-related projects and communities to provide the following functionality (more details are provided by the outputs of the Architecture and FAIR Working Groups):

- **A shared Open Science policy framework**, which effectively embeds a data compliance framework for open / FAIR data. It defines and applies the rules of how the data elements are published, shared and reused.

- An instantiation of the **EOSC Interoperability Framework**, including:
  - **Authentication and authorisation infrastructure (AAI) framework**, a trust and identity service for researchers to seamlessly access any EOSC resource.[18] The AAI framework implements the AARC Blueprint Architecture [AARC_BPA] to provide a set of interoperable building blocks for international research collaborations. The EOSC-Core includes those elements that provide identity interfederation for establishing trusted communications between identity providers and service providers. Community-specific AAI services are necessary components of the EOSC AAI architecture, but they are not part of the EOSC-Core.

---

[18] EOSC Resource extract from definition in the EOSC Glossary: EOSC Resources include services, datasets, software, support, training, consultancy or any other asset [EOSC_Glossary].
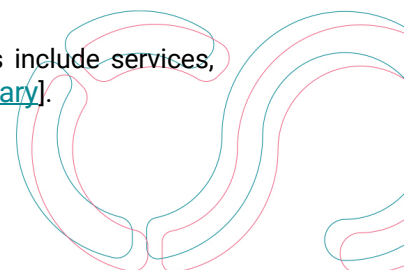
- o **Persistent identifiers (PIDs)**, services to generate, resolve and validate persistent identifiers.
- o **An interoperable metadata framework**, for ensuring openness and interoperability across disciplines while respecting privacy and security (copyright status, disclosure limitations, patents pending, other intellectual property rights (IPR) on the datasets or workflows, the existence of personal data, designation of data as Public Sector Information (PSI), etc.). Note that such a means of enabling interoperable metadata is a high priority for EOSC and is currently not addressed by the service providers consulted during the EOSC-Core Operational Costs Study.
- o **Data access framework**, whose primary role is to offer data as a service. It enables open interfaces where data consumers (users and machines) are able to discover and use data.
- o **Service management and access framework**, whose role is to provide a consistent and agreed-upon understanding of e-science services: what they offer, which science problem they address, what their operational capacity is, how they are accessed, who pays for them.
- o **An open metrics framework**, which sets the rules (usage, performance, value for money, user satisfaction) for the assessment of EOSC elements, i.e. policies, access framework, services, data, business, funding and usage models. This should include elements to facilitate the incentives and rewards mechanism for researchers, as recommended by the EC High-Level Expert Group on Next-Generation Metrics and the EOSC Pilot policy group [EC_NG-OS-Metrics].

- **Security policies and procedures** to ensure consistent and coordinated security operations across the federated services. This will include incident response policies and a service request and problem management scheme.

- **Operational support services** for the EOSC-Core and made available to those federating services connecting to the EOSC-Core. Support services related to the individual services accessible via the EOSC-Exchange or related to disciplinary data centres are not part of the EOSC-Core.

- **Web portal** with data and contents in multiple formats as well as supply-and-demand-facing services for accessing the EOSC resources. It is expected that other web portals will also exist and be developed outside of the EOSC-Core.

Building on the items listed above, the EOSC-Core will provide the means to operate the EOSC-Exchange as a digital marketplace of resources for publicly funded researchers.

4.5.3. EOSC-Exchange

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*.

Participation in the *EOSC-Exchange* as a service provider requires no registration fee. Service providers that do participate in the *EOSC-Exchange* will be required to conform to predefined Rules of Participation. While the technical requirements for participation in the *EOSC-Exchange* will be the same for all services, there may be differences in the legal and policy requirements for freely available and payment-based services.

A key objective of EOSC is to overcome existing national and disciplinary fragmentation in order to promote open research across Europe; consequently, the *MVE* should be as widely used as possible. However, it is recognised that controlled (authenticated and possibly

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 70 –

![EOSC logo]
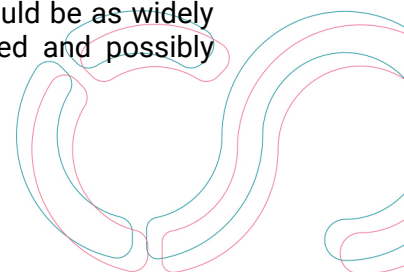
authorised)[19] access may be required in order to respect ethical, legal, social or commercial aspects and that the access policy for a resource may change during the research lifecycle. Such an access policy choice is a decision to be made by the resource provider but usage metrics should be tracked so that an impact assessment of EOSC can be made.

Licensing policies can also affect the adoption of the *MVE* and its impact. The *EOSC-Core* should have clearly defined requirements on the licensing policies of its components and their interfaces to ensure they remain openly accessible and cannot be controlled by a dominant party. Licensing policies for the federated infrastructures and contents of the *EOSC-Exchange* can be more tolerant in order to facilitate participation in EOSC but should still adhere to the FAIR principles.

4.5.4. Federated data and services

The Landscape Working Group established by the EOSC Executive Board has surveyed and documented the landscape of infrastructures, initiatives and policies across Europe relating to the development of EOSC [EOSC_Landscape]. Information has been collated on 47 European countries (EU Member States, Associated Countries and others). The country sheets [EOSC_Landscape_CS] and resulting Working Group report offer a snapshot of the state of play in 2020. However, it is clear that the types of information collected have potential value during the initial phases of EOSC implementation, to support ongoing monitoring of EOSC readiness and participation across different stakeholder communities. The findings were reviewed by the major stakeholders in the first validation workshop, which discussed a draft of the Landscape WG report 'Landscape of EOSC-Related Infrastructures and Initiatives'. While the participants of the validation workshop agreed that the country sheets have great value, they also felt that a more dynamic approach to populating them and keeping them up to date was needed. Though the disparity between the various countries is not as pronounced in terms of policies for data/services as it is for Open Science and FAIR data, it is still apparent that the landscape is very diverse in terms of available infrastructures. It appears that some Member States are currently in a more advanced state of EOSC readiness than Associated Countries. A detailed analysis is currently being prepared. With regard to specific references to EOSC in the policies of Member States and Associated Countries, 21% of respondents' policies currently mention EOSC while 43% state that this is in the planning stage. When it comes to funding for EOSC, the picture changes slightly to only two respondents' policies mentioning funding (4%) while 26% are in the planning stage.

The Landscape report also found that while many RIs (in particular the European Strategy Forum on Research Infrastructures (ESFRI) RIs) are leaders in data-driven science and are at the forefront of establishing good practice in relation to data science, there have not always been clearly defined data policies in place to govern the generation, management and sharing of research data. Several of the EOSC Cluster projects[20] are working to define common data policies. In addition, there is an apparent need for a wider range of stakeholders across the research ecosystem to be involved in providing and maintaining this key information. Securing participation from different stakeholders will be vital to ensure that the profiles can be refined to better reflect some of the potential indicators emerging across the various EOSC groups, ESFRI, the Member States and the European Commission, and are incorporated into the evolution of the EOSC strategy in a timely manner. It is desirable that as the EOSC ecosystem matures, the content of policies is considered in addition to their existence. This will stipulate the evolution of national research environments, as the harmonisation of RI data policies is a valuable step towards supporting EOSC readiness. A few countries noted that efforts are

---

[19] Authentication verifies you are who you say you are, while authorisation decides if you have permission to access a resource. If the policy of a service provider is to allow open access, then authorisation may not be required to access their services.

[20] [EOSC_Landscape] – section 3.5.1.

under way to establish national-level research data competence centres. Such initiatives could play a significant role in coordinating EOSC preparations across RIs in a national context and potentially have a key role in monitoring ongoing levels of participation and performance against emerging indicators.

Given that the landscape analysis indicates hundreds of infrastructure components available across Member States and Associated Countries that could potentially be federated, the EOSC Marketplace currently shows only a relatively small number of services per category. The reason for this is not clear. It may be the case that RIs that could provide services to EOSC have simply not yet completed the submission form required, and are waiting until they are sure of their readiness. While quality control measures are key for the longer-term delivery of the EOSC vision, it may mean that onboarding of services takes some time.

A current offer of services and resources is managed by the EOSC Portal [EOSC_Portal] via a Catalogue and Marketplace [EOSC_Marketplace]. In that sense, the EOSC Portal serves as a possible future entry point to EOSC services and resources from many domains by enabling users to access and request e-infrastructures services and data supplied at institutional, national and regional levels, enabling them to process and analyse data in a distributed computing environment [EOSC_Svcs&Res]. In order to provide a rich platform offering a wide range of services and resources, the development of EOSC clearly requires the participation of service providers. Services and resources are provided and maintained by different providers under the observance of legal frameworks and under a variety of licences and access requirements [EOSC_Providers].

In spring 2020, the resources listed in the EOSC Portal Catalogue are offered by 73 service/resource providers and aggregators. Taken together, they would give access to:

- 254 services;
- 4.4M datasets;
- 141K software and applications;
- 34.6M publications;
- and 3M other research products.

National and pan-European research infrastructures and RI clusters are quality and purpose assessed and horizontally interlinked to be able to address globally important scientific and technological challenges. They have strong links with research communities and projects, manage significant data volumes and develop innovative data analytics tools, ensuring effective research data exploitation.

The potential scale and diversity of the services and resources implies that the operational and financial responsibility for federated services and data will remain with their existing funders and cannot be transferred to a central EOSC entity. The investment in federated services and resources by Member States needs to be measured and acknowledged as an in-kind contribution to the overall EOSC funding model.

EOSC can provide an environment driven by societal challenges for public and private sectors to co-design innovative data-rich services and, in turn, increase Europe's technological sovereignty in key enabling technologies and infrastructures for the data economy.

In a second iteration, the MVE can be expanded with additional functionality and services dedicated to the requirements of end users from the public sector,[21] who are not involved in research activities but want to exploit open access to research data.

For example, EOSC can offer assistance to the public sector in relation to the implementation into national law by Member States of the Open Data Directive [EC_PSI] by July 2021. The

---

[21] In this document the term 'public sector' refers to all bodies governed by public law as defined in public procurement of services: Council Directive 92/50/EEC [EC_Procurement].
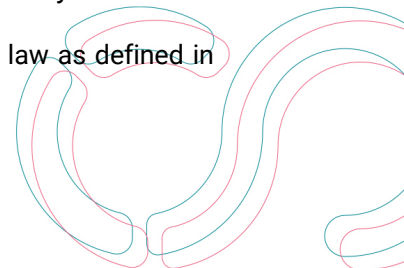
scope of the Open Data Directive includes research data resulting from public funding and focuses on the economic aspects of the reuse of information.

EOSC can also assist with the publishing of dynamic data, the uptake of application programming interfaces (APIs) and address the transparency requirements for public–private agreements involving public sector information, avoiding exclusive arrangements. The monitoring functions of EOSC could also help Member States identify high-quality datasets associated with important benefits for the society and economy.

Enabling the private sector to make use of EOSC resources greatly increases the potential for innovation and economic impact of EOSC. Therefore, in a third iteration, the MVE can be expanded with additional functionality and services dedicated to the requirements of end users from the private sector, so that they can exploit the FAIR data and associated services for commercial gain without distorting market competition.

4.5.5. Future Outlook

For EOSC to be a success, it must be widely adopted by researchers. This implies that EOSC must provide access to services that allow researchers to pursue their research activities more effectively through faster and seamless sharing of publications, data, software and other digital research outputs.

While the services to be provided to researchers via EOSC are expected to be *free at the point of use*,[22] they are not without significant cost to build, maintain and operate.

Researchers are practically minded and will only adopt EOSC if it provides interoperable services that make their research practices simpler and more effective, i.e. they need to be easy to use and need to support all phases of the research lifecycle. Readily available training and documentation, employing the latest digital learning tools, will be needed to reduce the barriers to adoption.

Therefore, the basic condition of success in ensuring EOSC sustainability is performance: how EOSC, as an ecosystem, operates and how the resources are provided, used and acknowledged by the users. Awareness needs to be raised among the EOSC stakeholder community of what is in EOSC and what is not, at all levels of the Interoperability Framework: technical, semantic, organisational and legal [EOSC_IF].

In order to gradually achieve interoperability of the services and compatibility of the data federated via the EOSC framework, standards and interfaces are needed, and the current activities and plans should put emphasis on developing those standards and interfaces step by step. This may involve revisiting and adjusting the datasets and e-infrastructures involved in the ongoing EOSC-related projects.

> **Use Case: OpenBioMaps –a discipline-independent tool developed to foster relationships between science and applied fields**

---

[22] *Free at the point of use* does not imply *free of charge*. *Free at the point of use* means the end user does not pay directly for the service when it is delivered, but their consumption will be paid for by other means. For example, an end user would not need to use a credit card to pay for a service but their employer may receive an annual bill from the service provider, or the employer may have arranged a suitable subscription.

OpenBioMaps (OBM) maintains an open and free biological database service and develops software applications for handling biological data. It offers scientists and conservationists a customisable toolset that facilitates easy access and management of data. By combining data from different fields and offering easy-to-use analysis tools, it widens uptake beyond academic research to applied fields and citizen science.

In the majority of use cases, OpenBioMaps builds a connection between conservation and conservation biology. Data may be used to designate protected areas, track bird populations in cities, or to build road safety prediction models as a public service for some road sections. The data are also used in several citizen science and educational projects.

OpenBioMaps has recently been enabled to use distributed computing resources from the EGI federated cloud through an EOSC service as part of the EOSC Early Adopter Programme [EOSC_EAP] to develop and maintain a service layer that allows OBM projects to run scientific analyses. As part of this programme, OBM is developing an interactive web interface that enables collaborative work on developing analyses. The project is also developing a portable computational package format (Docker container) which lets projects share analyses together with their environment so remote computational servers can run the analysis through these packages.

The adoption of services available in the EOSC Marketplace will allow OBM to widen its service offering to users (researchers in conservation biology and ecology), enhancing the analysis capability and enabling greater cross-disciplinary applications of the data.

Additionally, standard ways of calculating costs should be created for services that also include margins and returns for service aggregators and other 'intermediaries' that are needed to de-risk the quality of EOSC services and cohesion between operators.

Further consideration should be given to procurement processes (such as pre-commercial procurement and others) that could be used in order to eventually ensure that EOSC itself is able to buy all administration (and not only technical) services that it would need to survive. Procurement would be the mechanism determining the issue of intellectual property (IP) developed by EOSC projects. For projects funded through a grant mechanism with EU funding, the IP resides with the beneficiary who has generated the results (e.g. a university). However, it will be important that retention of results is with EOSC itself, to ensure its sustainability as well as the trust of the user community in the EOSC ecosystem.

Finally, the success of EOSC depends not only on sound funding models encompassing the financial, legal and governance aspects to create added value for the stakeholders but, in accordance with preliminary feasibility investigations, also on the incentives and rewards for researchers that encourage them to participate in a culture of sharing the results of their research. Without such incentives and rewards it is possible that the uptake of EOSC could be jeopardised by lack of engagement from researchers.

All of this will take time and cannot happen overnight. Implementing the set of iterations described in this document will take the EOSC schedule far beyond the end of 2020 and extend it over the full length of Horizon Europe, as depicted in Figure 4.2 below.
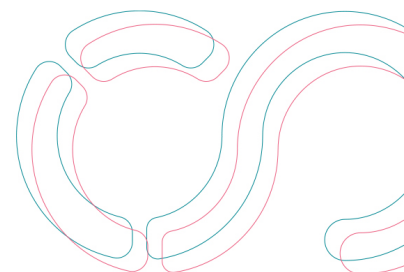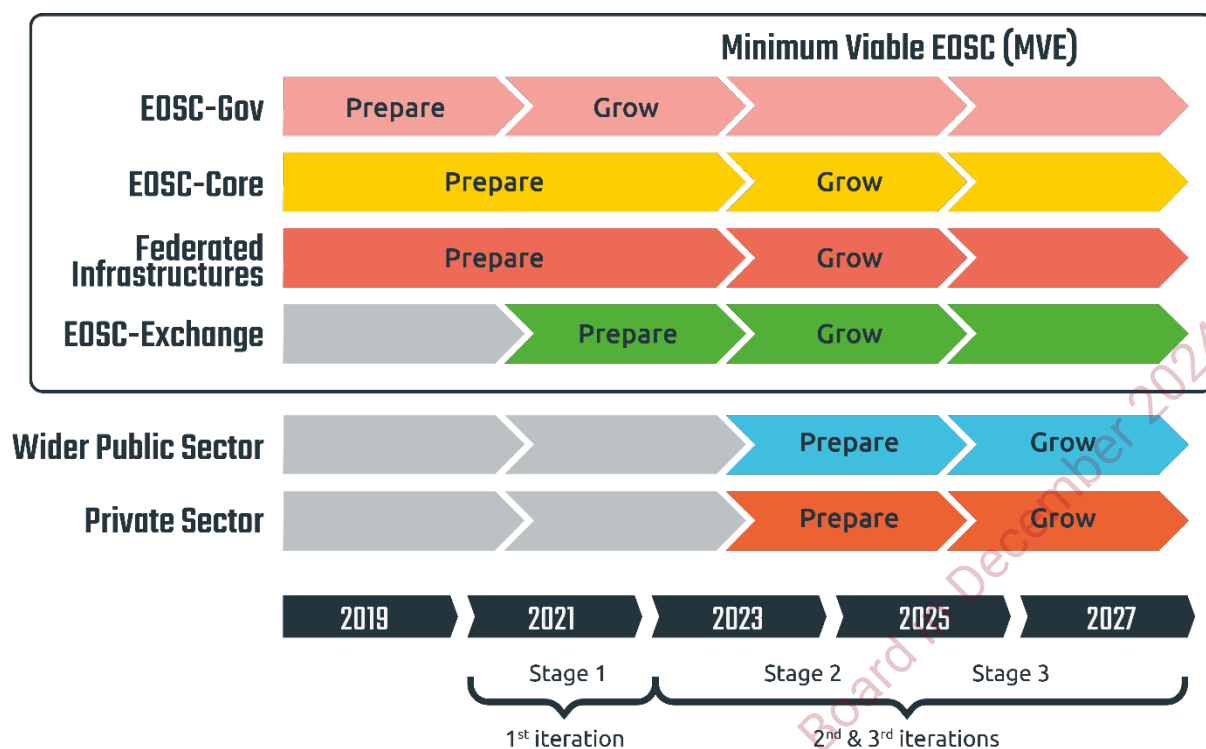
Figure 4.2: Schematic representation of timelines of EOSC iterations

Consequently, it is recommended that a transition period of 3 years (2021–2023) be anticipated to establish the MVE, building on projects to be funded via INFRAEOSC-03-2020 and INFRAEOSC-07-2020 funding calls, and by means such as open calls related to the EOSC Partnership, other EOSC-related projects as well as contributions committed by the EOSC Association members.

As of 2024–2025, EOSC will gradually open up to end users beyond the research community to develop and deploy services that will serve society at large, with significant contributions from the private sector. Specific activities and the timeline will be further elaborated taking into account the results of the previous period. It is expected that EOSC deployment will create market opportunities for new innovative companies to engage in the deployment of Open Science. The EOSC Partnership will also address the differences in economic development in the research and innovation sector by creating equitable access to data and services from both users and providers. Researchers and innovators will be able to jointly create innovative new technologies and services, which in turn will lead to the creation of new jobs and markets. The education, training and support needed to develop the necessary expertise will be facilitated by the use of virtual, shared environments.

4.6.   Open Science services: machines in support of people

The rise of the digital age creates new avenues for the development of Open Science, improving knowledge sharing between scientists. Digital technologies also allow new challenges related to the abundance of research outputs created around the world to be faced. Scientific activities have grown in volume and complexity in many ways. Machines are needed to help scientists face these new challenges.

The volume of scientific results produced every day has grown significantly. Even within a single discipline, it has become impossible for any scientist to read all the publications related to her/his research. When it comes to multi-disciplinary research activities, the scope of knowledge is beyond reach for a single individual; teamwork is no longer an option.
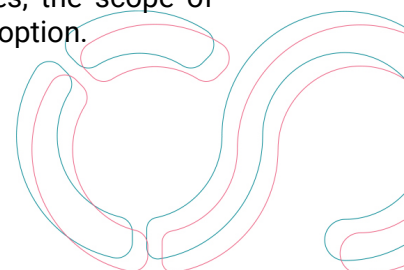
As the deployment of the internet extends the scope of research artefacts to publications, data and software, the volume of information available can no longer be managed by a research team.

As an obvious consequence, research can no longer be done without the use of machine-driven systems (hardware and software).

EOSC has to help scientists exploit those systems to perform their activities. Table 4.1 highlights the variety of systems that are commonly used and positions EOSC as offering an integrated view of those systems by federating existing infrastructures.

4.6.1. Digital systems for Science

Systems/Users	Hardware	Software
Individual Scientist	Personal Workstations, Tablets, Smartphones, Specific Devices, …	Generic & Specific Applications Generic Software (e.g. operating systems, programming languages environments), …
Research Team	Computing & Storage Servers, Specific Equipment, …	Databases, Shared Repositories, Shared Applications, Shared Libraries, …
Research Organisation	Large Computing & Storage Servers, Large Specific Equipment, …	Development Platforms, Shared Repositories, …
Research Infrastructure	High Performance Computing, Very Large Storage, High Performance Equipment, …	General Purpose Applications & Platforms, Very Large Repositories (publications, data, software)

Table 4.1: Commonly used systems, of which EOSC offers an integrated view by federation

4.6.2. Hardware

Digital hardware is managed like other research equipment. Sharing hardware has become common practice within laboratories, universities or research centres. Infrastructures allow resources to be shared at national or thematic levels.

The deployment of EOSC requires sharing of resources across borders and across disciplines. Nowadays, thanks to the availability of internet-based infrastructures, the technical aspects of sharing resources can be addressed. The challenges to achieving the EOSC vision with regard to these hardware resources will mostly be at the legal, financial and organisational levels.
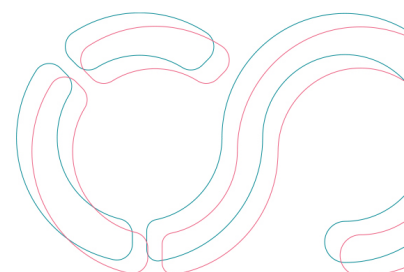
The multi-stakeholder approach is essential to address these challenges. Agile agreements, shared funding models, deployed Rules of Participation are therefore foundational for EOSC to deliver its full potential.

4.6.3. Software

In order to implement machine actionability, software is used at multiple levels of the research environment, which can be split into two categories:

- Research software: software used by scientists themselves to manage experiments, collect data, exploit results, check hypotheses, etc.
- Infrastructure software: software used to manage infrastructures at the service of scientists.

Each of these categories is described below, together with the software services on which the EOSC-Core itself is based.

4.6.3.1. Research software as research artefacts

In order for digital systems to help deliver their value to scientists, research artefacts have to be machine-actionable. As described in Section 4.4, research data have to comply with FAIR principles in order for digital systems to be able to find, access and reuse those data. Sharing research publications also benefits from the deployment of digital services built on top of the World Wide Web, which was introduced originally as a Web of documents.

While publications benefit from the Web of documents and the tools and practices that have been developed over the last thirty years, and research data benefit from special attention after the emergence of the FAIR principles, research software has started to receive attention only in the last few years.

Research software does not benefit from similar opportunities for a variety of reasons:

Research software has received recent attention

It is common for scientists to evolve research software in order to conduct derivative research initiatives. Therefore, software, as for any other research artefacts, has to be archived, referenced and described in order to be reused. Reproducibility of science requires the availability of the exact software version that has been used by prior experiments.

Publishers have started to include research software in their repositories. Data repositories have started to include software next to their datasets. From these pioneering efforts, a systematic approach to managing research software has to be put in place. During its transition phase, EOSC has recognised this situation. A task force on 'Scholarly Infrastructures for Research Software' has been organised by the Architecture Working Group and has delivered a report covering the state of the art, best practices and open issues, workflows and use cases before offering recommendations for next steps [WG_Arch_SIRS].

Research software benefits from generic software environments

On the other hand, software benefits from the open-source movement, which has been under way for decades. Open source allows software source code to be shared. As a consequence, one of the main characteristics of research software is that code reuse is considered normal practice.

Moreover, thanks to the deployment of the internet, cooperative software development efforts are improved by the use of software development platforms. It has therefore become possible to harvest open-source code and build open-source software archives. Also, software development platforms use version control systems which allow the software evolution to be archived. Users can therefore retrieve the exact version that was used to produce the research results.
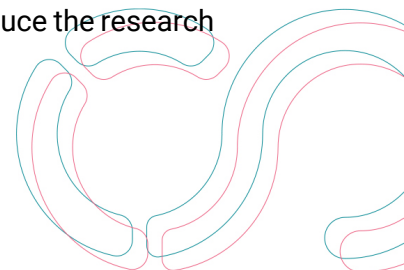
Research software needs metadata description standards

Research software is now recognised as playing a key role in research activities, as described by the CodeMeta project [CodeMeta]:

> *'Research relies heavily on scientific software, and a large and growing fraction of researchers are engaged in developing software as part of their own research. Despite this, infrastructure to support the preservation, discovery, reuse, and attribution of software lags substantially behind that of other research products such as journal articles and research data. This lag is driven not so much by a lack of technology as it is by a lack of unity: existing mechanisms to archive, document, index, share, discover, and cite software contributions are heterogeneous among both disciplines and archives and rarely meet best practices.'*

The deployment of the World Wide Web has provided the opportunity for the development of general schemes to describe information as shown in Figure 4.3. Research software developments have created their own schemes.
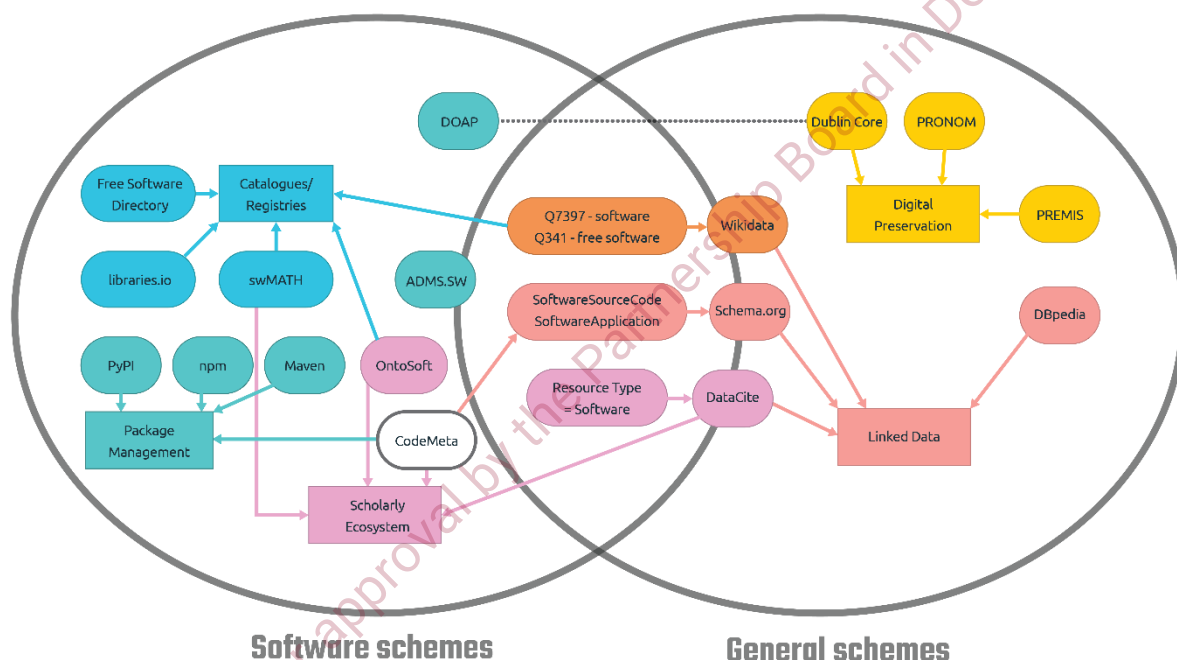
Figure 4.3: Software ontologies landscape derived from Pathways for Discovery of Free Software [Gruenpeter & Thornton] CC-by-4

After the creolisation period that has occurred recently, it is time to agree on metadata standards for software source code. Work is under way and EOSC will be able to both contribute to the standardisation and benefit from it.

4.6.3.2. Infrastructure software as service delivery

Over the last decades, a wide variety of research infrastructures has used the availability of new delivery models to develop new-generation infrastructures for the benefit of scientists. Those infrastructures are organised at national and thematic levels.

The Cloud Computing paradigm for sharing resources has developed and research infrastructures now have the choice when it comes to delivering their value to end users. Services can belong to one of the Cloud Computing layers as described in Figure 4.4.
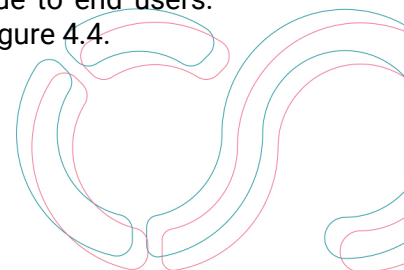
Cloud Computing Layers
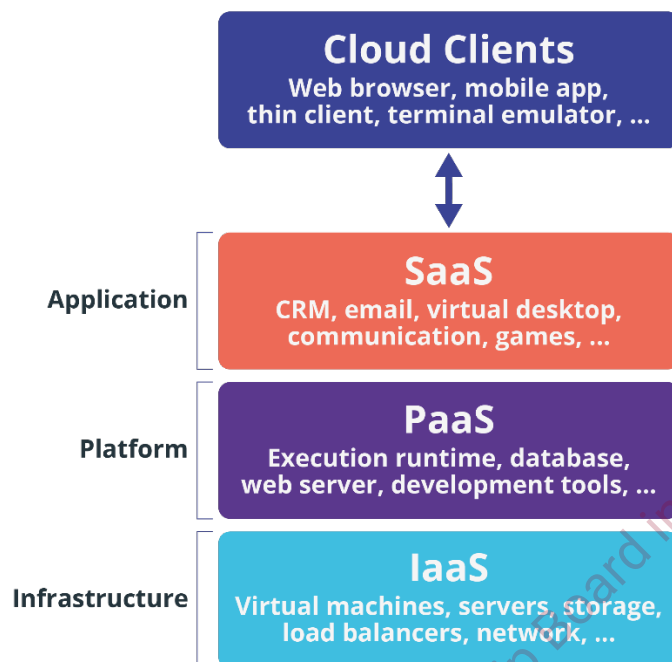
Figure 4.4: Cloud Computing layers (from Wikipedia)

When designing a new Cloud offering, the way resources are shared can belong to Cloud Computing types as shown in Figure 4.5.

Cloud Computing Types
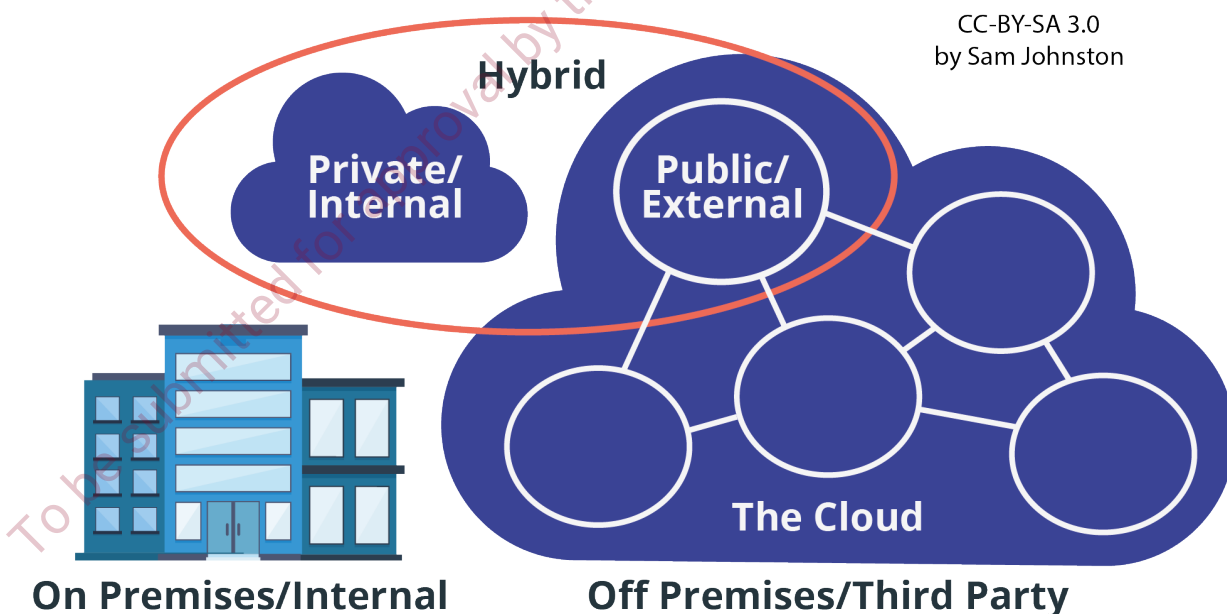
Figure 4.5: Cloud Computing types

There is no 'one size fits all' model for research infrastructures. Depending on the purpose of the infrastructure, the appropriate model will be chosen. As a consequence, EOSC will have to federate very different infrastructures.
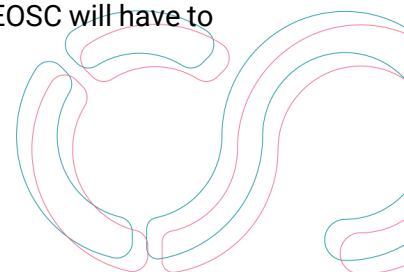
EOSC will be faced with the challenge of hiding the complexity and the diversity of services to the end user by providing a simple-to-use environment. Rules of Participation to the EOSC-Exchange are essential for resource providers to best deliver their value.

4.6.3.3. EOSC-Core federating

Last but not least, while as lightweight as possible, the EOSC-Core itself will be based on software services. In order to deploy EOSC services in a controllable manner, special attention has to be given to managing the software involved in implementing EOSC-Core functionalities.

The EOSC-Core needs to be exemplary in terms of openness at all levels:

- Open source code;
- Open interfaces;
- Open protocols;
- Open standards;
- …

While a reference implementation is critical to bootstrap EOSC deployment, the evolution of EOSC should be driven by innovation practices and allow multiple implementations to be welcome.

4.7. Recommendations

Starting from the guiding principles, it is possible to highlight recommendations for research communities and policy makers, to move them forward from the current state of the art towards an Open Science scholarly communication ecosystem that is based on, incentivises and facilitates Open Science principles and practices in performing and sharing science.

Research communities should:

- Normalise their Open Science processes (standards);
- Regulate them (policies);
- Facilitate their implementation (guidelines and frameworks, e.g. information models that describe flows and elements);
- Make sure their thematic services embed Open Science aspects by design (roadmaps).

The aim is twofold: to make the future scientific process

- As rigorous and automated as possible (e.g. services to FAIR-publish all outcomes on behalf of researchers);
- As transparent and reproducible as possible (e.g. tracking provenance, services, researchers, data, software, relationships, etc.).

Scientific communities should share a common understanding of the research products they manage, how these are semantically related, and how these should be published in order to maximise their discovery, access and reuse. For example, the concept of 'experiment' should be published, with all the elements necessary to ensure its reuse, replicability, reproducibility and repeatability by others.

In order to ensure widespread benefits of EOSC, improvements in Open Science practices are necessary. The first essential step is for the communities to develop a shared understanding of their internal needs for Open Science practices. Shared understanding could, in turn, motivate the development of agreed methodologies, standards, tools, policies and infrastructures. For example, generalising the deployment of FAIR data is a goal that cannot be achieved in one leap. Rather, it is a journey and each step, even a small one, is essential and valuable.
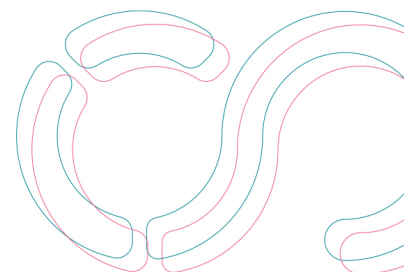
The EOSC FAIR Working Group investigated FAIR practice across disciplines and drafted a comprehensive study with recommendations [WG_FAIR_6Recs]. These acknowledge the importance of community practice and of devising a flexible architecture and set of rules in EOSC which facilitates uptake by all research groups. The recommendations echo previous priorities identified in the 'Turning FAIR into reality' Expert Group report, and are summarised in Table 4.2.

	EOSC	Research Funders	Institutions	Policy Makers	Coordination Fora	Standards Bodies	Data Service Providers	Publishers
1 Fund awareness-raising, training, education and community-specific support	✓	✓	✓					
2 Fund development, adoption and maintenance of community standards, tools and infrastructure	✓	✓			✓	✓	✓	
3 Incentivise development of community governance	✓	✓			✓			
4 Translate FAIR guidelines for other digital objects	✓	✓		✓	✓	✓		
5 Reward and recognise improvements of FAIR practice	✓	✓	✓	✓				
6 Develop and monitor adequate policies for FAIR data and research objects	✓	✓	✓	✓				✓

Table 4.2: Overview of recommendations and the stakeholder groups to which they apply

These recommendations are further developed below, indicating the key stakeholder groups tasked with applying each recommendation, and providing a short rationale and practical examples.

4.7.1. Fund awareness-raising, training, education and community-specific support

Stakeholders: EOSC, Research funders, Institutions

Rationale: Community-specific actions are needed because arguments and solutions that work for one community might not be the key drivers for another. Raising awareness is needed at all levels – from individual researchers through heads of institutions to policy makers – but in order to be meaningful it must be based on adequate, community-specific arguments. Awareness-raising, training, education and providing dedicated community-specific support take time and effort and thus such actions need to be financially supported. Funding pilot projects might be a useful mechanism to facilitate this.

Example: An initial pilot at Delft University of Technology (TU Delft) to fund data stewards with disciplinary knowledge helped communities realise the importance of FAIR practices, foster best practices and prompted them to appoint their data stewards as permanent members of staff [Plomp_2019]. Funding similar pilots could help other communities see the value of FAIR practices and drive the internal need for improvement.
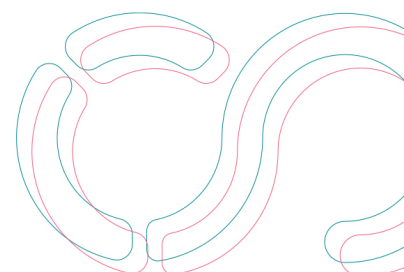
4.7.2. Fund development, adoption and maintenance of community standards, tools and infrastructure

Stakeholders: EOSC, Research funders, Coordination fora, Standards bodies, Data service providers

Rationale: It is difficult for communities to work without funds, on a best effort basis. The development of standards, methodologies and tools takes commitment and time.[23] However, this phase is essential for putting FAIR principles into practice. While it is important that community members actively contribute to standards development, leading such work requires dedicated resources. Funding of adoption efforts is also crucial, in order to avoid unnecessary overproliferation of standards and to facilitate alignment and interoperability between various communities. Implementation of standards also requires appropriate methodologies, tools and infrastructure (e.g. databases, repositories), tailored to community needs, and the development of these also needs to be funded. Standards, tools and infrastructure also have to be sustainably maintained and regularly revised to avoid depreciation, and this can only happen if communities see the value of such standardisation, are incentivised to do such work, and receive the necessary funding for this.

In addition, it is crucial that communities, especially those less experienced in FAIR practices, have access to people with expertise (for example, data stewards or ontology experts), who can help with development and adoption of standards and methodologies, provide best practice recommendations or case study examples, and offer tailored training. Such efforts have to be appropriately and sustainably funded and research institutions should be encouraged to take long-term responsibility for the availability of such support roles.

Examples: The Joint Programme on Wind Energy of the European Energy Research Alliance (EERA JPWind) received funding from the European Commission which allowed it to lead concentrated efforts that culminated in successful development of taxonomy and metadata for the wind energy sector [Sempreviva_2017].

Initiatives such as the Wellcome Trust's Open Research Fund,[24] or the EOSC Co-Creation Fund [EOSC_CCF], provide, amongst others, financial support for activities that aim at improving FAIRness of community practices.

The Research Data Alliance [RDA] is an example of an overarching coordination forum which plays an important role by offering a framework for communities who wish to work together, outputs to support standards development (e.g. FAIRsharing [FAIRsharing], which is a curated resource on data and metadata standards), or providing recommendations on best practices from various communities [RDA_Recs].

4.7.3. Incentivise development of community governance

Stakeholders: EOSC, Research funders, Coordination fora

---

[23] Those who successfully developed standards often cite years to ensure sufficient community consultation and co-development.

[24] For examples of projects funded by the Wellcome Trust Open Research Fund, see [Wellcome_ORF].

Rationale: Standards need to be developed by/with the community for them to be accepted and successfully implemented. For this to happen, clear community governance is essential to determine responsibilities and oversight of the different processes and to ensure a structured way of communicating feedback. Such efforts should be incentivised financially (e.g. the costs and time required to organise community consultations).

Examples: Astronomy is a discipline with strong community governance. The standard data format for astronomy was developed in 1981 and has been maintained by the International Astronomical Union [IAU_FITS]. The International Virtual Observatory Alliance (IVOA) develops and maintains the technical interoperability standards for astronomy. The IVOA does not have any formal funding, but benefits from in-kind contributions of community members [Genova_2017], which highlights the importance of advocacy and bottom-up level buy-in for such initiatives to be sustainable.

The wheat research community is an example of a community that used the framework offered by the Research Data Alliance and created a dedicated Wheat Data Interoperability Working Group to facilitate development of best practice standards in a structured manner (clear leadership of the group, clear ways of working and of providing community input, clear timelines and goals) [Dzale_2017]. The agriculture community set up an Interest Group at the early stages of the RDA which coordinates the discussion on future developments and Working Groups, and liaises with disciplinary international organisations such as the Food and Agriculture Organisation of the United Nations (FAO) [FAO] and Global Open Data for Agriculture and Nutrition (GODAN) [GODAN].

4.7.4. Translate FAIR guidelines for other digital objects

Stakeholders: EOSC, Research funders, Policy makers, Standards bodies

Rationale: Applying FAIR principles to the context of specific communities requires adoption/translation. This need is more obvious in the case of other (non-data) digital research objects where a direct mapping of the FAIR guiding principles may not be appropriate. The importance of each principle may depend on the priorities and maturity of the community in their use of certain research objects. This translation will need to be agreed in appropriate community fora, and such efforts should be incentivised financially (e.g. the costs and time required to organise community consultations).

Example: As part of the American Geophysical Union's (AGU) Make Data FAIR project [Enabling_FAIR] to enable FAIR data across the earth and space sciences, town-hall meetings [AGU_TH43B] and panels [AGU_U41A; AGU_IN41A] have addressed the challenges of making other research objects FAIR, including software, samples and workflows. This is beginning to lead to community-specific guidance around metadata and citation practices to improve software and service findability, accessibility and reusability [Hausman_2019].

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 83 –

4.7.5. Reward and recognise improvements of FAIR practice

Stakeholders: EOSC, Research funders, Policy makers, Institutions

Rationale: Efforts aiming at improvement of community FAIR practices are usually time-consuming and require a lot of dedication. Nevertheless, such efforts tend to be unnoticed in the current academic rewards system, unless linked to journal publications. To incentivise such work and to highlight its importance, it is essential that it is appropriately recognised and taken into account in evaluation, promotion and hiring criteria. This is a shared responsibility that needs a concerted approach between Institutions, Research funders and Policy makers at various levels. In addition, it is crucial that the needs of the most vulnerable communities, such as Early Career Researchers, are emphasised in the process. EOSC should play a supporting role.

This should go beyond merely recognising the time and efforts needed to make individual research outputs FAIR. Efforts aimed at greater community engagement, such as development of shared standards for FAIR practices and of the infrastructure, are crucial and need to be recognised as well. Furthermore, incentivising and rewarding FAIR practices should not be pursued in isolation, but rather be embedded in the broader discussion on responsible academic assessment and its role in improving the academic culture by, among other things, making room for the transition to Open Science, strengthening research ethics and integrity, and promoting a broad range of academic activities that goes well beyond the current focus on journal publications.

Examples: There are multiple examples of efforts undertaken by Research funders, Policy makers and Institutions towards better rewarding and recognising researchers for making individual research outputs more FAIR. The final report of the Open Science Policy Platform [OSPP_Report] offers a comprehensive set of recommendations for various stakeholder groups, reflecting the broader discussion on responsible academic assessment of which it is part. The Open Research Funders group developed the 'Incentivization Blueprint' [ORFG_IB], which provides concrete recommendations with a template specifically for research funders.

FAIRsharing is a resource that gathers community standards and credits record maintainers. However, the EOSC FAIR WG was not able to identify concrete examples where efforts aimed at improving FAIRness of community practices (thus, at a higher level than just making individual outputs FAIR) were explicitly mentioned in academic rewards and recognition policies. Interestingly, recommendations that such activities should be rewarded have been already articulated in the 'Turning FAIR into reality' report (Rec. 4, Action 4.1 and Rec. 6, Action 6.2) published in November 2018 [EC_EG_FAIR], suggesting that implementation of these recommendations did not happen and should be prioritised.

4.7.6. Develop and monitor adequate policies for FAIR data and research objects

Stakeholders: EOSC, Research funders, Policy makers, Publishers, Institutions

Rationale: Policies can be important drivers for FAIR data [Digital_Science_2019] and other research objects (software, workflows, models, protocols, etc.). Therefore,

it is essential that bottom-up, community-based efforts are coupled with top-down, policy-driven approaches. Policies should be developed collaboratively (ensuring that all relevant stakeholders are included [Stoy_2020]), they need to be explicit (e.g. clear roles and responsibilities, FAIR vs. open data, purpose and effects of FAIR metrics [Dillo_2020]), aligned with each other, and aligned with community practices and other relevant policies and regulations (e.g. research integrity). This applies to policies of Research funders, Publishers and Institutions. Proper implementation, monitoring and suitable incentives are also essential for the effectiveness of such policies. Implementation should be coordinated with institutional actors so that demands are not coming into effect without appropriate support and common understanding of means and goals.

Western European countries and Institutions have taken the lead in developing and implementing policies on FAIR. Therefore, dedicated efforts need to be focused on less advanced countries.

Examples: Finnish policies are highly coherent, which was achieved through coordination between the developments at a global level (OECD), European level (EOSC and the European Union), national level (Ministry of Education and Culture together with the Academy of Finland) and community-level (where both researchers and institutions are present) [FI_OS_Coord]. National Open Science working groups [FI_OS_WGs] comment on policies and ensure that national policy recommendations are taken into account in institutional policies. As a result, the national policy [FI_OS_Decl] has been developed by the community itself (through Open Science groups), but is at the same time in line with national and international requirements and funders' demands.

The research data policy of the Economic and Social Research Council (ESRC) in the UK [ESRC_Data_Policy] offers an example of a policy with consequences for non-compliance. It mentions that the ESRC has the right to apply sanctions, such as withholding the final payment of a grant, if data has not been archived within three months of the end of the grant.

The EOSC FAIR WG was not able to identify published examples of FAIR data policies being thoroughly and transparently monitored.

The above recommendations provide a basis for choosing the action areas that will be part of the EOSC programme over the next seven years, as well as identifying the requirements for those actions.

The EOSC action areas are described in more detail in the next two sections: Implementation challenges and Boundary conditions. For each, status, gaps and priorities are outlined.
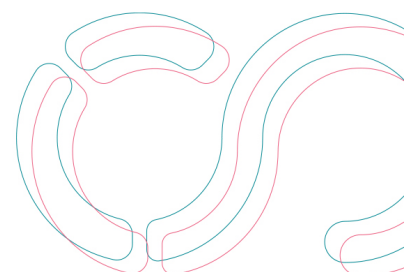
5 Implementation challenges

Based on the guiding principles and recommendations, the European Open Science Cloud governing bodies have identified fourteen action areas to help deploy the EOSC ecosystem. The seven areas relating to the primarily technical challenges and prerequisites to implementing the EOSC ecosystem are:

- Identifiers;
- Metadata and ontologies;
- FAIR metrics and certification;
- Authentication and authorisation infrastructure;
- User environments;
- Resource provider environments;
- EOSC Interoperability Framework.

This section describes each of those areas, giving an assessment of status, identifying gaps and proposing priorities. The remaining action areas are described in the next section, Boundary conditions.

5.1.   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.

5.1.1. Status

Systems that are based on an uncontrolled assignment of identifiers prove to be too unstable for trustworthy long-term identity management. In order to provide trusted PIDs that are usable, a combination of organisational and technical solutions needs to be supported. Services need to supply PIDs that are globally unique and have a stability of reference over time, and thus require organisational management and ongoing support. The EOSC PID policy [EOSC_PID_Policy] sets out the expectations on the use of PIDs and PID services by participants in EOSC.

Persistent identity is an established field, and mature technologies (e.g. Handle [Handle]), infrastructures (e.g. DOI [DOI]) and organisations (e.g. DataCite, ORCID, DONA, ePIC [DataCite; ORCID; DONA; ePIC]) already exist to support PIDs. The issuing of PIDs for publications (e.g. Crossref PIDs [Crossref]) and their use within citations has become standard practice. The use of PIDs for data citation (e.g. DataCite) and also unique references to people (e.g. ORCID) has been extensively developed over the last decade and has become widely accepted practice in the research community, although their uptake and use by the research community at large is not universal.

Beyond publications, data resources and researchers, PIDs are of value to identify all resources used or referred to in research data and accompanying metadata. These could include documents, data, people, organisations, projects, funding, software, services, instruments, samples, videos and other artefacts. Using PIDs for these resources gives more

reliable and semantically meaningful means to provide rich metadata to support research as well as properly attribute and track the use of valuable research objects. There are emerging technologies, standards and organisations to support many of these, although to date uptake has been limited due to a lack of commonly accepted approaches and clear business cases for their use.

Data generation and analysis applications are also increasingly being required to access and process data on a large scale and across distributed infrastructure. Software tools need to address data objects reliably and PIDs provide a means to do this. In these applications, PIDs need to be issued and accessed rapidly and at scale at the data generation stage and to be accessed across the research lifecycle. PIDs thus need to be assigned at an appropriate granularity for the application, to support the addressing of data objects within a larger aggregation. Support for versioning and tracking through the data lifecycle would also need to be supported to accurately record the provenance of data from raw through fully quality-assured data to actual results. The concept of a FAIR Digital Object has been developed, with an inherent use of PIDs, and standard specifications of PID Kernel Information and PID Type Registries published [RDA_PID_Kernel; RDA_PID_Registry]. Local handle systems have been provided to support these use cases. Nevertheless, this is an area that needs further refinement both in the applications where it is of most value and the practical technologies involved.

PIDs are thus an integral part of the research infrastructure, and play a key function in the data lifecycle, from the data generation and analysis stage to research output publication, curation and reuse. Tracking and connecting the use of PIDs in metadata and in citations, where they refer to one another, can form the basis of a rich, searchable resource for finding and contextualising resources. Tools that exploit this 'graph of research entities' include the Research Graph from OpenAIRE [OpenAIRE_RG] and the PID Graph from FREYA [FREYA_PG].

5.1.2. Gaps

PIDs are an established mechanism which has been used for nearly 20 years. However, there are still areas for further development.

- **Establishing mature and recognised PID infrastructures for emerging resource types**. There is a need to develop and establish trusted and widely used PID infrastructures for a wider range of resource types. In particular, instruments, software, organisations and services are types that would be of value in EOSC, although there is a wide range of further objects, such as physical infrastructure, physical samples, video recordings and theoretical concepts, some of which are domain specific. There are mature technologies for some of these resources, whilst others need development, and action on the adoption of all PID types is required. The PID types should then be used within core EOSC services, such as to register the scientific services in the EOSC-Exchange.

- **Support for machine-actionable PIDs**. Tools and standards supporting machine-actionable PIDs have been developed over recent years, including PID Kernel Information and PID Type Registries, but are not as yet mature or widespread. PID Kernel Information has been introduced as a small amount of standard metadata within the PID record to allow programmatic access and use [RDA_PID_Kernel]. PID Type Registries [RDA_PID_Registry] and accompanying Kernel profiles are not as yet standardised for different machine-readable data types and automated processing is largely missing or considered experimental. EOSC should consider the support of PID Type Registry services within EOSC, and develop services and use cases that exploit these services in automatic data analysis.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 87 –
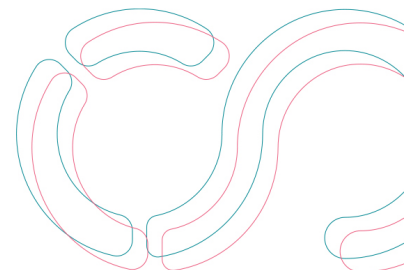
- **PID 'meta resolver'**. Each PID provider provides its own resolver, while a meta resolver could form a single service which can recognise different PID types and redirect to the appropriate resolver, regardless of issuer.

- **Standardising the PID graph**. Tools for connecting and searching across networks of PIDs are still prototypical, with several different approaches being explored (e.g. Research Graph, PID Graph, CERIF [CERIF]), and services developed to exploit this PID graph are still experimental and local. There is a need to standardise approaches across PID providers and for the uptake of tools built on this graph to become more widespread.

- **Integration of PIDs into FAIR data management**. The use of PIDs should be integrated into workflows that collect and analyse data to ensure that FAIR data is generated. PIDs need to be assigned early and potentially at scale (depending on the application). Collection of metadata associated with a PID needs to be automated close to where the data is generated, and integrated into data collection and processing workflows.

- **PIDs and sensitive data**. The FAIR principles would require PIDs to be used with sensitive data and this would reflect onto the PIDs themselves. This may lead to situations where access to parts of the metadata is restricted. EOSC would apply the principle of 'as open as possible, as closed as necessary'[25], and fine-grained access control for creating, updating and accessing PID records (Kernel information) may be needed.

- **Quality of service for PIDs**. The EOSC PID policy defines expectations on the quality of service of PID providers and services. The extent to which providers and services comply would need to be validated. The enforcement of policy is a governance rather than a technical issue for EOSC, but the governance may need to be supported by tools and processes to publish in a machine-readable form or validate the service.

- **New PID technologies**. New mechanisms and tools are appearing which support PIDs in novel ways. For example, some approaches do not require an authoritative certifying organisation. Intrinsic or smart PIDs are inferred (computed) from the form of the object (e.g. identifiers for software, or chemical objects). Others are decentralised, with no issuing authority but rather use distributed ledger technology to ensure their integrity. Further development and exploration should be encouraged within the EOSC programme.

5.1.3. Priorities

- Develop standardised identifiers for resource types that have not as yet become standard practice. For general research use, EOSC would 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 infrastructure against the EOSC PID policy.

---

[25] For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".

5.2. 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. The EOSC Interoperability Framework is taking a broader approach to identifying gaps and setting priorities relating to interoperability, and this section on metadata and ontologies fully aligns with this broader approach. Without improvements in interoperability, there will be no widespread adoption of metadata schemas and ontologies in European research activities, and Europe will fall short of fully realising Open Science.

The path towards better interoperability and adoption of existing metadata schemata 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, to drive the process of improved interoperability and increase adoption. Coordination with activities around metadata and ontologies outside of EOSC, for example in the Research Data Alliance (RDA), is of course essential. In addition, the interaction with the research information systems (CRIS) community and its metadata expertise will be of benefit in terms of enriching the research data with an elaborated set of additional research-related metadata (on researchers, projects, organisations, equipment, etc.) adding to the FAIR-ness of the data, while at the same time keeping the administrative burden for the research community as low as possible according to the 'only once' principle.

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.

5.2.1. Status

Scientific disciplines and communities have defined specific detailed metadata schemas and ontologies to describe community-owned data products. The adoption varies between research disciplines and is, for example, strong in the life sciences (e.g. DICOM [DICOM]), or astronomy (e.g. FITS [IAU_FITS]).

Metadata schemas describing resources that are not research outputs, e.g. organisations, instruments, samples, workflows or projects, have been developed in the research information community since the 1980s, resulting in standard metadata schemas that include projects, organisations and publications as well as persons (researchers), equipment and datasets, amongst others, while the schemas on samples, workflows or services are mostly an emerging activity.

Integration of discipline-specific metadata across communities and the aggregation of metadata derived from different metadata schemas and ontologies is still lagging.

Automatic metadata generation from instruments would be very beneficial, but is not yet common practice (though there is, for example, EXIF [EXIF], and again DICOM).
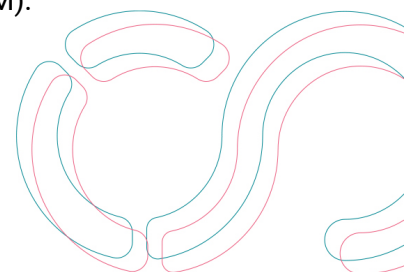
5.2.2. Gaps

Work on developing, improving and applying metadata schemas and ontologies – both for specific disciplines and for general use – is happening in many different places, but is often not well-coordinated, leading to a number of standards that are sometimes not well-aligned or that even conflict with each other.

Information about existing metadata schemas and ontologies is scattered across organisations and services, making it hard for users to find the relevant information. Such information is usually not described in a standardised way. The communities using a particular metadata schema are not always easy to identify.

Communities have defined crosswalks to map different metadata schemas and ontologies, but there is no standard way to describe or discover these existing crosswalks, nor to facilitate their maintenance when updates to the schema are applied.

Crosswalks between community-specific metadata and generic, common metadata, allowing the harmonisation of metadata for use cases such as discovery, have not been fully exploited, leading to silos of metadata that cannot be easily aligned.

The development of solutions in which community-specific metadata and generic, common metadata are integrated, resulting in an optimal application of metadata for use cases such as discovery, needs attention. In this respect one could build on existing work on metadata models, in which generic, discovery and subject or field-specific metadata are combined and integrated.

User-friendly tools to apply and maintain metadata for all types of research objects are not easy to find or are not available.

5.2.3. Priorities

- Develop governance structures for coordinating the work on metadata and ontologies within EOSC, both for specific disciplinary communities and for overall coordination.
- Provide or embrace/stimulate existing registries of metadata schemas and ontologies, defining clear protocols for federation/harvesting, crosswalks and tools for metadata management.
- Engage with the research information community in order to maximise the re-uptake of the information already residing in research information systems that communicate in a semantically interoperable manner based on standards.
- Develop EOSC guidelines for a minimum metadata description based on existing metadata schemas and tools 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 schemas across communities, sharing and community maintenance of crosswalks, measurement of metadata resources uptake across communities, validation of data sources against metadata schemas, etc.

5.3. FAIR metrics and certification

5.3.1. Status

The FAIR principles are a recent concept so metrics are still under definition. The principles were intentionally articulated broadly but this ambiguity leads to different interpretations and the risk that metrics do not fit different community practice. 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. The FAIR principles therefore need to be applied to all components of the ecosystem, since FAIR data maturity depends on the capabilities and trustworthiness of services such as repositories and persistent identifier systems.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 90 –
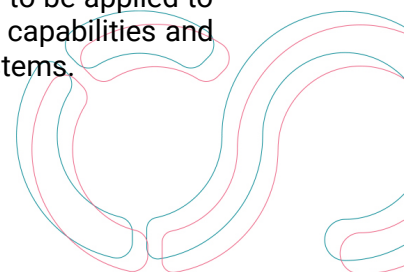
The definition of criteria potentially has very significant consequences if they are used to decide on participation or funding. Also, metrics are not meant to be a punitive method for direct comparison between datasets from different areas, because communities will arrive at optimal FAIRness in different ways. These risks are well understood by the community: the open consultation on the SRIA held during the summer of 2020 showed that metrics and certification are given a low priority, ranking second-to-last with 39% of votes in the feedback compared with 78% for the highest-ranked priority, metadata and ontologies. This has to be taken into account, by implementing them inclusively and progressively, taking into account also that FAIR is a journey, the diversity of community FAIR practices and the highly different stages of preparedness of the communities, to enable buy-in by a diversity of communities. It is essential to examine the criteria applicability and to gather feedback in a wide range of contexts.

The Metrics and Certification Task Force of the EOSC FAIR Working Group recommends that the definition of metrics should be a continuous process, regularly tested and iterated to minimise these risks. Inclusiveness should be a key attribute, to recognise the diversity of practice across communities and the different stages of FAIR maturity. Existing work, in particular by the international FAIR Data Maturity Model Working Group of the Research Data Alliance (RDA) [RDA_FAIR_DMMWG], should be built upon and tailored to the EOSC context. This forum also provides an appropriate international community to iterate and maintain the metrics, ensuring collective, community governance.

Status of metrics

The RDA FAIR Data Maturity Model Working Group has published a model with 41 criteria, allowing compliance of data with the FAIR principles to be assessed. A degree of priority – essential, important, useful – is attributed to each criterion. The Working Group has worked in a transparent way, and requested inputs and tests from the community throughout its eighteen-month time span. The model is being implemented, for instance, in FAIRsFAIR [FAIRsFAIR], which is progressively defining criteria to deal with use cases. The Metrics and Certification Task Force of the EOSC FAIR WG is proposing a set of possible EOSC metrics as a target, with a timeline towards progressive implementation, which requires extensive testing by a wide range of communities.

FAIRsFAIR produced a first assessment of FAIR semantics (semantics is discussed in Section 5.2 Metadata and ontologies) and high-level requirements for assessment frameworks, and an evaluation of how services influence data FAIRness.

Software is another important component of the FAIR ecosystem. A Working Group, FAIR 4 Research Software [FAIR4RS_WG], common to the RDA, Force11 and the Research Software Alliance, was created mid-2020, as a result of discussions held in many venues during recent years. Its aim is to define the FAIR principles for research software and provide guidelines on how to apply them. This WG should bring another key component to the FAIR ecosystem.

Status of certification

As stated in the 'Turning FAIR into reality' action plan [EC_EG_FAIR], there is a need for certification schemas to assess all components of the FAIR ecosystem. Significant work has been devoted to certification of data repositories, with an international landscape that includes in particular CoreTrustSeal [CoreTrustSeal], which provides a generic core framework for trustworthy repositories and has now certified an international set of trustworthy repositories in different disciplines, DIN 31644 (nestor Seal) [DIN_31644; nestor_Seal] and ISO 16363:2013 (also known as CCSDS 625.0-M-1 − Audit and certification of trustworthy digital repositories) [ISO_16363]. In parallel, ELIXIR is developing its own evaluation badges and processes [ELIXIR]. The availability of certification criteria is also an asset enabling repositories to self-evaluate and improve their practices and processes, even if they do not apply for formal certification.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 91 –
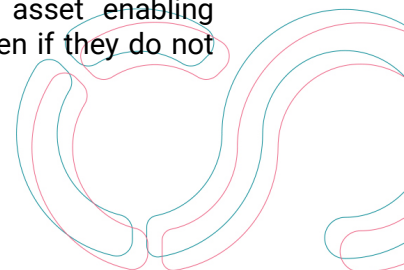
In the context of FAIR, work is ongoing, in particular in the FAIRsFAIR project, on FAIR alignment of repository certification schemas. This is complementary to the evaluation of the FAIRness of the data itself. More generally, the certification of FAIR-enabling services is also being studied – a service can enable, respect or reduce the FAIRness of its holdings.

5.3.2. Gaps

Existing work on FAIR metrics and certification should be extended under the next framework programme to ensure applicability across disciplines and support implementation. FAIR assessment should be inclusive and progressive, and its usage should take the specific context and needs into account. Several gaps and potential opportunities for extension are noted below:

- Metrics should be combined with a FAIR assessment framework that reflects the needs of different communities while offering comparable methods to assess FAIRness.
- The present checks are good for a proof of concept, but to make general rules for inclusion the scope of the tests was not broad enough; it has to be expanded considerably to explore potential problems and fine-tune the recommendations.
- Different communities attach different weights to the criteria, in particular but not only to interoperability, which has to be fully taken into account.
- The individual assessment models and metrics should be aligned with RDA core metrics and should not hinder a comparative evaluation.
- The model can already be used to measure progress on the path to FAIRness, but care should be taken before applying the model for pass-or-fail measurements.
- The need to develop automated evaluation tools for scalability is recognised but there are risks associated with the tool biases.
- Alignment of repository certification schemas with FAIR is under way but needs to be further developed and tested.
- Other critical elements include PID services, semantics and registries, for which assessment frameworks have yet to be defined.
- All the assessment frameworks have to be maintained over time, taking into account feedback from implementation and evolving requirements; the FAIR principles themselves may have to be maintained.

5.3.3. Priorities

Significant progress has been made on defining FAIR metrics for data and certification schemas for repositories. This should continue to be built on rather than reinventing the wheel, particularly given the global input and consensus fostered via the Research Data Alliance on these topics. Priorities for FAIR metrics lie in implementation and robustly testing across research communities. For certification of services, support is needed in aligning frameworks with FAIR, developing models for certifying core services such as PIDs, and enabling uptake.

Priorities for FAIR metrics

Priority 1: Support the assessment and improvement of the RDA FAIR Data Maturity Model.

Priority 1.1: Support disciplinary communities to clarify their requirements with respect to FAIR and identify cross-community use cases.

Priority 1.2: 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 and guidance, assess the degree of priorities, identify adverse consequences and apply corrections.

Priority 2: Assess and test the proposed EOSC FAIR data metrics in a neutral forum, which could be a Working Group set up by the RDA Global Open Research Commons Interest

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 92 –

Group, to seek global agreement with the international EOSC counterparts, in addition to any EOSC-specific Task Force or Working Group addressing FAIR metrics.

Priority 3: Support the definition and implementation of evaluation tools; their thorough assessment and evaluation, including inclusiveness; comparison of tools (manual, automated); identification of their biases and applicability in many different contexts, including thematic ones.

Priority 4: Support the definition of FAIR for software and of the assessment framework for key elements of the FAIR ecosystem, in the first instance PID services and semantics.

Priority 5: Define and implement governance of the principles, assessment frameworks and metrics, adapted to each specific case.

Priority 6: Provide guidance for and support to implementation: support data and service providers to progress in the FAIRness of their holdings.

Priorities for FAIR Certification

FAIRsFAIR is working on Priority 1, is also active in Priorities 2 and 3 with a set of repositories, and is working on a framework for FAIRness of services.

Priority 1: Support the current efforts to align certification standards and assessment schemas with FAIR.

Priority 2: Test the proposed schemas in a variety of communities to gather feedback and update the proposed framework accordingly.

Priority 3: Provide support, methodologically as well as financially, to data and service providers to progress towards certification.

Priority 4: Monitor the progress of certification, assess the maturity of the certification landscape, and take appropriate action if fields or regions are lagging behind.

Priority 5: Support the establishment of core criteria and methodology to certify other key elements of the FAIR ecosystem, in particular in the first instance PID services and vocabulary repositories / metadata registries, and test them extensively.

Priority 6: Support the establishment and maintenance of registries of certified components of the ecosystem; if several registries are available for a given component, they should be harvestable and included in registries of registries.

Priority 7: Establish a Working Group under the EOSC Stakeholder Forum to ensure the implementation and further development of recommendations in the 'Recommendations on certifying services required to enable FAIR within EOSC' report [WG_FAIR_CSRecs].

5.4. 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 for the security and trust of any collaboration. AAI for e-science is developed not in a vacuum but in the context of a global marketplace of AAI products and services which typically focuses on the consumer-business relationship.

The goal of the EOSC AAI is to build a foundation for e-science AAI which will ensure long-term availability of the aspects of digital identity that are unique to scientific collaborations

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 93 –

and which are often hard or even impossible to achieve using the tools and design patterns used to provide enterprise or consumer identity.

5.4.1. Status

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. The AAI for EOSC can build on a large body of existing work that has been carried out in the Federated Identity Management for Research [FIM4R] activity and the AARC and AARC2 projects [AARC] and its governance spin-off AEGIS [AEGIS], in which a large number of e-infrastructures and research infrastructures are represented. Most notable is the AARC Blueprint Architecture [AARC_BPA], which has been embraced by most large research collaborations worldwide and which describes the components of an interoperable AAI for research collaborations. The AARC BPA describes how community AAIs and infrastructure proxies can leverage eduGAIN [eduGAIN], the federation of national research and education (R&E) identity federations and other sources of identity for global science collaboration.

5.4.2. Gaps

Despite more than a decade of development in the field of global AAI for the research and education community, a period that has included establishing large-scale global systems such as eduGAIN and eduroam, both the user experience and the service provider experience remain confusing for large parts of the R&E AAI ecosystem.

The EOSC effort provides a unique opportunity to address these challenges. To guide this work, the SRIA authors have turned to the first principles of the EOSC AAI:

- User experience is the only touchstone;
- All trust flows from communities;
- There is no centre in a distributed system.

From these first principles the following problem statements have been derived:

- There is no consistent user experience for AAI across the e-science ecosystem;
- There is no consistent interface for service providers in the e-science ecosystem;
- The AAI ecosystem must grow to match the growth of EOSC beyond the research and education community.

Each of these problem statements is expanded below.

There is no consistent user experience for AAI across the e-science ecosystem

Currently the user experience for authentication and identification is fragmented. A user authenticating to several services cannot count on any aspect of that behaviour to be consistent, except possibly for the login screen of the home organisation identity provider (IdP) (if the user ever gets that far, that is).

In order to successfully identify to a service, a user must:

- Be able to identify the correct gesture to initiate a login flow – in other words, be able to find the login button on the page in the case of a web application;
- Be able to find her login provider (home organisation) among the offered alternatives;
- Have access to a login provider that offers a combination of authentication and identity assurance that matches the requirements of the service;
- Be able to understand what the login process entails in terms of authentication options, credentials, tokens, gestures, etc.;
- Have the appropriate association with the chosen identity provider (employee, student, etc.).

In summary:
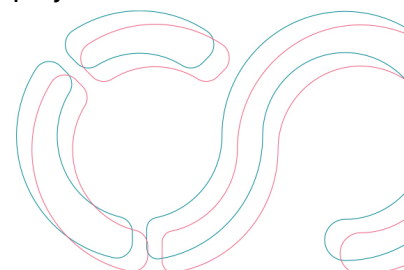
1. Services must be universally reachable, in the sense that users should be able to either successfully authenticate to all services or understand why they are not permitted access.
2. All participating identity providers must participate in a common framework for managing attributes across the ecosystem.

There is no consistent interface for service providers in the e-science ecosystem

Currently service providers and identity providers alike primarily interact and interface with national research and education federations, typically operated by National Research and Education Networks (NRENs). These are branded entities, designed (primarily) to support service delivery for campuses – mostly mission-critical services for administrative and student processes.

The needs of the e-science community are quite different from the 'bread-and-butter' IT services that make up the majority of the services (by use) of the current national identity federations.

In order to successfully integrate with the e-science AAI ecosystem, a service must currently:

- Implement an identity federation protocol;
- Register as a service provider with one or more identity federations;
- Convince a number of identity providers to provide attributes;
- Support additional service requirements such as security requirements, etc.

While this is simple enough to do for services where there is a clear relationship with the business goals of universities, these goals have proven to be almost insurmountable for many e-science service providers.

The reason for this is debated in the community but it is likely that a contributing factor is the fact that even the most well-organised and well-funded research project only has a very small number of contributors at any given university. The distributed nature of research ironically means that the more successful a project the more likely it is to look small – measured in terms of active users – from the point of view of any one university. The needs of services that only affect a small number of users are never prioritised by IT organisations who typically are responsible for the campus AAI.

Thus, in order to achieve the goals of EOSC it is necessary to establish a mechanism for connecting services to the AAI ecosystem that is better aligned with the needs of research and one that avoids the failure modes described above.

Specifically, EOSC should:

1. Scale the BPA (proxy) architecture and supporting infrastructure;
2. Establish clear rules of participation for services that foster cross-disciplinary interoperability between e-science services.

The AAI ecosystem must grow to match the growth of EOSC beyond the research and education community

The AARC BPA has proven itself from an architectural point of view in an environment with a limited number of research communities. With the expected growth of EOSC, models and policies must be developed to onboard communities and services at scale. A wide variety of sources of identity, from government and industry in addition to those from the R&E sector, must be supported, and both large and small and long and short collaborations must be made available. A particular challenge is the emergence of new technologies and paradigms that are recently getting some attention (mainly in the consumer identity space) and that may become valuable also in the field of R&E and enterprise identity.
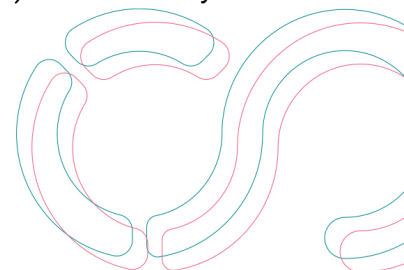
5.4.3. Priorities

Summarising the above, the following priorities have been identified:

- 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.

5.5. User environments

Users are those individuals who access and benefit from the resources exposed through EOSC. They may not be those agreeing or commissioning resources (the customers) but they are the ones interacting with them. In other words, EOSC users and providers include all actors in the scientific lifecycle, such as researchers, service providers, developers, funders, organisations, citizens, small and medium-sized enterprises (SMEs), etc.

The nature of EOSC is to establish a distributed, federated and clustered architecture. One of the main drivers is to make it possible for users to continually improve their own journey, including by giving EOSC feedback on possible bottlenecks, etc.

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. They may also include other environments yet to be created, both those as part of the central part of EOSC or those created by thematic or regional communities or even external interfaces created by start-ups/SMEs.

5.5.1. Status

Discovery of EOSC and resources

In order to benefit from EOSC, users must be able to discover research artefacts and services. Discovery implies the promotion, communication and presentation of the user environments.

Currently, promotion of EOSC is largely through projects working in the environment, so the current set of stakeholders is not fully inclusive. It is expanding, for instance through the thematic and regional INFRAEOSC projects, but this is still a subset of the European Research Area. Future projects and other initiatives, as well as clarified sustainability and governance structures for EOSC, will increase knowledge of EOSC.

Once users have identified relevant catalogues or portals, they must be able to use them to discover resources of interest to them. In the EOSC context, resources include computing, storage, data sources and scientific products such as literature, research data, software, experiments, documentation, etc. This implies effective cataloguing, tagging, search, discovery and suggestion mechanisms.

Present discovery of resources occurs primarily through the EOSC Portal [EOSC_Portal]. Discovery in the Portal is based on categorisation of the services. The EOSC Portal services are currently classified into the following categories: networking, compute, storage, sharing and discovery, data management, processing analysis, security and operations, training and

support. The Portal also includes tags and text search of submitted information. Rating of services is implemented but unused.

Other services exist in thematic and regional portals, and in future these are intended to be connected to the platform behind the EOSC Portal, but this has not yet occurred. Hence they are currently islands, with some services duplicated across them, rather than an interconnected system of systems.

The vision for the future is that resources can be discovered either through the EOSC Portal or through the other portals.

Composing resources in a user environment

Beyond the initial discovery and subsequent ordering of and AAI-managed access to resources, the aspiration of EOSC is that resources can not only be found and used, but also be 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 environment to generate new scientific outputs. Such composition can be facilitated by the science gateways, a well-established concept of user-friendly interfaces (suites of applications and tools) – researchers' work environments. 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.

Composability of resources is an aspiration of EOSC that in general has not yet been implemented. At present there is integration between researcher-facing services and core services, but this is not the same thing. There are some efforts to compose services coming from the EOSC-hub competence centres (e.g. deploying a workload management service from a community over a high-throughput computing service to compose a community-specific service) but they are limited. There are some examples of the user-community-specific science gateways, but not of common-use gateways.

Community of practice of EOSC researchers

To add value for the research domain, EOSC should not only bring together resource providers to work more closely together in support of composability, but also bring together users to enable and promote excellent research. Actions and functions that promote communication between users, especially those who are not from the same community or domain, establishing communities of practice, will support the success of EOSC.

Efforts exist within the projects constructing EOSC to build communities of practice, and the thematic and regional EOSC projects represent the construction of specific communities, but the larger community of practice of EOSC users is not yet a reality.

Some 'hooks' exist for these functions within the EOSC Portal, such as resource rating within the Marketplace, but the richer features are not yet there. More features may be seen in some regional and thematic portals, but these are also based on existing communities brought together online, rather than being created in the EOSC user environment.

5.5.2. Gaps

Discovery of EOSC and resources

All expected users for all user groups should be able to find the EOSC services and resources they need, but at present EOSC awareness is correlated with EOSC projects. In the next phase of building EOSC there must be ways to expose the wider community to EOSC. This may involve showing the benefits of EOSC to groups already using local or thematic user environments, as well as offering EOSC as a user environment for new groups who do not yet have their own effective user environment.
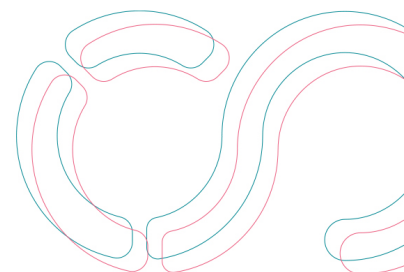
EOSC should also offer users functionalities to discover resources from the service providers of the distributed architecture. FAIR principles must be implemented where eligible. Possible tools for this are meta catalogues which aggregate information from the resource catalogues of the service providers; the EOSC Portal should function in this way, but does not today. These services are possible if autonomous service providers offer their catalogue information in the open interface for developers and expert users. Meta catalogues should offer the information to portals in the structured format and in the open interface. This should not only allow the EOSC Portal to offer an integrated meta catalogue by pulling resources from other catalogues, but also allow other catalogues to pull resource listings from the central meta catalogue. This interaction must be based on common agreement to use shared formats for resource description, and on application programming interfaces (APIs).

As part of this, categories must be rethought, as they have been inherited from prior efforts. They must be revised with community input, with a mapping to allow older entries to be recategorised. Tags should also be considered, to allow a relatively modest set of categories, for simplicity, and to offer indications of what resources are in terminology that makes sense to different user groups.

Composing resources in a user environment

To compose resources from autonomous and distributed service provider federations in a user environment requires a legal and organisational framework. This is needed for ensuring the position of the users and their work. This is not yet in place and is not fundamentally in the work plans of the current EOSC projects. Future EOSC projects must incentivise and encourage composability, both technically for specific pilot cases and at the organisational and managerial layer, to push providers into the choices that allow services to be composed. This implies both technical and policy-level convergence. Further expansion of the science gateway technologies in terms of the functionalities, EOSC services interoperability and towards new appliances and communities can be seen as one of the directions.

In the case of EOSC services, user requirements, usability and good user experience are critical aspects. These have to be a driver of the distributed EOSC service development. Development of EOSC and its services has to be continuous, agile and science-output driven. This is especially important for added-value services, applications and tools (the EOSC-Exchange) supporting the full cycle of scientific workflows. EOSC itself has a role as a usability evaluator.

Community of practice of EOSC researchers

More serious attempts must be made to support the creation of communities of practice, as they offer some of the clearest added value of EOSC, much as European funding drives the creation of communities of practice in research across the European Research Area. These must not be, for instance, simple 'forums' which users will not use, but must be naturally combined with user environments to drive uptake. Community of practice should involve both horizontal and vertical collaboration in EOSC.

Users must have clear feedback channels to EOSC and connecting points to services. For instance, when suggesting resources, workflows related to them could also be suggested, and other users who created or used those workflows highlighted, naturally funnelling users to spaces where they can communicate and share with peer researchers. Users in these communities have their role in setting requirements, targets and priorities.

5.5.3. Priorities

Several priorities have been identified to address the current gaps:

- Integration of existing catalogues and portals should be addressed to ensure users can find the services and resources they need.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 98 –
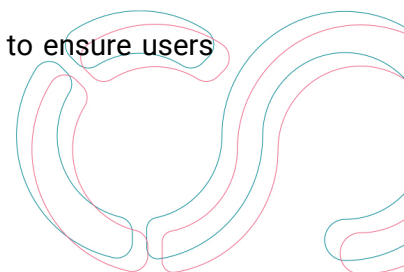
- In line with the definition of a minimum metadata framework, information about resources should be aggregated to enhance discovery.
- The distributed architecture model must address legal and organisational frameworks to enable researchers to compose resources.
- Strong engagement and consultation with the EOSC community of researchers is required to ensure interoperability and integration of portals, thematic and regional community services and resources.

5.6. Resource provider environments

EOSC is not a single monolithic organisation or resource provider but is rather a federation built out of many independent organisations and resource providers as in a system of systems approach. As such, it ensures the independence and autonomy of resource providers. Resource providers are widely distributed across Europe, have the mandate to serve one or more research disciplines and have to comply with different national and European legislations.

5.6.1. Status

The EOSC resource provider landscape is highly distributed and diverse. Resource providers are distributed across all European Member States and vary widely, with a number of resource providers dedicated to a specific scientific discipline or research community. Furthermore, there are generic resource providers serving national, regional and/or institutional research communities. Notwithstanding these challenges, the EOSC platform aims at gradually developing into a more mature offering. The EOSC Portal provides a mechanism for discovering the resource provider environments and requesting onboarding. Up to now, the groups onboarding and entering the resource provider environment have been driven by project membership, personal connections and some political considerations, but as the Portal and EOSC mature, a much wider uptake of the opportunities offered to resource providers is expected.

In the last decades, research infrastructures (RIs) and e-infrastructures have built service infrastructures to address their users' needs. The research infrastructures have been adopting, to a limited extent, common services provided by e-infrastructures. Because of the nature of how RI and e-infrastructure services have been developed to provide bespoke solutions, some level of composability between RI and e-infrastructure services exists, for example between community workflows, high-performance computing and/or cloud computing and data services.

To increase the value of funding and efforts previously invested in developing technologies, the approach taken was to reuse services and technologies as much as possible and adapt these to the requirements of a user or community. Services are composed with a community focus, therefore adapting services to another community is challenging. Due to community particularities, semantic differences, defined standards, use of APIs, use of different tools and services, such adaptations may sometimes be impossible. The development of bespoke solutions drove the proliferation of the standards and APIs in use by resource providers, limiting the interoperability and reusability of resources from an EOSC perspective.

One of the main challenges for EOSC will be to move beyond project-based partnerships and collaboration models to more sustainable long-term operations, such that resource providers can offer resources to any researcher in Europe and are assured that the resources consumed by researchers from outside their targeted user community are consumed in a financially sustainable way. This requires legal and organisational interoperability between resource provider organisations with sustainable funding mechanisms through which the costs can be recovered.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 99 –
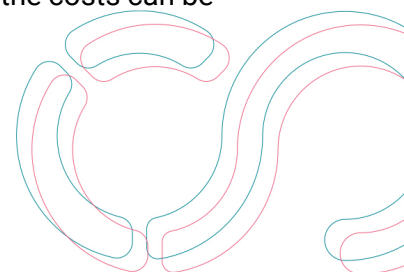
5.6.2. Gaps

EOSC should facilitate the work of resource providers in defining and adhering to a common interoperability framework. The framework would define policies (e.g. usage of PIDs for research entities, such as organisations, authors, services, data sources) as well as the information models and standards required to describe and monitor usage of resources, e.g. profiles for resources, relationships between resources, usage statistics, etc. By adhering to such a framework, resource providers will make their resources (i.e. research data, software, services) more findable and accessible, to some extent interoperable and reusable, but most importantly monitorable. Metadata about resources, their interlinking and their usage by users (services or researchers) will enable the definition of new indicators to measure both fulfilment of Open Science criteria (openness, FAIRness) and quality of science for all stakeholders, by considering the full production of science (not just the publications), the supporting services and facilities, and the investment made by the funders.

Resource providers should be incentivised to produce and operate resources that are Open Science by design, i.e. adhere to such a framework to support monitoring, sharing, and reuse of scientific outputs and reproducibility of science as a whole. Machines should support people, i.e. scientists, in the process of generating outcomes of science in such a way that FAIRness and openness (but as closed as necessary) are respected. The amount of manual work scientists will have to face to implement Open Science will otherwise risk being the most prominent barrier.

Communities, RIs and e-infrastructures have been developing interoperability frameworks and guidelines specific to their community and resource domain. While these frameworks are well known within their domain, they are in general unknown to the average user not belonging to the targeted user domain of the resource provider. To mitigate the problem of lack of awareness, EOSC can provide a platform through which communities, RIs and e-infrastructures can promote their interoperability frameworks and guidelines. Another part of the solution could be to ask resource providers to maintain a guide for using the resources and publish the standards that are used. At this moment, EOSC is missing a forum that stimulates the definition and evolution of adopted interoperability frameworks beyond the community and infrastructure domains.

Even if technical challenges are overcome, it is not obvious that, due to legal, organisational and/or financial constraints, researchers are allowed to access a service, data source and/or research product to which they have no direct access. These constraints are commonly set by non-technical boundaries defined in partnership agreements to which a resource provider belongs, in the mandate given to the resource provider or in national legislation and/or regulations.

For EOSC to achieve its vision (see Section 2 EOSC in the making) it is necessary to overcome not only the technical but also the non-technical challenges for resource providers to provide access to resources to any researcher within Europe in a sustainable way. The EOSC Rules of Participation Working Group has been given the task of specifying the initial conditions for resource providers to participate in EOSC. The rules are expected 'to set out in a transparent and inclusive manner the rights, obligations and accountability of the different stakeholders taking part in EOSC'.

The EOSC Rules of Participation and Interoperability Framework should contain legal and organisational aspects that allow resource provisioning in a sustainable way across resource providers and across national, community and partnership boundaries. They must also be sufficiently concrete for resource providers to understand how to comply and for EOSC to validate.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 100 –
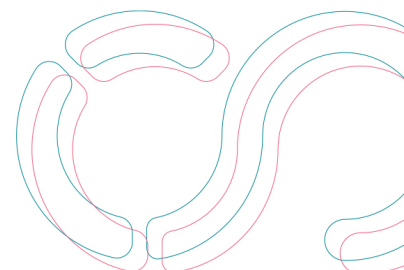
5.6.3. Priorities

Priorities for this action area are:

- Ensure more efficient onboarding of resources and integration with existing research community catalogues and repositories.
- Enable the composability of resources and across resource providers. This should be optimised and automated as far as possible, for example in resource delivery (e.g. on-demand and self-serve) and composability via the EOSC Interoperability Framework.
- Ensure the guidelines for resource providers are appropriate to and respectful of the existing interoperability frameworks available at a community level.
- Incentivise resource providers to produce and operate resources that are Open Science by design.

5.7. EOSC Interoperability Framework

5.7.1. Status

Achieving a good level of interoperability within EOSC is essential to federate data and services and provide added value for EOSC users, across disciplines, countries and sectors. In the context of the FAIR principles, interoperability is discussed in relation to the fact that 'research data usually need to be integrated with other data'. Standards are critical to achieve this, at both the disciplinary and cross-domain level, and implementation must build on existing research culture and practices, as well as existing technologies such as the Semantic Web, linked data and knowledge graphs. Efforts should also focus on addressing gaps where standards do not yet exist, to avoid the risk of leaving certain research communities behind.

Full interoperability, between data sources and services using different standards and semantic artefacts, is difficult to achieve at this point in time, but through EOSC the use of standards is being encouraged/required to enable crosswalks and as much interoperability as possible. In addition, the data need to interoperate with applications or workflows for analysis, storage and processing. The EOSC view on interoperability should consider not only data but also the many other research artefacts that may be used in the context of research activity, such as software code, scientific workflows, laboratory protocols, open hardware designs, as well as the services that allow the handling of such data. The current EOSC Interoperability Framework focuses mostly on the digital object level and recommendations are made for expanding this in the next phase of work to address services and other components too. For example, rules for service operation should require a level of reliability and availability to guarantee stable service levels.

In terms of EOSC, the 'I' of FAIR is the critical aspect, as interoperability is the glue that allows EOSC to function. In order to enable data to be discovered and accessible, a minimum set of metadata, common standards and, preferably, machine-readable semantic artefacts that can interoperate needs to be agreed. Interoperability across countries, data repositories and disciplines is fundamental to the EOSC vision and a prerequisite for the federated approach. Some work has been started in this regard. The FAIRsFAIR project has been reviewing generic metadata standards to recommend approaches for common discovery in EOSC, a co-creation project has been funded to review the DDI-CDI standard and RDA Working Groups are considering standards such as schema.org. A comprehensive review of all possible generic standards to adopt within EOSC must be conducted and in-depth consultation with the full range of research communities must take place to determine which will be most appropriate to apply for broad uptake.

5.7.2. Gaps

This section is organised according to the different layers of interoperability that are identified by the European Interoperability Framework: technical, semantic, organisational and legal. In

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 101 –

addition, there are three overarching activity areas under which these various gaps and associated priorities fall:

- **Support for standards development and adoption**
  EOSC cannot enable FAIR and support interoperability without standards to describe and understand digital objects. Many of the gaps identified address a lack of standards or low levels of adoption, both of which need to be addressed incrementally to enable the full benefits of FAIR to be realised. Once research community standards are in place, work can be performed to map between these, enabling data and services to be used in wider contexts. Turning FAIR principles into practice requires an enormous amount of human skills and support, as well as the standards and technological resources. This gap is even larger if the data coming from the long tail of science are taken into consideration, so work to professionalise data stewardship roles and ensure appropriate levels of support and services are in place is key.

- **Engagement with research communities**
  FAIR should be implemented according to the subsidiarity principle, preferencing standards of research disciplines over more generic, less rich metadata. Engagement with professional scientific unions or scholarly societies, research infrastructures, data stewards and software engineers that work closely with research communities and represent their needs at an institutional, European and global level is key to ensure standards have wide applicability and adoption. Fora such as W3C, the RDA and other bodies defining standards at the global level also play an important role here.

- **Robust governance and implementation**
  A clear governance framework is required for implementation to specify how the different levels of interoperability will be handled across organisations and user communities. Policies from funders and institutions should require and/or incentivise the curation and use of agreed standards. Moreover, recommendations can be made to ensure common services, such as PID resolution, function consistently irrespective of the type of identifier used.

At the technical level, the main gaps with regard to achieving better interoperability in EOSC include the following:

- When trying to work with infrastructures or services across communities, **authentication and authorisation often needs to be performed separately for each community/service**.
- Research data may be made available in **multiple general-purpose formats** (CSV, Excel, database dumps, JSON, XML, shapefiles, etc.) **or community-based models** (e.g. Darwin Core, VOTable and VOResource, FITS, NetCDF), which are usually hard to align when reusing datasets across communities.
- Coarse-grained or fine-grained research data from other communities may be difficult to find, given the **lack of knowledge about how to query their repositories**.
- Multiple service providers for **different types of PIDs exist**. As a result, different sets of policies are enforced to varying degrees, and sometimes the identifiers are not even resolvable.

At the semantic level, many of the interoperability gaps have already been identified in Section 5.2 of this document. The primary issues are as follows:

- Need for **principled approaches and tools for ontology and metadata schema** creation, maintenance, governance and use. Different communities are using different tools and representation models for their semantic artefacts. Some communities have no agreed standards and no strategies for bridging that gap.
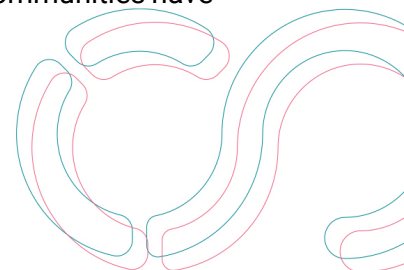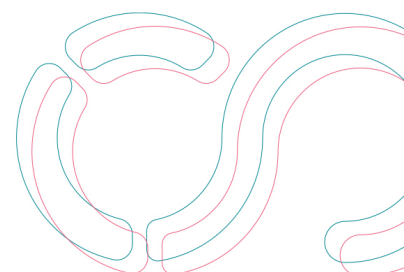
- Need for **harmonisation across disciplines**. It should be possible for a user from one community to add metadata to existing items (data and semantic artefacts) according to their own research discipline practices (e.g. for a social scientist to add DDI-based metadata to a dataset coming from an environmental scientist). Allow a researcher to transform metadata (or data) from one discipline's format/annotations to another's.
- Need to **harmonise the same type of data** (e.g. observational data in environmental sciences, as is being done in the I-ADOPT RDA WG).
- Need for **federated access over existing research data repositories** (both inside a discipline and across disciplines). Ability to support discovery of data on the basis of a high-level description, and possibly also on more details such as concepts related to observations and variables.
- Lack of tools for deduplication of legacy metadata records and their quality validation.

Use Case: Astronomy – interfacing an existing, global, widely used disciplinary data-sharing framework with EOSC

Astronomy has been a pioneer of open data sharing and remains at the forefront. The astronomical Virtual Observatory (VO) is the disciplinary interoperability framework which enables astronomers to discover, find, access, interoperate and reuse the data they need for their research. It is defined at the international level by the International Virtual Observatory Alliance (IVOA) and is widely used by data providers worldwide. It provides seamless access to resources and is thus almost invisible to the astronomers but underlies some of the most-used tools. The IVOA Registry of Resources lists more than 100 'authorities' that provide at least one VO-enabled resource.

Interfacing the IVOA with EOSC would allow the astronomy community to use resources provided by EOSC as seamlessly as possible, for instance computing resources or EOSC-Core services such as AAI. For example, the IVOA Registry of Resources should be integrated with an EOSC catalogue of services, and the semantic vocabularies provided by the IVOA recommendations should be interfaced or integrated in the relevant EOSC services. Another aspect is to assess how VO-enabled resources could be 'onboarded' as EOSC services, in a way compatible with the VO service description and deployment architecture.

The integration of VO data and services into EOSC is being assessed in the ESCAPE cluster, which focuses on connecting ESFRI projects to EOSC. The IVOA Registry of Resources, which is built on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and includes the Dublin Core plus disciplinary extensions in its metadata schema, has been included in EUDAT B2FIND [EUDAT_B2FIND]. B2FIND exposes the resources and allows filtering by keywords, which works well for the high-level resource description. Figure 5.1 below gives a composite of screen copies of searches for IVOA resources in B2FIND.
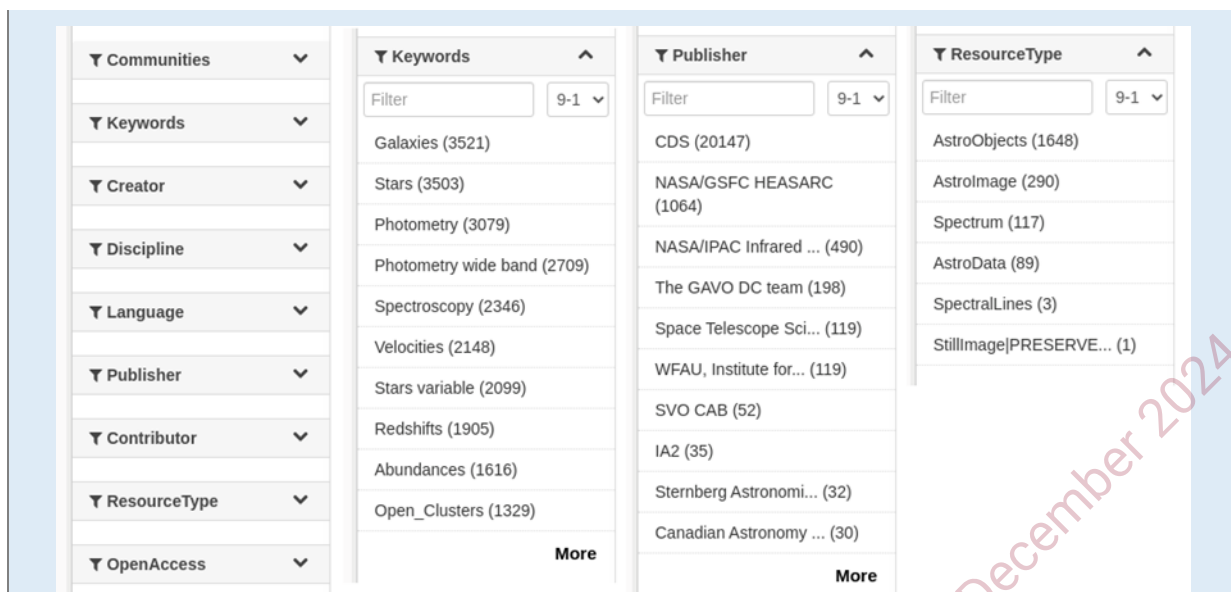
Figure 5.1: Composite image of searches for IVOA resources in B2FIND

This integration exposes the global astronomical data resources in the B2FIND interdisciplinary environment. It can be seen as a first step on the way to federating into EOSC but lots remains to be done as the EOSC structures mature. In order to be accepted and bought into by the astronomical community, any interface should preserve the existing, widely used capabilities provided by the IVOA framework, respect the disciplinary culture, which includes open access to data, and should not require a high overhead from the data providers. Ongoing consultation with research communities is therefore critical as EOSC Rules of Participation, minimum metadata frameworks and interoperability requirements are specified.

At the organisational level, the following gaps have been identified:

- Need for a **clear governance framework** that includes clear instructions on how the other levels of interoperability will be handled across organisations and user communities (data formats, AAI services, metadata schemas, ontologies, etc.).
- Need for **documents explaining terms and conditions and acceptable use policies for services providing interoperability**. For instance, providing clear descriptions of the service-level agreements (SLAs) of those providing catalogues and registries of semantic artefacts, or providing systems to overcome semantic differences between different data sources, or alignments between models.
- Need for **interoperability certification mechanisms for service providers**, so that service users can set their own expectations about the support for interoperability of those services.

At the legal level, the following gaps have been identified:

- **Lack of clear statements of rights** or information on the legal conditions under which data can be accessed. In effect, much content is shared without a usage licence, let alone a standardised, machine-readable licence.
- **Wide adoption of standard open licences** for data, code and other outputs to ensure data can be combined without conflicts in licence terms.
- National copyright varies across countries so there is a need for clear licences or a good understanding of how and which type of **data can be exchanged, taking into account different jurisdictions**.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 104 –

eosc
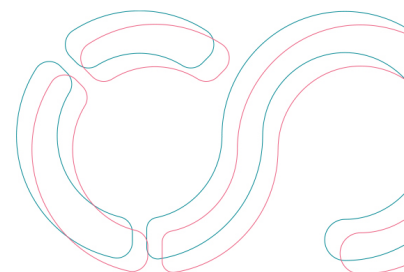
- **Consistent, machine-readable consent agreements** to ensure permissions to access or use data are clear and regulations such as the General Data Protection Regulation are met.
- Need for a **repository of machine-readable licences** that can be associated to different types of research outputs.
- Need for machine-readable schemas for the representation of the main characteristics of **service-level agreements**.

5.7.3. Priorities

As a result of the previous analysis of gaps and needs, the following recommendations can be made to include as priorities for further development of interoperability in EOSC at the technical level:

- **Use open specifications**, where available, to ensure technical interoperability when establishing EOSC services.
- Define a **common security and privacy framework** and establish processes for EOSC services, to ensure secure and trustworthy data exchange between all involved parties.
- Define an AAI process for EOSC that is common across communities, easy to implement by resource providers and easy to understand by users.
- Ensure service-level agreements for all EOSC resource providers are easy to understand by users from different communities.
- Enable discovery of data sources available in different formats, either generic or community-based, to facilitate overcoming their heterogeneity, and provide **easy access and tools to integrate data across communities**, enabling the usage of these data.
- Provide tools for quality validation of metadata records and content of digital objects.
- **Make available 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.
- Implement the EOSC PID policy, accommodating any appropriate PID usage, recognising that established practices are at different levels of maturity for different resources and that new PID types may emerge.

At the semantic level, the following priorities have been identified:

- EOSC should provide support for the **maintenance of repositories of semantic artefacts, and governance frameworks** for such repositories, taking into account common practices and stages of semantic resource development and usage of different communities.
- EOSC should define clear protocols and building blocks for the **federation/harvesting of these repositories of semantic artefacts**.
- Research communities should be well supported (independently of their current state of semantic artefact adoption) so as to generate **clear and precise definitions for the terms** they use, as well as **for their metadata and data schemas** (and to incorporate those that they are already using) and **their documentation**. EOSC should provide support to make these definitions publicly available and referenceable by persistent identifiers for machine readability.
- Urgent, additional resources (financial, but also skills and training) should be dedicated 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.
- EOSC should propose a minimum vocabulary to allow **discovery over federated research artefacts (data, software, publications, etc.)** across scientific communities,

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 105 –

and based on existing metadata models (e.g. DCAT, DDI 4 Core, DataCite core schema, OpenAIRE Guidelines). There should be some alignment among them, and this vocabulary should be extensible, to allow for disciplinary metadata that is typical for some research communities.

At the organisational level, the following priorities have been identified:

- The **current set of Rules of Participation recommendations should be completed with aspects related to interoperability**. For instance, for data providers this may include asking explicitly that data is published according to specific data formats and/or vocabularies for a specific community.
- The same is applicable to **services**, which may be recommended to ingest or output data according to such standardised data formats and/or vocabularies, and to their corresponding metadata, with some level of quality.

Finally, at the legal level of interoperability, the following priorities have been identified:

- A **list of EOSC-recommended licences** and their compatibility with Member States' recommended licences should be provided to data producers, right-holders and users, so as to avoid an inadvertent breach of copyright and with a view to harmonising and reducing the overall number of recommended licences.
- EOSC should seek to develop and implement **minimum standardised, human- and machine-readable expressions of right statements and use conditions**, to be included in metadata and be used by all repositories regardless of discipline.
- Need for metadata schemas for service-level agreements.
- EOSC should consider developing a centralised source of knowledge and support on copyright and licences to users and data generators and to address common Q&A.

As noted in Section 5.3 FAIR metrics and certification, these recommendations necessitate investment in the development of data standards, crosswalks and registries to support a FAIR ecosystem. Two key areas of activity for the next framework programme are the support of community standards and the proposal of a minimum vocabulary to allow discovery over federated research artefacts (data, software, publications, etc.) across scientific communities.

This section has described the seven action areas relating to the primarily technical challenges and prerequisites to implementing the EOSC ecosystem. The following section addresses the challenges and prerequisites relating to boundary conditions.

6 Boundary conditions

Seven action areas have been identified by the EOSC governing bodies to help deploy the EOSC ecosystem which relate to the social, financial, legal, educational and cultural challenges and prerequisites to its implementation. These are classified as boundary conditions and are:

- Rules of Participation;
- Landscape monitoring;
- Business models;
- Skills and training;
- Rewards and recognition;
- Communication;
- Widening to public and private sectors and going global.

This section describes each of those areas, for each one providing an assessment of status, identifying gaps, proposing priorities and, where appropriate, describing some further considerations which should be taken into account.

6.1. Rules of Participation

6.1.1. Status

In the current European research landscape, Open Science practices are not yet the norm amongst many researchers. Data and other digital research objects are not consistently findable, accessible, interoperable and reusable (FAIR), and the current landscape of regional, national, European (and international) research data infrastructures (RDIs) is distributed, diverse and fragmented. This presents barriers to the open sharing of scientific results.

Research collaboration in Europe and globally can be further increased to realise more and better science if changes are implemented to adopt Open Science practices, make digital research objects FAIR and federate RDIs. EOSC aims to achieve this by stimulating widespread changes in the research environment. The Rules of Participation (RoP) [EOSC_RoP] provide transparent and consistent terms for participation in EOSC, helping to build the trust and confidence required to support this process of change. These RoP are set at a level to encourage wide participation, including from less advanced research communities.

The EOSC legal entity, the EOSC Association, will be responsible for the RoP, including their monitoring, enforcement and periodic review and updating, to ensure their impact is understood and that they respond to the requirements of the maturing EOSC.

6.1.2. Gaps

The current EOSC has evolved through research and development activities undertaken in a number of projects that have progressed largely independently. Each project has proposed and followed its own work plan. However, these work plans have not been as well coordinated as they might have been. For example, schedules of delivery of services could be more coordinated, onboarding requirements could be better aligned, and greater coordination could help avoid gaps and synchronise on overlaps.

Whilst project-level governance can monitor compliance of projects against their own objectives and planned activities, unless those activities are aligned a priori, project-level monitoring has little leverage on coordination across projects. It has been observed that it is difficult to build a coherent infrastructure through a collection of independent projects.

To build a coherent infrastructure that removes silos and provides integration requires tighter coordination across projects. It is not possible to build a highly interconnected road network
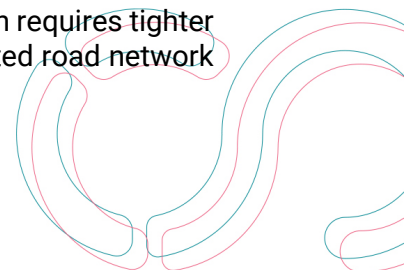
with contributions from several partners without joint planning of where roads end in one area and begin in another. Thus, whilst local requirements – whether regional or disciplinary – are best served through local planning, there needs to be wider agreement about where and how the local arrangements will integrate. In addition, there needs to be international agreement about the interfaces across national boundaries.

For EOSC, this is partly about architectural standards that enable integration at a technical level, about data standards that enable sharing of data, and also about stakeholder engagement to bring about standardisation of policies, processes and procedures. The RoP will define the policies, processes and procedures required to provide assurance of sustainability, transparency, quality and trust in the practices and services offered through voluntary participation in EOSC.

Many of the qualities the RoP may ideally require of EOSC participants are not yet widespread and are not universally available. For example, many digital research outputs are not fully FAIR; many repositories are not yet certified; metadata is not fully standardised and a common metadata framework to support discovery in EOSC has not been defined; many services are not yet interoperable; authentication and authorisation infrastructure (AAI) is not yet globally recognised/interoperable; persistent identifiers (PIDs) are not yet universally assigned and unique.

Diversity is a major challenge for defining RoP. The evolution of the RoP will be an iterative process, achieved through dialogue with the community and enforced with the consent of the community. It is also important to develop rules that encourage EOSC users and suppliers in the desired directions for EOSC to achieve its objectives, whilst not imposing requirements that are so onerous as to discourage use of and supply to EOSC. The Rules will also need to reflect changes in the wider environment, such as the development of the GAIA-X initiative [GAIA-X].

In the beginning, therefore, many of the Rules will need to provide encouragement rather than impose strict requirements, but can develop over time to include more stringent conditions. This approach is evident in the proposed FAIR metrics and Interoperability Framework, which define different target levels for each time period, incrementally increasing expected levels of FAIRness and standardisation.

6.1.3. Priorities

- The RoP define standards for policy, processes and procedures that provide assurance of quality and trust in the services offered through EOSC.
- The RoP apply to all digital resources made accessible via EOSC, including data and services. They define a minimum set of rights, obligations and accountability 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.
- The RoP assume that the governance structure for EOSC will include a governance framework involving the appropriate stakeholders, which includes a legal entity that will assume ownership of the RoP and provide a decision and revision process for them.
- The RoP may evolve in the future to incorporate elements arising from the FAIR, Architecture and Sustainability Working Groups (WGs), which are developing recommendations in their respective domains.
- The RoP provide a conceptual framework for policies and documents relating to issues such as Terms and Conditions and Acceptable Use Policies. These will need to be further elaborated and reviewed with respect to legal regulations before the RoP are finalised.
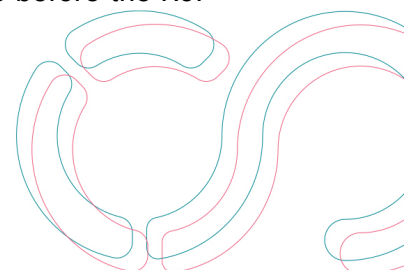
- RoP are about governance, oversight and authority. Without RoP, EOSC becomes no more than a search engine over an unmanaged collection of resources.
- It is essential that there is a framework where RoP can be defined, maintained and enforced.
- If EOSC is to be delivered through a programme of projects, far greater control over these projects is required.

6.1.4. Considerations

Further considerations that should be taken into account with regard to Rules of Participation relate to:

- Registration and discoverability;
- Transparent subsidiarity;
- Federated services;
- Federating services;
- Global agreement;
- EOSC compliance for external services.

Each of these is discussed below.

6.1.4.1. *Registration and discoverability*

EOSC will be primarily a federation of existing data and services where data remain in their current repositories and EOSC provides a means to make those data more broadly discoverable and interoperable. To enable this federation, EOSC must recognise resources, or collections of resources, through registration of those resources in an EOSC catalogue. Participation in EOSC is therefore defined by registration of resources as EOSC resources or in an EOSC-recognised collection of resources. Although somewhat tautological, this definition acknowledges the fact that participation works on a voluntary basis; if and when a provider chooses to register a resource with EOSC, it becomes discoverable and accessible through EOSC.

A digital resource is therefore considered to be an EOSC resource if, and only if, it is registered in an EOSC-recognised catalogue of resources. Registration of resources also indicates compliance with the EOSC RoP and use of EOSC branding is available only to registered resources.

6.1.4.2. *Transparent subsidiarity*

While participating as a data provider in EOSC implies commitment to the principles of openness described above, custodianship of the data remains with the data provider. Thus, individual data providers determine the precise conditions under which the data they expose through EOSC may be accessed and used, provided that these do not contradict the underlying principle of openness. Such resource-specific terms of use may, for example, require users to inform the data providers of the purpose for which the data will be used.

In line with the principle of transparency, data providers will clearly define and publish any such terms of use for the data they provide. These will include any licensing information, whether access requires authentication and/or authorisation, and any conditions regarding how data can be processed, changed and redistributed by users.

Users of EOSC resources will also need assurance of the quality of the resources they consume. This applies not only to data and services, but to all resources, including software and training resources, etc. As participation in EOSC is voluntary, it is difficult to impose uniform quality standards across the infrastructure. This may not even be desirable, as quality and value do not necessarily correlate: a certain resource may be very valuable even though it is of low quality because it is unique, whereas for other resources that are more widely

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 109 –
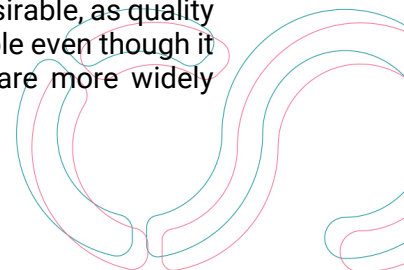
available, quality may be the value-defining characteristic. Here, transparency and community governance are the best way forward. If quality metrics are monitored and made openly available, resource users can ascertain and assess whether a particular resource is valuable for their use. To enable this transparency, the RoP will require resource providers' cooperation with regard to the monitoring and reporting of their resource.

6.1.4.3. Federated services

As with data, in order to be available to EOSC users, services that are federated in EOSC need to be registered in a service catalogue that is itself registered with EOSC. This is not to say that users will necessarily access these services through a generic EOSC gateway. Rather, researchers may continue to access resources through their existing field-specific portal, with these portals being enhanced through access to a wider range of resources, mediated and adapted by the providers of the domain-specific resource. As with many forms of infrastructure, providers of existing portals may be able to hide the technical details of how services are delivered and seamlessly present new functionality in a way that is tailored to communities in their specific fields.

For such an invisible infrastructure to be achievable and maintainable, service descriptions and protocols will need to be provided in both human- and machine-readable forms. The metadata supporting this may include: parameters relating to terms of use, including any accessibility constraints and/or quotas; the means of accounting and monitoring; measures concerning verifiability and quality of service, including any service levels; definitions for technical interoperability such as application programming interface (API) descriptions; and declarations relating to liability.

For these metadata to be machine processable without the need for software to be hard-coded to particular schemes requires the definition and agreement of the metadata schema and vocabularies to be used. While it is unrealistic, in the short term, to expect all communities to agree on a single, universal metadata scheme, it is feasible to envisage adoption of a registration service for schemata with the individual schema being agreed within specific communities through global consensus-building activities such as those supported by the Research Data Alliance (RDA) [RDA].

6.1.4.4. Federating services

The EOSC federating services, guiding those who operate EOSC, are those services that are required to support the functioning of EOSC itself, enabling it to function as a federation. Such federating services include those concerning: authentication, authorisation and accounting; registration of users, organisations and projects; monitoring and accounting of usage; and service and data catalogues. Central to this suite of services, and also underpinning findability and accessibility, are the persistent identifier services that can provide some necessary stability and provenance in an otherwise highly dynamic and flexible environment.

These federating services will necessarily be subject to more stringent requirements in order to support the levels of availability and reliability that users will expect from a functioning research infrastructure. Unlike the federated services, each of which will have their own independent community-focused funding mechanisms and metrics for success, the federating services are generic in nature and will therefore be more directly linked to the EOSC governance framework through qualitative and quantitative service level agreements.

6.1.4.5. Global agreement

It is crucial that EOSC interoperates with other open research support environments within and outside Europe. Research is global; therefore, research infrastructures need to support global communication and collaborations. Global reciprocity agreements and discussions,

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 110 –
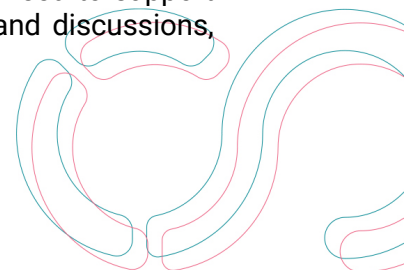
such as those provided by the RDA Interest Group on Global Open Research Commons, are an essential component for establishing common principles.

6.1.4.6. EOSC compliance for external services

It should be recognised, however, that EOSC will never provide, nor should it attempt to provide, all the services, resources and tools that will be used by researchers. Many tools, such as internet search engines, social media communication channels and office systems tools, are currently provided, and will continue to be provided, by suppliers external to EOSC.

An important consideration for EOSC will be how to accommodate use of such external tools into research workflows, and whether a notion of EOSC compliance needs to be developed for such external tools and services.

6.2. Landscape monitoring

6.2.1. Status

The Landscape Working Group established by the EOSC Executive Board [EOSC_WG_Land] set out to survey and document the landscape of infrastructures, initiatives and policies across Europe relating to the development of the European Open Science Cloud, as at the beginning of 2020. The resulting report [EOSC_Landscape] describes 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. Although the report only provides a snapshot of current policies on Open Science, it shows that the majority of the countries surveyed have either already adopted a respective policy or are getting ready to adopt relevant practices in this regard. Encouragingly, some of the countries that have already adopted policies have also defined the target year by when the policy should be fully implemented, which means that they have also defined detailed action plans.

In most countries, Open Science requirements are reflected in existing research evaluation policies. However, the majority of these policies (61%) only cover requirements for open access to scholarly publications; requirements for FAIR data are only in place in the policies of 34% of the surveyed countries. While it is encouraging that 44% of countries indicated they were planning or currently developing policies covering FAIR data as well, it is obvious that only a small number of countries seem ready to mandate that research data should automatically be made open. Equally problematic is the finding that relatively few countries (21%) mention EOSC in their policies. While this will also change in the future, as 38% indicated they plan to do so in future, only three countries (Bulgaria, Denmark, Romania) currently include reference to EOSC in their criteria for funding. More than half of the surveyed countries have nominated contact points for Open Science (53%) and for EOSC (42%).

A follow-up to this activity, utilising a more dynamic approach, was identified as an urgent need by the stakeholders during the validation workshop in early 2020. A monitoring exercise, able to capture the overall development of Open Science implementation in Europe, is an important asset for the first phases of EOSC implementation.

6.2.2. Gaps

There are many monitoring mechanisms in Europe, offered by, for example, OpenAIRE, DCC, SPARC Europe, GÉANT, etc. Nevertheless, none of them provides a complex view of the landscape with a particular focus on EOSC. WG Landscape provided a snapshot of the EOSC-related environment in MS and AC at the national and institutional level. The report gives an overall description of the landscape of the European infrastructures.
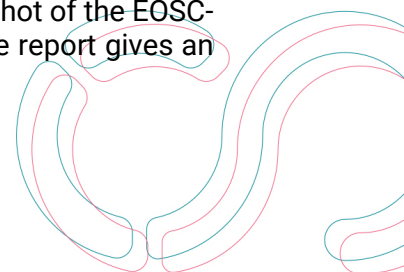
By the end of 2020, all MS should be prepared for joining EOSC. This accelerates the current development within this area. In light of these facts, the Landscape Report is of value but the information it provides will quickly be out of date. Regular updating will support the purpose of the report and users will be provided with relevant information on the preparedness and readiness of MS/AC and stakeholders for joining EOSC.

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.

6.2.2.1. Monitoring areas

The Landscape Report structure provides a framework for future monitoring exercises and for the structure of the areas to be monitored. The areas that require monitoring include the following.

A. The infrastructure landscape
   a. A description of the backbone infrastructure(s) at national level that is/are already contributing to EOSC services (e.g. relevant data infrastructures, e-infrastructure, other services and data management cycle).
   b. A description of the supporting infrastructure(s) contributing to Open Science targets at the national or regional level (e.g. universities, public and private research-performing organisations, thematic infrastructures, etc.).
   c. An overall description of the remaining research environment relevant to EOSC, including the private sector.

B. The organisational landscape
   a. A description of the institutional structure(s) at national level accountable for defining and implementing EOSC-related policies and strategies, including their hierarchical structure.
   b. A description of the EOSC-related policies and strategies.

C. The strategic landscape
   a. A description of the institutional structure(s) at national level accountable for defining and implementing EOSC-related policies and strategies, including their hierarchical structure.
   b. A description of the EOSC-related policies and strategies with direct and indirect impact on EOSC.
   c. A description of various EOSC-supportive measures taken at the national, regional or institutional level (programmes, projects and their harmonisation, financial and other incentives, etc.).

D. The strategic outlook
   a. An assessment of the level of preparedness at national, regional and institutional level to join, support or interact with EOSC (e.g. not only research data but also data-related algorithms, tools, workflows, protocols, services and other kinds of digital research objects, as well as remote access to research infrastructures).
   b. A part of the monitoring exercise should focus on updating the list of infrastructures, including all stakeholders and services, and various scientific disciplines, that have already reached a certain level of EOSC implementation.
   c. A description of any relevant trend in the evolution of the research environment (e.g. scientific domain in the context of EOSC development).
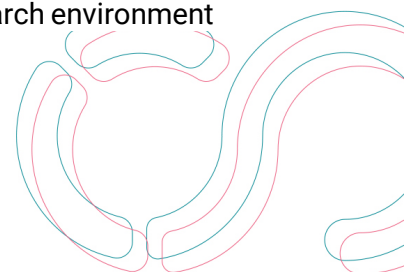
Before conducting the monitoring exercise, it is critical to define the purpose and aim of the monitoring process (i.e. the monitoring methodology) and to identify the right tools for gathering data. Considerations regarding gathering and maintaining the information to ensure the sustainability of the datasets, in terms of both internal consistency and persistence, are another indispensable prerequisite for a good monitoring process. Sufficient and sustainable funding concepts shall be developed and aligned to the identified monitoring methodology and data maintenance. The monitoring methodology must be developed to take into account the needs of the envisaged European EOSC implementation architecture, but at the same time it must have the flexibility to accommodate national specificities, and, in addition, all the stakeholders must be consulted. This applies in particular to the key performance indicators, which, as stated above, must be designed, selected and approved with all the major stakeholders. KPIs cannot replace the expertise and knowledge of the evaluation/monitoring panel, and the monitoring cannot be reduced to administrative procedures only. KPIs should comply with well-proven criteria for defining indicators and measures.

6.2.3. Priorities

It is important to elaborate a thorough monitoring methodology to define not only the criteria and indicators, but also process and responsibilities. Given the self-governance model chosen for EOSC implementation, this must be driven bottom-up to meet the varying needs of the different stakeholders' communities, as well as to encourage harmonisation of the national and regional priorities with pan-European development, with only light supervision from the EC. Priority shall be given to the description of the full set of actors and actions, ranging from compliance with FAIR principles in the internal strategies and policies of the individual institutions (universities, research-performing organisations, research infrastructures, e-infrastructures, etc.), up to the monitoring of the overall environment of the national landscape (national policies and strategies, research-funding organisation actions and other measures supporting Open Science, etc.). The monitoring shall comprise an assessment of both the societal and the technical aspects of EOSC implementation readiness.

The following priority areas have been identified:

- 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.

- 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.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 113 –
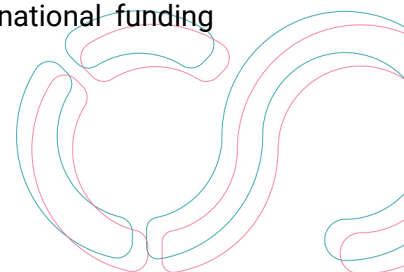
- ○ Stimulate dedicated funding streams or other measures (programmes, grant schemes, project support, financial and other incentives) that target the promotion and/or implementation of Open Science principles at institutional level.
- ○ Stimulate funding investments and operational costs of infrastructure(s) at national level contributing to EOSC.

6.3. Funding models

6.3.1. Status

Viable funding models are an essential element of ensuring an operational, scalable and sustainable EOSC federation after 2020. The Sustainability Working Group [EOSC_WG_Sustain] has taken an iterative approach to identifying funding 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 [WG_Sustain_Strawman] in September 2019 on which community feedback was gathered, leading to a 'tinman' report [WG_Sustain_Tinman], 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 [EOSC-Core_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 systems 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 includes 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 explore funding models for the full Minimum Viable EOSC, have been used to develop a third document, referred to as the FAIR Lady report, published in October 2020 [WG_Sustain_FAIRLady].

A unique added value of EOSC is its ability to provide support and access for researchers to reuse data alongside services through the same portal and this can only be achieved by bringing together all the elements of the Minimum Viable EOSC (MVE). Consequently, looking for sustainability in only part of the ecosystem would be a high-risk strategy and a missed opportunity to pursue the value-driven approach typical of platforms that has led to their fast growth in terms of impact.

6.3.2. Gaps

The EOSC-Core operational costs study and the use cases examined by EOSC-hub highlighted the fragmented and complex nature of the European research-funding landscape and the associated difficulties involved in attempting to provision services across borders. The majority of research in Europe is funded nationally. Funding sources are varied, complex and involve a large number of different rules, which contributes to suboptimal use of the combined Member States' investment in research resources. However, the demand for cross-border use

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 114 –
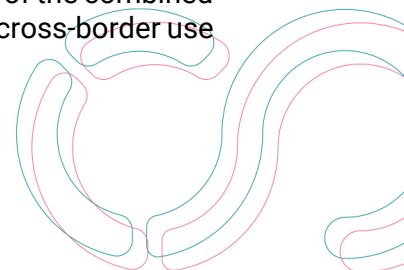
of research resources clearly does exist and will continue to grow, notably to address the Sustainable Development Goals supported by the OECD and UN [OECD_SDG; UN_SDG].

As stated in 'Prompting an EOSC in practice: Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC)' [EC_EG2_EOSC], the EOSC funding model is a critical non-technical element that will determine the success of the EOSC vision. The MVE, including the EOSC-Core, federated data and services and the EOSC-Exchange, is considered as an ecosystem to be sustained by a combination of platform-funding models. Platform-funding models create value by facilitating exchanges between two or more interdependent groups.

Two families of funding models need to co-exist, potentially applied to different sides of the platform or targeting different clusters of roles and players, in order to sustain EOSC: transaction-based models and patronage/membership-based 'learning' funding models.

Transaction-based models are widely known and build on the perceived value in interactions between different entities. The platform facilitates transactions, reducing their costs and/or by enabling externalised innovation. Use cases analysed by the EOSC-hub project [EOSC-hub_CBSvcs] highlighted that complex information needs to be accessed and exchanged before transactions between users and suppliers of research data, resources and services can be concluded.

EOSC will add value by providing frictionless, easy access to data and related services so that research communities can better connect with suppliers, users and funders. EOSC can also promote a cross-fertilising multi-disciplinary environment where investments can be efficiently leveraged and benefit from economies of scale.

Patronage/membership-based 'learning' funding models promote the perceived value based on being part of a community and finding help and support or networking capabilities for their members. For example, in EOSC this could mean offering private dashboards to each research organisation through which they can track their consumption over longer periods, allowing them to negotiate better terms with the resource providers. Similarly, resource providers would benefit from continuous interactions with (potential) users, generating a private flow of data and insights to better tailor their future offers.

6.3.3. Priorities

A workable funding model for EOSC leading to sustainable funding must be prioritised in the next framework programme. The funding models are currently under-developed, specifically in terms of enabling cross-border use of data and services, which will jeopardise uptake. If services are to be free at point of use, a national/EU funding model must be in place to ensure the costs incurred are recovered by the providers. It is not clear how any transactional model with service charges across borders will facilitate use and it could create an unsustainable overhead and barrier for users and providers alike. At the initial stage, the funding solution needs to be simple and effective, but still compliant within relevant regulations. In-depth studies and piloting such as the following are urgently needed.

The following activities have been identified as priorities in order to define a viable funding model:

- Perform cost assessments.
  - Assess cost estimates associated with the EOSC-Core services.
  - Assess cost estimates associated with the full Minimum Viable EOSC (MVE).
- Ensure sustainable financing for EOSC.
  - Develop financing schemes for EOSC.
  - Develop monitoring schemes for the in-kind contribution of members.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 115 –
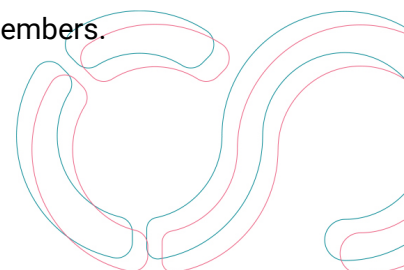
- ○ Develop synergies between national and EC funding streams as well as a higher level of coherence in the funding from different chapters (RTD, CONNECT) of the framework programme, and across the three pillars of Horizon Europe.

6.4. Skills and training

6.4.1. Status

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. The vision of a strong EOSC ecosystem that exploits digital technologies and has data and software at its core necessitates a comprehensive skills and education strategy. Skills and training are indeed essential for mainstreaming Open Science practices in research and thus essential for enhancing its quality and efficiency, leading to more new breakthroughs, sparking innovation and ultimately generating growth in the economy. It is therefore important to overcome existing gaps and barriers in the necessary skills and training quickly, to reduce the risk of Europe losing a leading position in Open Science.

A sustainable EOSC skills and training strategy must address different professional and research roles as well as their functioning in an organisational or team setting. The diagram in Figure 6.1 shows the various actors in the EOSC ecosystem and describes the main tasks related to each role. The report 'Digital skills for FAIR and open science' by the EOSC Working Group on Skills and Training details the digital skills required for those actors to practise or enable FAIR and Open Science [WG_Skills_DigitalSkills]. Digital skills are defined by the OECD as a range of abilities to use digital devices, communication applications, and networks to access and manage information [OECD_BDWC]. In the context of FAIR and Open Science, these skills include an understanding of data, software, tools and frameworks. Workforce capacity development is relevant to individual and institutional actors. Individual-level skills and competences form the basis, but 'the group as a whole is more than the sum of its parts'.[26] Research data, for instance, require collaboration across different roles and responsibilities. Understanding the EOSC ecosystem and the skills challenge calls for a clear definition of the appropriate profiles required to cover the complete research lifecycle and EOSC added value.

---

[26] Angus Whyte, Jerry de Vries, Rahul Thorat et al., D7.3: Skills and Capability Framework, 2018, p. 13 [EOSCpilot_D7.3].
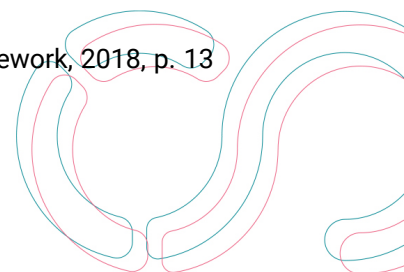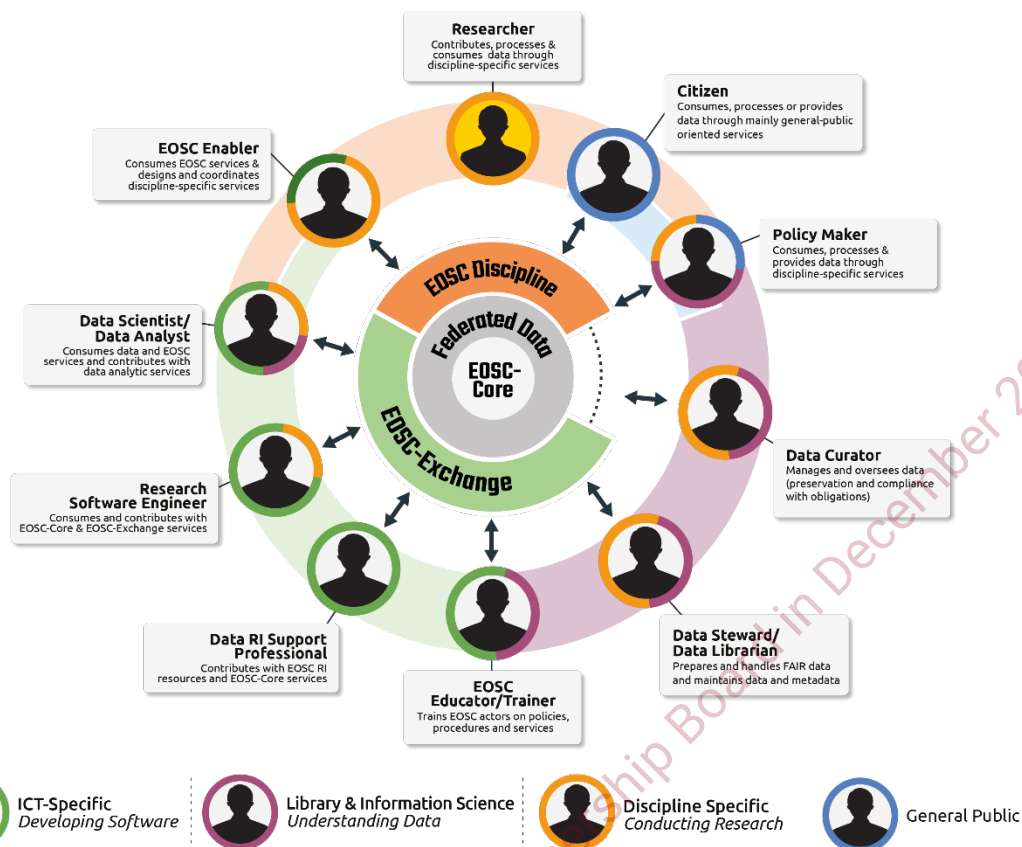
Figure 6.1: Actors in the EOSC ecosystem: roles and interactions

6.4.2. Gaps

Lack of Open Science and data expertise

Using or developing tools for handling data is becoming an increasingly important part of research.[27] However, at the moment, there are not enough adequately trained people to meet current demand for open and data-intensive science needs, let alone to meet increasing demand and diversity goals. Legal/IPR and data ethics expertise is a challenge even among the FAIR data / Open Science experts, while the research community is not equipped to explore opportunities presented in an interdisciplinary environment.

There is also insufficient support for the technological development of 'FAIR by design' needed for digital research object acquisition in all the research infrastructures and laboratories ('smart technologies'). This is a key activity to enable open data, open source software and FAIR paradigms to become a reality on a large scale, and in the near future. Interdisciplinarity, coordinated and coherent approaches to skills and competences building and of education and training provision is another area of concern. There is a need for a baseline approach for data stewardship.

Lack of a clear definition of data professional profiles and career paths for these roles

Data scientists, data stewards, data curators and research software engineers are some of the different actors needed for the development of data-driven, data-intensive science. It is necessary to provide recognition for these roles and define career paths that make them a viable choice.

---

[27] OECD report: 'The Digitalisation of Science, Technology and Innovation' [OECD_DSTI].
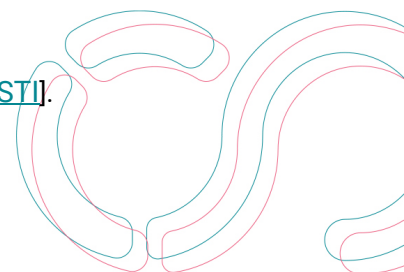
Although the reliance on the emerging new scholarly data and software support profiles are cornerstone elements in the implementation of FAIR data mandates, a very diverse and uneven picture is seen across Europe. A coordinated and coherent approach is needed to address this gap.

Fragmentation in training resources

Quality and FAIRness of training and learning resources remains a challenge. Fragmentation of existing training initiatives also reduces impact and there is a need to establish coordination with EOSC.

6.4.3. Priorities

To realise this vision of a strong research ecosystem with data and software at its core, EOSC has an important role to play in ensuring recognition of data professionals, propagating Open Science skills by aligning curricula, coordinating the collection of relevant learning materials and influencing strategic agendas at the national and European level. Four key priority areas have been identified, with a phasing plan in line with the development of EOSC:

1. Developing the next generation of Open Science and data professionals;
2. Coordinating training and aligning curricula for students and researchers;
3. Building a trusted and long-lasting knowledge hub of learning materials and related tools;
4. Influencing national Open Science policy for skills by supporting strategic leaders.

Priority 1: Developing the next generation of Open Science and data professionals

Developing the next generation of data and software professionals necessitates a set of activities which EOSC needs to actively pursue:

- **Enhance professional data career paths** with appropriate rewards and recognition.
- **Develop data skills profiles** through community understanding and consensus mechanisms.
- **Recognise data skills**.
- Provide a **quality assurance framework** and certification mechanisms for trainers and trainees.
- Facilitate and simplify **lifelong learning mechanisms** for up-skilling.
- **Establish Data Stewardship Competence Centres** (DSCC) through the establishment of cross-disciplinary and cross-national networks of experts, leveraging existing international models.

Priority 2: Bridging the education gap: coordinating and aligning curricula for students and researchers

Researchers are at the centre of EOSC and one of the most urgent priorities is to **equip them with the relevant digital skills to practice FAIR and Open Science.** Being self-sufficient to work with data is not the same as having self-service data and analytics. No matter how consumable the data is, researchers need to be curious and capable of understanding, questioning and taking the right action based on the insights delivered. This, in turn, improves their experience of and confidence in using data.

> **Use Case: Emerging approaches towards implementing FAIR data stewardship in Europe**
>
> In the past years, it has become clear that there is a large need for, and shortage of, individuals with Open Science and data stewardship expertise within research organisations. This is evident in all research domains and also transcends the institutional and even the national level. For this reason, a lot of effort has recently been put into professionalising data stewardship, bringing together national and global approaches, and both discipline-specific and discipline-agnostic efforts. However, the implementation of these data stewardship

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 118 –

competences, education and certification in higher education institutions (HEIs) and research-performing organisations (RPOs) is still in an early stage.

Substantial work related to data stewardship competences, training and education has already been done in two European countries, which may serve as the basis for the EOSC Skills agenda for data stewardship. In Denmark, national recommendations have been provided to gather evidence for supporting pre-qualifications for data steward education at universities [Wildgaard_2020] and the first Master in Data Stewardship is expected to start in September 2021. In the Netherlands, a community-endorsed data stewardship competency framework was developed and recommendations for national implementation of a national FAIR Data Stewardship Roadmap have been formulated, which can now be taken further by the national stakeholders [Scholtens_2019; Jetten_2021].

The roles and stakeholders in the data stewardship landscape that emerged from both efforts align very well and are shown in Figure 6.2 below.
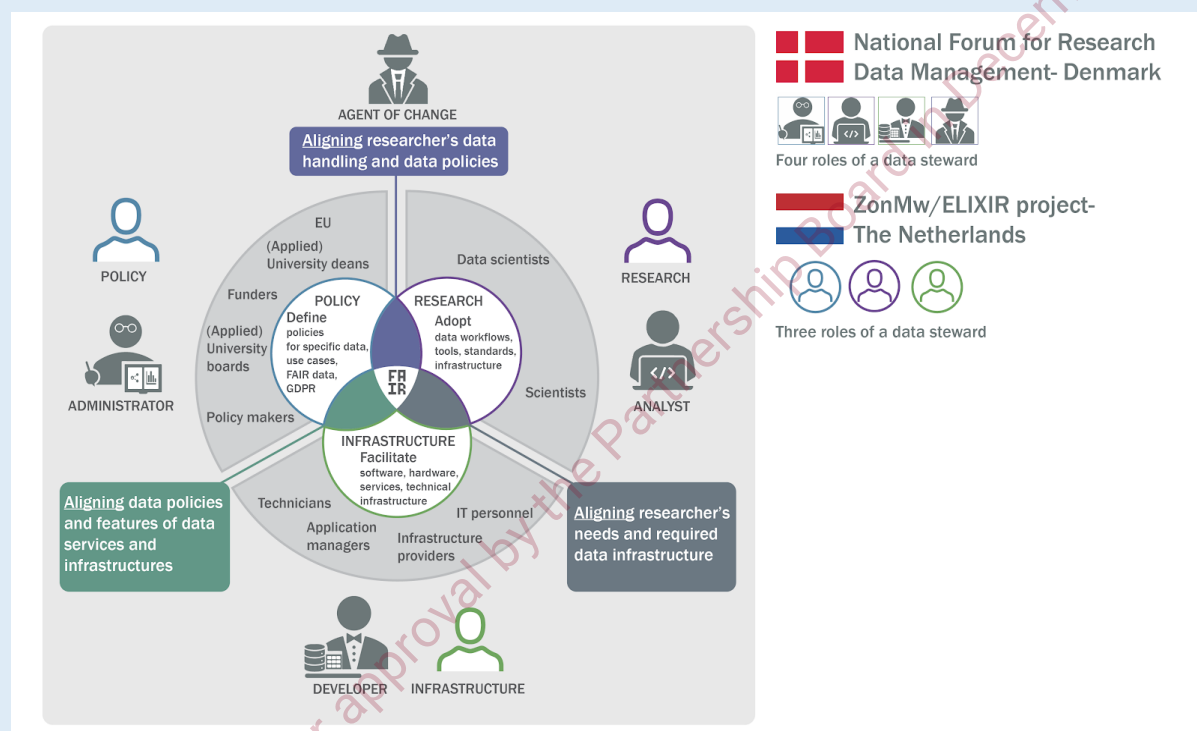
Figure 6.2: Roles of data stewards in the data stewardship landscape in Denmark and the Netherlands [DS_Roles]

EOSC is in an excellent position to coordinate and to support Member States in building the data stewardship capacity that is needed. Two major complementary actions are required to close the skills gap with respect to data stewardship, and EOSC can play a crucial role in both. First, it is essential to train future data stewards based on an accredited curriculum in Europe's universities. Second, current professionals in the field need to be (up-)skilled and proper career paths for data stewardship need to be defined and acknowledged. By ensuring research groups access the help of data stewards, research outputs will be more effectively managed and shared, ensuring higher quality, reproducibility and more consistent long-term sharing and reuse.

Activities to be considered within the realm of EOSC include:

- Align curricula and training with demand.
- Support communities to tailor generic materials to be more relevant to specific disciplinary and professional practices.
- Support and align with other EU and national programmes (e.g. Erasmus+) to help organisations (e.g. libraries, NGOs) engage and up-skill all levels of researchers (e.g. supervisors and mid-career researchers, teachers in higher education institutions) as well as the public (citizen scientists).
- Explore and align mechanics in rewarding early career researchers for Open Science practices in evaluation processes and awarding efforts with ECTS[28] or other formal certificates.
- Build and operate a network of researchers-champions in open science.
- Promote and support advanced learning environments as part of the broader Open Science agenda.

Priority 3: Building a trusted and long-lasting knowledge hub of learning materials and related tools

A key goal for EOSC is to build capacities to sustain learning corpora for data skills and tooling, with activities that include:

- Develop a quality assurance and certification framework for learning material.
- Devise a common framework for learning pathways for different Open Science and data-related profiles.
- Support the development of an EOSC Knowledge/Education Hub as a set of interconnected and decentralised learning platforms and living repositories for knowledge sharing.
- Facilitate the adoption of open learning environments.
- Promote and support innovative ways of learning by employing creative methodologies and technology for teaching/training at all levels, from awareness to focused expertise.[29]

Priority 4: Influencing national Open Science policy for skills by supporting strategic leaders

In order for EOSC to achieve its aims it is necessary to ensure both the availability of highly and appropriately skilled people, and a policy environment that supports digital skills for FAIR and Open Science. Steering the skills agenda at a national and European level requires a community of influencers with the ability to affect change within their environments to ensure that policy decisions deliver success.

This priority is unique in its focus on influencing national policies. This will help countries to create, update and coordinate their national Open Science and data skills policies and activities. But most importantly, it will create the desired culture shift in the policy-making community, creating a generation of visionaries, able to make the connections with emerging technologies (e.g. AI, HPC) and to quickly adapt curricula and training. It will draw on existing Open Science leadership, ambassador and champion programmes such as those operated by YERUN, LERU, IARU, LIBER, SPARC Europe, OpenAIRE and RDA, etc. These advocates will provide a bridge between Open Science best practices and the policy as influenced by the senior political and administrative leaders targeted by this activity.

Its aims will include:

---

[28] European Credit Transfer and Accumulation System, usually used for students (sometimes PhD level, but not everywhere) [ECTS].

[29] Example Google game for data science/Machine Learning [What_If].
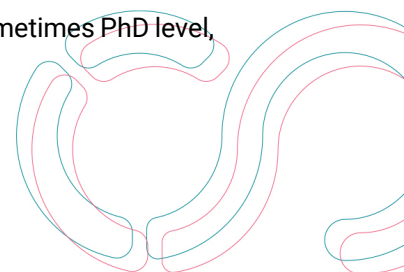
- **Leveraging existing Open Science ambassador programmes** and champions to provide senior leaders with relevant insights to influence national policy.
- Providing a **peer network and forum for senior policy leaders**, to enable them to lead transformation initiatives in their country or their organisation. Alignment with the CoNOSC group is recommended here.
- Assisting national stakeholders to **align skills initiatives across Europe**, within academia but also with other sectors.

6.5. Rewards and recognition

6.5.1. Status

Present-day rewards and recognition (R&R) systems are largely shaped by government-mandated national and institutional policies and regulations, but they are also stimulated by the competitive environment in which researchers and research-performing institutions compete for funding and other resources. In addition, the driving forces coming from ranking the results of institutions may lead to stimulations related to the metrics used by the rankers. Many R&R systems currently used by research-performing and research-funding organisations tend to incentivise and reward too narrow a range of academic activities – e.g. publishing in prestigious 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 [Cohen_2019; EUA_RATOS] and facilitates further competition (and waste of resources) instead of collaboration. Moreover, these R&R systems focus on past performance, often not addressing future potential. Finally, individual excellence is often emphasised over collective excellence or team science.

Although guiding principles for good practices in research evaluation have been developed, the practical implementation of these is often lacking. Examples are the Leiden Manifesto [Leiden_Manifesto] presented in 2015, and a recent initiative from the Declaration on Research Assessment (DORA), summarising five design principles [DORA_1] that help institutions experiment with and develop better research assessment practices. The DORA initiative maintains a curated list of good practice examples of implementation from institutions showing leadership in this area [DORA_2].

Other initiatives re-evaluate the whole process of research assessment, such as the Association of Universities in the Netherlands (VSNU) position paper 'Room for everyone's talent' [VSNU_PP].

6.5.2. Gaps

To advance from the status quo, a culture change needs to be realised as current R&R systems poorly reward and recognise not only Open Science and FAIR practices, but also aspects of education, research, impact, science communication 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 researchers should have a better balance in valuing achievements in education (where appropriate); research; influence (on science, society, economy, environment and teaching); organisation and leadership. In particular, and of direct relevance to EOSC, recruitment, evaluation and promotion criteria should recognise openness and FAIR practices, as this is vital for advancing the good practices in research data management that underpin EOSC [CESAER_WP_RDM].

In addition to the evaluation of researchers and research there is an ongoing discussion in the community about 'the third space' – 'the in-betweeners' – people who do not fit the traditional categorisation of support staff vs. academics, in roles such as data stewards, data managers

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 121 –

and research software engineers. They often have a research background (many have PhDs and postdoctoral experience) but have different skills and choose different career paths. However, they clearly contribute to the team science's collective success, albeit in a different manner. This divide between support staff and academic staff affects the perception of people and the contributions they make. It may, however, also result in issues such as difficulty with residence permits for support staff, difficulty with visas, tax issues, inability for support staff to apply for grants, etc. This is an issue not only for Open Science. However, the new professions resulting from the development of Open Science increase this dilemma.

All these developments are connected to the broader topic of supporting and rewarding modern research (and research support) careers, where further gap analysis combined with practical guidance and best practice examples are available from a range of sources. Notable examples include the advice paper 'Open Science and its role in universities: A roadmap for cultural change' [LERU_AP24_OS], the position paper 'Room for everyone's talent' [VSNU_PP], the Dutch 'Strategy Evaluation Protocol 2021-2027' [VSNU_SEP], and the white papers 'Sharing Experiences with the Human Resources Strategy for Researchers' [CESAER_WP_HRS4R] and 'Boost the Careers of Early-Stage Researchers' [CESAER_WP_BCESR].

6.5.3. Priorities

Overall, two sets of priorities can be distinguished: internal priorities for organisations engaging directly with rewards and recognition systems (including research-performing and research-funding organisations) and external priorities for organisations that set broader rules and framework conditions (e.g. EU institutions and national and regional governments).

For internal priorities, several broad lines of action can be identified, including: (i) demonstrating leadership in enacting change towards a culture of trust, openness and risk taking; (ii) preparing Human Resources (HR) to adjust R&R structures (including approaches to recruitment and promotion) using (iii) next generation and progressive metrics [CESAER_WP_NGM].

For external priorities, broad lines of action include: (i) safeguarding institutional autonomy; (ii) identification of legal obstacles; and (iii) removal of legal obstacles to empower researchers, research-performing organisations and research-funding organisations to experiment with, develop and refine better R&R systems. Examples include ensuring sustainable funding levels and that labour laws, migration rules, social security schemes and pension systems support, and do not hinder, modern and diverse careers within academia.

Organisational rankers (e.g. those who publish 'best university' lists) is a category that falls between the two above, as they themselves rarely interact directly with researchers and their R&R systems, but they also do not have the formal mandate of governments. Nevertheless, rankings have an outsized impact on the behaviour of the research community, and EOSC should engage closely with rankers to ensure that any influential rankings that still use narrow and outdated criteria embrace instead a modern and diverse understanding of what makes excellent research and research teams. Moreover, their handling of data is in most cases not in line with sound scientific practices (cf. the findings of the INORMS Research Evaluation Working Group [INORMS_RtR]).

EOSC, as one of the organisations stimulating Open Science, can help in modernising R&R systems aligned with the priorities outlined above by direct action and by promoting action from – and working with – other actors.

EOSC should work with **researchers, research-performing organisations and their leadership** to:
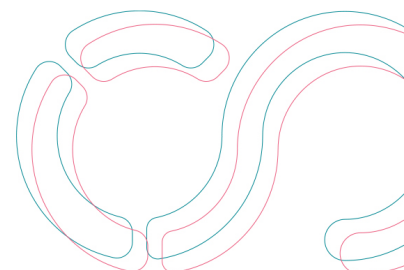
- Change the narrative around careers in academia that currently focus on individual competitiveness to focus instead on collaborative and team-based approaches, including the importance of promoting equality, diversity and inclusion;
- Embrace the five design principles outlined in the one-page DORA briefing paper [DORA_1] to help institutions experiment with and develop better research assessment practices: (i) instil standards and structure into research assessment processes; (ii) foster a sense of personal accountability in faculty and staff; (iii) prioritise equity and transparency of research assessment processes; (iv) take a big picture or portfolio view towards researcher contributions; and (v) refine research assessment processes through iterative feedback;
- Embrace a culture of quality, trust and risk taking and reduce the focus on 'narrow metrics';
- Involve early-stage researchers in scientific leadership and governance, for example as outlined in the white paper 'Boost the Careers of Early-Stage Researchers' [CESAER_WP_BCESR];
- Follow best practices such as the examples in the white paper 'Sharing Experiences with the Human Resources Strategy for Researchers' [CESAER_WP_HRS4R];
- Sign and /or implement the Declaration on Research Assessment (DORA) [DORA]; the Leiden Manifesto [Leiden_Manifesto] and the Hong Kong Principles [WCRI_HKP].

EOSC should work with **organisational rankers** to:

- Make their methodologies and data-wrangling processes more transparent, verifiable and thus reproducible;
- Reduce the focus on 'narrow metrics' to promote a culture of quality, trust and risk taking;
- Change the narrative in rankings around 'excellence', which is currently often defined based on individual competitiveness, to focus instead on collaborative and team-based approaches, including the importance of promoting equality, diversity and inclusion.

EOSC should work with the **EU institutions** to:

- Make an inventory of legal obstacles to improving researcher careers and mobility at the European, national and regional levels;
- Empower individuals, universities and other research-performing organisations to challenge and remove barriers; for instance, following the example of the copyright retention strategy of cOAlition S [Plan_S];
- Provide support for pilot projects to develop new approaches to evaluating academic performance and to improve staff employability (especially for early-stage researchers);
- Promote wider transformation of national frameworks, using EU funding programmes and various soft instruments as a 'lever' to promote open, transparent and merit-based recruitment processes and recognition for all aspects of academic work;
- Create interaction between countries on the topic of R&R systems to facilitate learning and sharing of good practice.

EOSC should work with regional and national governments to:

- Produce at least a country-level inclusive approach to research evaluation;
- Provide resources and sustainable funding levels to ensure attractive, stable and stimulating working conditions for researchers, teachers and support staff;
- Direct national funding agencies to adopt best practices in modern research evaluation (examples include the Leiden Manifesto [Leiden_Manifesto], DORA [DORA] and the Hong Kong Principles [WCRI_HKP]).
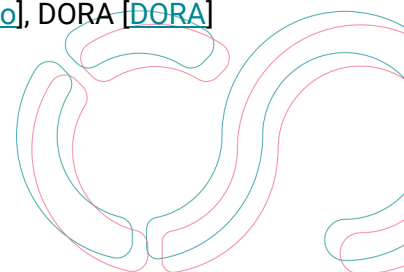
6.6. Communication

The Executive Board established a Task Force on Communication to provide clarity on the why, how and what of EOSC, and to set up these messages in a consistent way. The Communication Task Force consisted of members coming from the Executive Board, Governance Board, and communication experts from the EOSC Secretariat, European Commission and elsewhere. There are three areas of importance in communications: stakeholder engagement, content production, and branding and positioning of EOSC for the different stakeholders.

6.6.1. Status

There is a toolkit for communication [EOSC_Comms_Toolkit], including a new template for (PowerPoint) presentations, using the current logo. The logo can be used by EC, EOSC and projects and initiatives that work on EOSC. There is a protocol for the use of the EOSC logo by (EC) projects and external entities [EOSC_Brand_Guidelines]. This policy was approved by the Executive Board, and published on the website of the EOSC Secretariat. A standard presentation on explaining EOSC for different stakeholders is also available.

Nine different groups of stakeholders have been distinguished [EOSC_Landscape]. For communication purposes, these groups can be aggregated into three main stakeholder groups:

- 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 and research initiatives, such as RDA, offering global platforms for sharing expertise;
  - Cloud providers, including commercial parties such as Amazon, offering services to research;
  - Cloud community.
- Research Performers:
  - Research communities;
  - Research-performing organisations.
- Research Funders:
  - Research funders;
  - Policy makers.

In a later phase, these will be expanded to include citizens, small and medium-sized enterprises (SMEs) and other societal groups as users and enablers.

6.6.2. Gaps

The following gaps and requirements have been identified:

- EOSC addresses researchers, but also policy advisors and research funders, as it comprises a system change. This diversity of stakeholders requires a communication policy that addresses the different needs of these groups.
- Each stakeholder group described above 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 and efficiency issues, whereas researchers and providers focus on functionalities.
- The focus should be on further content production and elaborating key messages for the different stakeholder groups. One of the required actions will be to explain EOSC via Q&As, use cases, best practices, etc.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 124 –
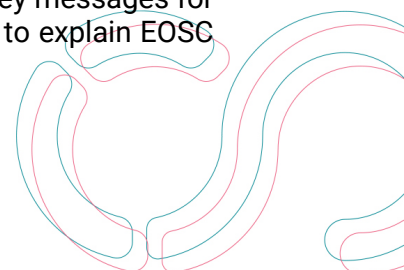
- Now that the EOSC legal entity has been established, consideration will need to be given to ownership, licensing and liability issues, both within the legal entity and in its relationships with EC projects, service providers and other contributors to EOSC.
- A communication strategy will need to be prepared, including branding, positioning and promoting tools and services once these become available.

6.6.3. Priorities

Current priorities are taken from the EOSC Work Plan 2020, the Task Force Agenda and the Open Consultation on EOSC:

- Set up a Strategic Communication Plan.
- Inform stakeholders about the developments of EOSC:
  - Perform an in-depth stakeholder analysis;
  - Provide key messages for the different stakeholders;
  - Develop stakeholder messaging that is impactful (addressing the why) and functional (addressing the how and the what);
  - Engage with these communities.
- Develop and deploy communication channels:
  - Including the Web Portal (cf. EOSC-Core description in the FAIR Lady report).
- Work on licensing and ownership issues.

6.6.4. Considerations

Further considerations that should be taken into account with regard to communication relate to:

- Communication strategy;
- Stakeholder analysis;
- Communication channels;
- Licensing.

Each of these is discussed below.

6.6.4.1. Communication strategy

The EOSC legal entity will produce a strategy and multiple annual roadmaps. Based on these documents, EOSC can set up a communication strategy to build an engaged user community and connect with users, its service providers and the membership, including funders.

6.6.4.2. Stakeholder analysis

To provide adequate information and content, an in-depth analysis of the three major stakeholder groups is required. It should result in key messages that are tailored to the different groups and be the start of building communities.

6.6.4.3. Communication channels

EOSC will develop and deploy a variety of channels (listed below) tailored to its different stakeholders and taking into account different levels of readiness. The core communication channel will be the EOSC website.

- Website
  The EOSC AISBL legal entity acquired the domain name eosc.eu. The EOSC website will be a gateway for information on EOSC and how to participate, will direct to the EOSC services, and will also integrate social media channels.
- Social media channels
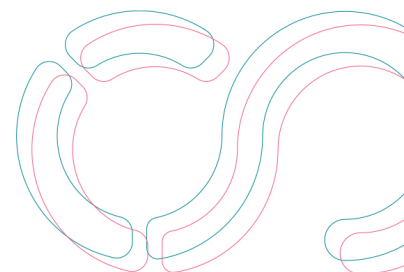
EOSC will focus on Twitter and LinkedIn. Twitter will be used for high-frequency but brief messages providing a link to the website for more information. LinkedIn can be used for less frequent but in-depth stories that can stand on their own.

- Blogs
Blogs will be used to communicate opinions from the EOSC Secretary General, President and other Board members. They will be published through LinkedIn, Twitter and the website. These can be combined with Videos for marketing.
- Webinars
A series of EOSC webinars will be set up, to which early-adopting service providers will be invited, to show and explain EOSC services, initiate discussion on upcoming topics, provide clinics (cf. EOSC Symposium 2020).
- Annual EOSC Symposium
The annual symposium will offer stakeholders the opportunity to co-locate (when face-to-face meetings are possible).

6.6.4.4. Licensing

As EOSC will provide information on services, data and tools, it must ensure that IPR and publishing content are secured. This applies to (data) catalogues containing metadata, controlled vocabularies and thesauri. EOSC may also consider trademarking its name and some of the key tools.

6.7. Widening to public and private sectors and going global

To realise its full potential, EOSC will be widened beyond the European research community to the broader public sector and to the private sector, while also developing its global reach.

6.7.1. Widening to public and private sectors

6.7.1.1. Status

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.

EOSC should aim to expand to include public and private stakeholders who form part of the wider EOSC knowledge ecosystem, thus enabling further excellence by the European research community. The affordances created through the expansion of EOSC would:

- Stimulate novel research methodologies and support research excellence;
- Enhance existing research practices through greater access to data based on FAIR data principles;
- Encourage the development of novel Open Science research exchanges that allow interdisciplinary and international collaborations, and open up new categories and fields of knowledge;
- Establish a mechanism for a technology transfer that is grounded in the research communities' ethos and principles;
- Enable a marketplace for exchange of knowledge and datasets, established upon FAIR data implementation guided by the research communities' values.

A targeted study has been conducted by the Industry Commons Foundation / MTF Labs AB on behalf of the Sustainability Working Group with the objective to deliver practical, actionable advice and models for technology transfer and engagement with existing and potential scientific research user groups outside of academia, and to scale the impact of EOSC and further incentivise and reward its community of researchers and research institutions.

The study identified use cases resulting from cross-domain, data-driven applications created in pan-European collaborations by research communities, citizen scientists, public sector

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 126 –
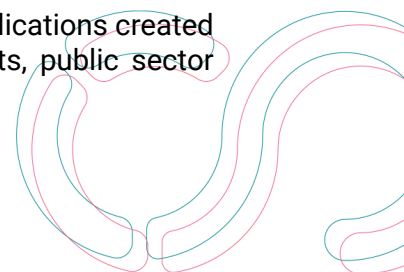
organisations and industry. Furthermore, new and emerging case studies from grassroots innovation communities, industry demonstrators and European projects were selected in close collaboration with the Sustainability WG. Following stakeholder feedback, additional valuable use cases from a variety of domains and areas of application have been included. This has resulted in a total of 23 use cases, of which a selection is highlighted below.

Big Data Value Association (BDVA): ICE Datacenter Gold i-Space

The ICE Datacenter Gold i-Space provides testing in a flexible full-scale datacentre – without large-scale investment, with access to massive amounts of research data and with an on-call team of world-leading scientists who can contribute to an organisation's innovation activities. Green Computing,[30] along with space data and other types of datasets, demonstrate the potential for industry engagement with e-infrastructures and a model for working with large research datasets for the private sector.

COVID-19

Europe's readiness for future pandemics is of utmost importance and should be addressed to ensure the preparedness of infrastructure, building on already-existing frameworks, such as the COVID-19 Data Portal, for broader use such as EOSC. The COVID-19 use case supports the widening of EOSC to the public and private sectors and helps fast-track the global visibility of EOSC. Integration of molecular research data with sensitive patient and clinical data will ensure that patients benefit directly from the research supported by EOSC. Europe's industry, including SMEs, will access data and deposit data in the public domain. Cross-linking with socio-economic, societal response and other social science and humanities will promote an integrated understanding of European outbreak response and preparedness, and demonstrate the value of FAIR data to society and public engagement during a global public health crisis.

Human Rights Data: Cambridge Whisper

This use case features the collection and processing of highly sensitive and confidential data through interviews with refugees about their personal experiences of human rights abuses. It demonstrates the potential for unique tools that build upon the EOSC framework and portal, allowing specific scenarios with software requirements that model best practice in the tools themselves.

Industry OntoCommons: Siemens Complex Equipment

This use case describes and analyses the digital twin of products/industrial assets in manufacturing and energy industry across their lifecycle from design to service, based on IT systems. It demonstrates the importance and centrality of FAIR data in industry and the potential for EOSC to act as a Web of FAIR Data in a context within which industry is developing ontological interoperability.

Neurofeedback Patient Data

Clinicians collect and process large amounts of patient data from EEG brain wave monitoring. There are significant challenges in storage and analysis of this data and enormous potential for anonymised data sharing that would reveal larger patterns and more nuanced

---

[30] A number of Sustainable Development Goals are addressed by the European Union Green Deal with respect to computing: Affordable and Clean Energy – since energy is central to nearly every major challenge and opportunity in computing infrastructure. Industry, Innovation and Infrastructure – since investments in green infrastructure are crucial to achieving sustainable development. Sustainable Cities and Communities – since there needs to be a future in which cities provide opportunities for all, with access to basic services, energy, housing, transportation and more, and it will be based on computing, including both large datacentre infrastructures and network edge and device computing.
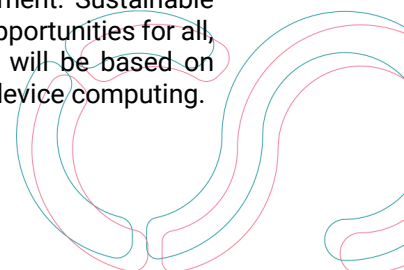
understanding. This use case highlights the potential for EOSC to act as an intermediary Web of FAIR Data verification platform between non-academic professional researchers.

Ocean Data

This use case concerns navigating complex datasets and studies across a wide range of disciplines in the EU Oceans Mission in order to initiate agile and adaptive prototyping projects that give both citizens and industry the tools and autonomy to engage with and respond to a richer understanding of seas and oceans. It demonstrates the potential for academic research to engage with citizen users in order to collaboratively address local challenges as well as those that affect industry and the environment.

Open Media – European Broadcasting Union

This use case concerns the promotion of EU digital sovereignty and means of preserving and promoting the cultural and historic value of European public media archives. It describes a multiplier effect for news gathering and provision by providing instantaneous translation and targeted news aggregation and verification. The use case raises questions about the e-infrastructure offering of data storage and processing at scale in competition with commercial providers for use in a public service media context.

PaNOSC

This use case contributes to the realisation of a data commons for neutron and photon science, providing open data services and tools for data storage, analysis and simulation, for the many scientists from existing and future disciplines using data from photon and neutron sources. It demonstrates the potential for innovative SME bridging organisations to translate large amounts of specialist scientific data to meet the needs of industry research and product development, and the potential for new markets to emerge based on European research.

Sentinel Hub: BlueDot Observatory

This use case features SMEs leveraging global monitoring of water bodies on a shoestring through API access. It highlights the commercial and societal potential for European open research data, but also the challenges faced by EOSC to act as an intermediary and an enabler in this context.

Višnjan Observatory: Citizen science

As a member of the International Asteroid Warning Network (IAWN), Višnjan is amongst the top five observatories in the world in collecting more near-Earth object (NEO) measurements to determine if they are a threat to Earth. Without these follow-up and confirmation measurements the majority of newly discovered asteroids that are daily discovered, mainly from Hawaii, would get lost in a day or even in a matter of hours. Measurements are taken to ascertain if the discovered object is really there, calculate its trajectory and verify whether it is a potential threat. Višnjan is a member of Spaceguard Foundation, an association that supports the creation of a system to discover celestial bodies that could potentially be a threat to life on Earth. The use case demonstrates the impact and scientific gravitas of citizen science projects that exist outside academia and the potential for recognition and support through non-monetary incentivisation mechanisms and acknowledgement.

6.7.1.2. Gaps

The key findings of the study include the following, which need to be further developed:

- Industry feedback indicates that EOSC should act as the validating organisation for industrial FAIR data as well as for data produced and used by research communities.
- The addition of JUST (judicious, unbiased, safe and transparent), which highlights accountability by a responsible researcher, has been equally well-received by all interviewed stakeholders.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 128 –
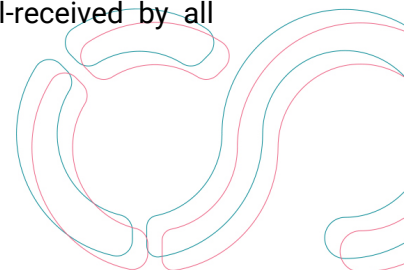
- The broader academic research community has requested that the EOSC front end be a live, audiovisual platform for remote collaboration, inclusive of access to research data and value-added services (which can be added at a premium).
- An additional important stakeholder group has been identified in professionals working with large valuable datasets (e.g. clinicians) who wish to be part of the EOSC marketplace.
- The strategy for EOSC expansion based on knowledge circles has been universally supported by all interviewed stakeholders.

The study results provide the foundations for the definition and programming of reward systems (ontological and programmatic), financial sustainability and business models for FAIR data services beyond the Minimum Viable EOSC (MVE). 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 inner circles of EU consortia, PPPs, to sector-specific and citizen bodies, and further on to citizen engagement groups.

The study produced the following recommendations.

Web of FAIR data

A key recommendation emerging from this study is that for EOSC to have the greatest impact and reach to external stakeholders it must establish itself as the Web of FAIR Data as its primary USP. Validation and interoperability of data in knowledge transfer and technology transfer are key to its centrality in the application (and collection) of research data from beyond the realms of academia. Note that this also works in the clinical example as well as industry to industry and in all cases where SMEs could build innovation on top of existing data. It also provides an incentive and an imperative to make as much European research data as possible – both new and historical – available in this ecosystem. The expertise of FAIRification should be a standard for all European Marketplaces including GAIA-X, Industry Commons and the new planned EIC marketplace, thereby supporting EOSC's key role and future sustainability.

EOSC-Future (INFRAEOSC-03)

The INFRAEOSC-03 funded project should be used to initiate, implement or prototype, as appropriate, a series of recommended actions. The following are potential examples of what can be tested through this project:

- PaNOSC value-added SME application for industry use.
- EOSC as Community Engagement Platform: Pan European Association of Citizen Scientists.
- A One-Health approach to the COVID-19 pandemic building on the latest technological advances, e.g. federating research, patient and clinical data between national centres.
- Dynamic multi-modal tools for online collaboration (with optional added-value applications and e-infrastructure provision).
- A marketplace for pan-EU media applications in partnership with EBU.
- Creation of SME-led automatisation and customisation layers on top of EOSC e-infrastructure (e.g. AirBnB for compute services).
- Integration of intellectual property tracking.

6.7.1.3. Priorities

The SRIA consultation exercise placed this Action Area lowest in terms of relevance for the immediate future. This aligns with plans to only widen EOSC after the programme has successfully engaged and delivered a functioning platform to European research communities first and foremost.
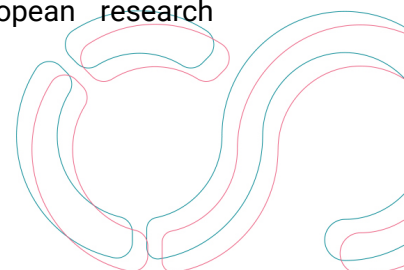
The following priorities have been identified:

- 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 , 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.

6.7.2. Going global

6.7.2.1. Status

As noted in Section 2.6 International dimension, EOSC operates in a global ecosystem with the clear aim to promote the 'Open Science, Open Innovation and Open to the World' principles in its international activities. Around the world, regional and national Open Research Data Commons and Open Science Clouds are being developed and several important international policy agreements and initiatives have been established that demonstrate the importance of international cooperation in Open Science. These include the G7 Expert Group on Open Science [G7_OS], the OECD Principles and Guidelines for Access to Research Data from Public Funding [OECD_ARDPF], which is currently being updated, and the Research Data Alliance [RDA] and GO FAIR [GO FAIR] initiatives. The common vision embodied across these international developments enables Europe to enhance scientific cooperation and collaboration with other parts of the world and drive a cultural change towards Open Science based on agreed principles.

6.7.2.2. Gaps

Through global cooperation, Open Science has the potential to effectively address many new scientific questions, as well as revisting some long-standing problems. This is particularly true for a number of pressing contemporary challenges. The following are areas where international cooperation is of particular importance:

- Activities with special relevance to complex societal challenges such as climate and sustainable development goals;
- Issues of scarcity such as limited and sporadic amounts of data (e.g. rare diseases); limited availability of the research subject (e.g. rare-earth elements or metals); or a small talent pool in a unique research field (e.g. ITER, black holes, etc.);
- Research fields where the talent pool is very dispersed (e.g. Arctic research);
- Screening for unique solutions developed by local communities, such as indigenous groups;
- Scientific observations resulting from synergies between enabling technologies (e.g. from sensor to satellite such as oceanography).
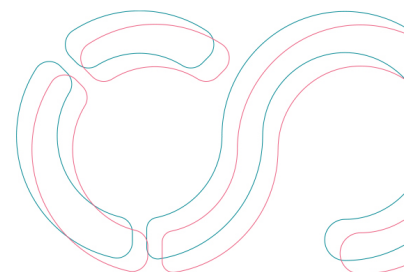
6.7.2.3. Priorities

For the above issues to be addressed, and the full potential of EOSC and Open Science to be realised through a global approach, the following priorities have been identified, taking into account the need to adapt to and consider diverse capabilities and demands, and the principles outlined in Section 2.6.

- Promote an international Open Science culture and the need for change in the reward systems to support the transition of other world regions towards Open Science, where certain regions with less developed research ecosystems could leapfrog. EOSC members, especially infrastructures with already existing international cooperation, are particularly suited to address this.
- Initiate an international data steward network across domains to exchange best practices and success stories.
- Promote the uptake of the building blocks of EOSC (such as the EU ICT technical specifications [EC_ICT_TechSpec], the rolling plan for ICT standardisation [EC_ICT_Standard], FAIR, PIDs, AAI, APIs, CoreTrustSeal, etc.) and open source solutions abroad, given that formal standardisation is difficult in the current fast-changing, open source environment of research.
- Promote the EOSC service portfolio abroad, such as the EOSC-EarthOb, which will enable the use of Copernicus and Galileo data more easily, particularly relevant for third countries.
- Provide state-of-the-art trainings on technical requirements of the Horizon Europe calls, such as data management plan (DMP), FAIR, Open Access, to enhance third-country participation and success in Horizon Europe calls.
- Initiate EOSC Rules of Participation (RoP) for service providers from third countries, noting that compliance with applicable legislation is a prerequisite beyond the RoP.
- Develop value propositions to third country service providers, to widen the EOSC portfolio.
- Encourage emerging regional Open Data Commons in countries/regions with commitment to Open Science, eligible for the EU Development Funds.
- Initiate partnerships via Memoranda of Understanding (MoUs) with other Open Data Commons that enable users of each initiative to access the resources of the others. Cooperation with these initiatives should be found at an institutional level, to establish a level playing field, and enable a good user experience.
- Propose a Global Open Data Commons Charter which paves the way to a Global Open Data Commons. This should be developed in close cooperation with the RDA Global Open Research Commons Interest Group, which brings together a number of actors from the relevant initiatives.
- Give support to and cooperate with existing international projects and initiatives, build on their work and contribute to their mission (e.g. Data Together which is comprised of CODATA, GO FAIR, RDA & WDS).
- Enable the formation of international consortia for Horizon Europe calls.
- Systematically embed the sustainable development goals (SDGs) into the EOSC Annual Work Plan and activities, as well as the overall strategic goals of the Horizon Europe programme.

Sections 5 and 6 have described key action areas involved in deploying the EOSC ecosystem. The next section considers the benefits that are expected to result from its deployment.
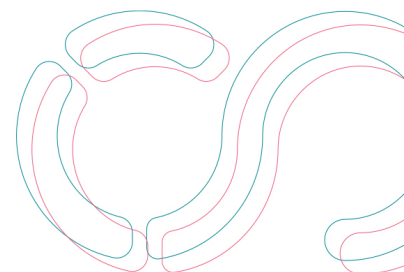
7 Expected impacts

The climate crisis, the extinction of species, global poverty and social inequality are only a few of the challenges that humankind is facing in the 21st century [EC_HE_Missions]. Research plays a crucial role in addressing these challenges and, against this background, EOSC will be a major European vehicle for joining forces to help transform individual research efforts into collective efforts. EOSC will also help to fill infrastructure as well as social gaps in unstructured areas, and play a significant role in raising to the most advanced level the science domains that have unsatisfied e-needs, with the target to increase levels of integration.

Recalling the Objectives Tree presented in Section 3, the final row of the tree describes the benefits of EOSC for the three dimensions of Science, Industry and Society:

European Open Science Cloud Objectives Tree

PROBLEMS	Public & private sectors do not exploit Open Science for improving quality and productivity of research	Researchers do not combine and build upon ever-growing available scientific results	National, European and global infrastructures do not share Open Science standards and practices
	PEOPLE	DATA	INFRASTRUCTURES
BARRIERS	Absence of incentives, rewards and skills for open sharing stifles the uptake of Open Science	Scientific results are unfindable, inaccessible, not interoperable, and often used only once	The scientific landscape consists of national and disciplinary research silos and infrastructures
	OPEN	FAIR	FEDERATION
OBJECTIVES	Open Science practices and skills are rewarded and taught, becoming the 'new normal'	Standards, tools and services allow researchers to find, access, reuse and combine results	Sustainable and federated infrastructures enable open sharing of scientific results
	SCIENCE	INDUSTRY	SOCIETY
BENEFITS	Improved trust, quality and productivity in science	Development of innovative services and products	Improved impact of research in addressing societal challenges

Figure 7.1: EOSC Objectives Tree – benefits

This section considers the impact of EOSC on each of these areas. Section 7.1 addresses the impact of EOSC in improving trust, quality and productivity in science; Section 7.2 looks at the development of innovative services and products; and Section 7.3 discusses the role of research in addressing societal challenges. To conclude, Section 7.4 summarises the critical success factors that must be in place for these benefits to be realised.

7.1. Improved trust, quality and productivity in science

Encouraging collaboration and openness

EOSC will stimulate the cultural changes in the entire research ecosystem. Open Science, which is realised with the help of EOSC, is striving for better horizontal and vertical links between scientists, scientific institutions, research and data infrastructures, and interconnecting scientific disciplines. It equilibrates the traditional research outputs, such as publications, patents, etc., with other forms of research outputs, including, for example, data, software, including models, simulations and methodologies. Making these outputs as

findable, accessible, interoperable and reusable (FAIR) as possible is therefore a key requirement in measuring and rewarding the contribution of research.

Open Science and EOSC will have a significant structural effect with the potential not only to change the way research is performed, by creating a pan-European, multi-disciplinary federation of research infrastructures supporting a broad range of a researcher's data and computing needs, but also to enable new mechanisms for communication and evaluation of research, motivating researchers, institutions and national research systems to open their research outputs.

Trusted frameworks for data availability and security

The foundational fact – indeed, prerequisite – that EOSC provides a secure, safe and transparently trusted virtual environment where scientific outputs can be deposited and found according to the FAIR principles, represents a significant change that will impact the overall quality of research. It unlocks the full value of research and, by developing certified services and standards, will enhance the quality of knowledge management, data discoverability and reuse. EOSC will also underpin the development of new ways to deal with open access to all forms of research outputs, with automated access guided by clear and transparent Rules of Participation that ensure trust in the quality of data and the function of data access services. Researchers will therefore be able to make their data open in the knowledge that their work will be acknowledged, their intellectual property (IP) will be protected where appropriate, and that sensitive data will also be appropriately protected where necessary.

Infrastructure planning

Alongside the direct impact on science, EOSC will also contribute to the quality of research by reducing the disparities in the Open Science readiness in different countries, reducing the divide across regions and mobilising important resources that will federate national data systems, enabling new actors to foster data interoperability with a high level of interdisciplinary research. The pan-European EOSC will also positively influence the planning of institutional and national infrastructures by developing synergies and compatibility schemes with other existing infrastructures, improving the quality of the integrated research landscape, increasing researchers' ability to provide science-based solutions to complex societal challenges.

Broadening discoverability

EOSC will facilitate integration not only within scientific domains but also across domains, offering a trusted and stable ecosystem for linked Open Science. Even within their own field, researchers face challenges in discovering, locating, accessing and reusing relevant data. EOSC will address these challenges in two ways: first, by making data FAIR to enhance discoverability; and second, by federating research infrastructures so that relevant datasets and thematic services from particular fields are more widely exposed, encouraging multi-disciplinary research.

Making new connections

EOSC will enable the 'intelligence' and processing power of machines to be utilised to uncover connections and related relevant material that may not be put together otherwise. Metadata is a central tenet of FAIR. All digital objects require persistent identifiers and rich contextual information to enable discovery and reuse. EOSC will provide a context where this metadata can be standardised in machine-readable formats so it can be processed at scale by computers, thereby alleviating some limitations of human searching and maximising the potential of machine searching.

Addressing global challenges

Societal and global challenges demand cross-disciplinary research, and thus datasets from different disciplines must be interoperable. By federating scientific data infrastructures and

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 133 –
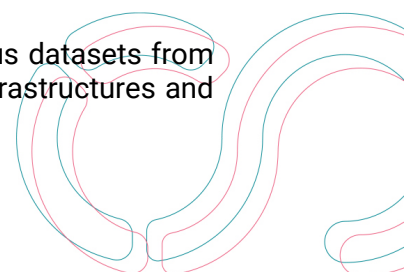
overcoming fragmentation, access and reuse of data will become easier and more efficient. EOSC will integrate the landscape of research data repositories in Europe, which is currently highly fragmented. By far the largest part of the relevant research data are not stored in repositories. For those data that are stored in local, institutional or disciplinary repositories, they form disconnected research data silos where data are largely unfindable, thus inaccessible and definitely not interoperable. This clearly hampers data reuse, knowledge circulation and, more importantly, it reduces significantly the impact science could have on society in the broadest sense.

> ### Example: Addressing the COVID-19 Pandemic
>
> When addressing global challenges, multiple streams of data from different fields are needed. COVID-19 is a case in point. To address the pandemic, epidemiological data to track the spread of the disease, understand patterns of transmission and support contact tracing were naturally at the fore. The various applicati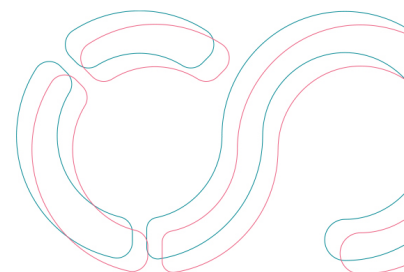ons released to gather these surveillance data raised many social and ethical questions about appropriate access and reuse, requiring strong governance controls and robust authentication and authorisation infrastructure (AAI). Person-level clinical data on patients, such as virology test results and imaging data such as lung scans, as well as sequence and metabolomics data were also needed. To implement effective policy measures, these medical data need to be combined with a much wider range of inputs such as real-time travel information, economic analyses and social insights into likely public responses to proposed measures.
>
> The European COVID-19 Data Platform, coordinated by the European Commission (EC) and European Molecular Biology Laboratory (EMBL), enables the rapid collection and comprehensive data sharing of available research data and tools on COVID-19 from different sources for European and global research communities. Practically, this enables researchers to upload, access and analyse COVID-19-related reference data and specialist datasets. A data portal provides the primary entry point into the functions of the data platform, which in turn forms an entry point into the future EOSC.
>
> Expanding the European COVID-19 Data Platform to enable the integration of molecular research data with patient and clinical data will ensure that patients benefit directly from the research supported by EOSC. Cross-linking with socio-economic, societal response and other social science and humanities will promote an integrated understanding of European outbreak response and preparedness and demonstrate the value of FAIR data to society and public engagement.

Through the federation of data and research infrastructures, EOSC will enable the creation of new opportunities and solutions in key thematic sectors such as health, food, transport or environment. EOSC will allow researchers from different countries and disciplines to verify, combine and build upon existing scientific data, addressing questions that cannot be addressed in isolation. In order for EOSC to achieve these goals, there is an onus on researchers to adopt relevant community standards and for the curation community to develop crosswalks for interoperability. Research communities need to be supported to define and adopt data standards, sharing agreements, services tools and know-how to facilitate the reuse of data. Some, such as astronomy, life sciences and linguistics, have self-organised, but many others require support to avoid widening the gap between the research communities active in EOSC or the range of content and resources that are available for multi-disciplinary reuse.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 134 –

Enhancing reproducibility

Reproducibility of research results is an essential aspect of research. It encourages objectivity and self-correction as well as discouraging scientific misconduct and fraud. However, it is widely recognised that today many research results are not reproducible. Opening up research processes and outputs is an important way to aid reproducibility. This is true not just for data, though this is critical, but also for all the processes and tools used in the research lifecycle, including software, methodologies, instruments, simulations, and analysis and workflows. EOSC will provide researchers with the means to access complete datasets and analysis platforms and provide services that support reproducibility, as well as ensuring long-term preservation and long-term availability of these research data and tools.

Reproducibility also requires a stable and trustworthy IT infrastructure, which is not, in general, provided by an individual researcher's desktop folders and analysis codes, often developed for one-off use. Where research is undertaken by large teams, this may exist already so that the team can work together. Where research is undertaken by an individual or small team, this is less often the case. EOSC will provide a sustained and stable infrastructure for research, with a multitude of readily available research datasets and tools, thereby encouraging researchers to develop their own research environment on this platform, encompassing reusing existing components, rather than building one-off, non-reusable tools in their own personal IT space.

7.2. Development of innovative services and products

Europe is undergoing a digital transformation in all sectors to foster innovation. In science, EOSC will lead to a fundamental revolution in the way researchers, companies and government agencies share and exploit research outputs, somewhat similar to how 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.

Within the scope of the Co-programmed European Partnership, EOSC will also address the differences in economic development in the research and innovation sector by creating more equitable access to data and services from both users and providers. Researchers and innovators will be able to jointly create innovative new technologies and services, which in turn will lead to the creation of new jobs and markets.

Opportunities to improve support for researchers

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.

By enabling access to data and services at European level, EOSC will facilitate and widen the opportunities for researchers to collaborate, and will enable them to start new research activities in their home country without relocating. EOSC will therefore further strengthen a balanced and fair 'brain circulation' and achieve a more symmetric mobility of researchers.

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 be managed in a holistic, synergistic way, as interrelated digital objects, in order to optimise the reuse of research results.

The EOSC Web of FAIR Data and Services 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.
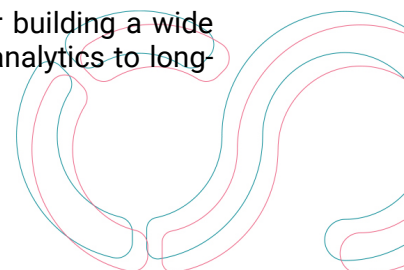
The consolidation of (FAIR) data commons and the interconnection of research data silos will also enable the creation of new opportunities and new solutions in key thematic sectors such as health, food, transport or environment.

To encourage the development of innovative services supporting FAIR principles, as well as data stewardship and preservation across different phases of the research lifecycle, dedicated incentives schemes funded by the EC are foreseen that would use the EOSC-Exchange as a distribution channel. For example:

- Research and Innovation action grants to develop services to be made available via the EOSC-Exchange;
- Pre-Commercial Procurement / Public Procurement of Innovation Solutions (PCP/PPI) co-funding financial instrument for innovative services to be co-developed with the private sector, procured jointly by public authorities and commercialised via the EOSC-Exchange.

All such innovation incentives would require developments to adhere to Rules of Participation resulting in production-quality services (Technology Readiness Levels 7–9) to be included in the EOSC-Exchange with associated training material.

Opportunities to improve support for the private and public sector

EOSC will enable the additional functionalities and services that it provides to serve not only the research community but also the public and the private sector so that they can exploit open data and associated services in such a manner that greatly increases the potential for innovation and economic impact in Europe. EOSC will bring more actors and investments into the research and innovation process.

EOSC will be instrumental in stimulating many areas of the European private sector, for example, the cloud and artificial intelligence (AI) industries, that are 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'[31].

---

[31] For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".
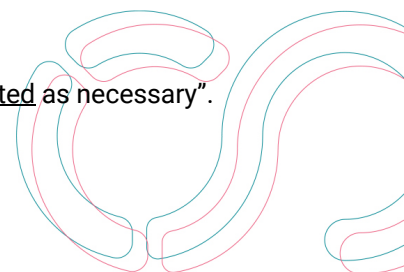
> **Example: The Copernicus Data and Information Access Services**
>
> One inspirational example is the Copernicus Data and Information Access Services (DIAS), which provide access, tools and processing capabilities for scientists and innovators to exploit Sentinel data. The five DIAS online platforms are operated by the industry and allow users to discover, manipulate, process and download Copernicus data and information. All DIAS platforms provide access to Copernicus Sentinel data, as well as to the information products from Copernicus' six operational services, together with cloud-based tools (open source and/or on a pay-per-use basis).
>
> Federating Copernicus data and DIAS added-value services into EOSC will leverage the existing EC investments for the benefit of multiple science and innovation communities. This will reduce the burden on scientific institutes to engage in complex procurement processes, support cross-analysis of data from heterogeneous sources, create market opportunities for research data services and represent a demand-side stimulus for the commercial DIAS.

Opportunities to improve European leadership and collaboration in a global setting

The EOSC Partnership will increase European leadership in Open Science and provide opportunities to strengthen international cooperation. EOSC has begun as a European initiative, federating research data repositories and infrastructures across Europe, but the ultimate goal of EOSC is to lead the development of a Global Open Research Commons, of which EOSC will form the European component.

EOSC will be European and open to the world, reaching out over time to relevant global research partners and initiatives so that by 2027 there can be alignment and interoperability of infrastructures to promote Open Science globally. Coordination fora including COAR, CODATA, RDA and WDS [COAR; CODATA; RDA; WDS] provide an environment where the different layers of interoperability (legal, organisational, semantic and technical)[32] can be discussed with partners from around the world. There is a clear willingness to collaborate and it is expected that the first agreements will be put in place during the first iteration of EOSC.

7.3. Improved impact of research in addressing societal challenges

Research in society

Through the introduction of EOSC, research will gain public awareness and will meet the public need to trust scientific facts. Against this background, empathy, transparency and the mediation of research ethics will have as big an impact on the public status of research as will data quality or quantitative ways of measuring impact, whilst both concepts will enhance societal resilience and meet socio-economic needs. EOSC will make possible a much higher level of interdisciplinarity and scientific evidence in decision making, planning and strategy at societal level.

EOSC will help Open Science to become the new normal. EOSC envisions a sustainable and federated infrastructure that offers standards, tools and services, allowing researchers to find, access, reuse, and combine scientific results, and in which these researchers are trained and rewarded for Open Science. This will improve the quality and productivity of science, with researchers being able to access and exploit other research as well as collaborate with other researchers, and will increase public trust in science as an open and evidence-based enterprise for society. This renewed trust in science is crucial given the rise of fake news and

---

[32] Layers of the interoperability model defined in the European Interoperability Framework [EC_Interoperability].

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 137 –
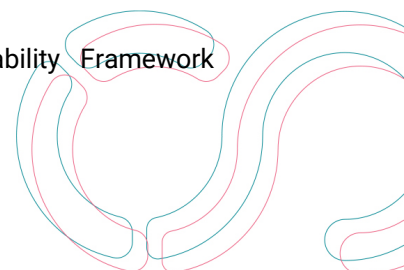
loss of trust in experts. This will also stimulate the development of innovative services and products arising from scientific breakthroughs, further stimulating scientific advancement and fuelling the economy by stimulating market competition, creating jobs and encouraging consumer spending. These objectives and benefits, in turn, improve the impact of research in addressing the global societal challenges of the times and give a return on the public investment in science.

Supporting international collaboration

As the COVID-19 pandemic has dramatically shown, immediate and open access to scientific research is crucial to deal with urgent societal challenges. EOSC will ensure that scientific publications, data and software relating to urgent societal problems are discoverable, accessible and reusable for other researchers to speed up breakthroughs, such as finding a solution to halting the spread of and ultimately vaccinating against COVID-19.
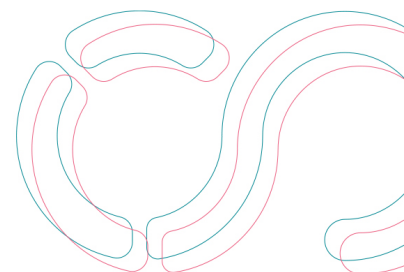
Better and faster sharing of research will naturally strengthen collaboration among researchers and disciplines as well as create opportunities for new levels of integration. The interoperability of data will also lead to unexpected links across disciplines as well as stimulate and support multi-disciplinary research. EOSC will, in effect, bring researchers within and across disciplines together and help science become more of a team enterprise. This is crucial for successfully tackling large-scale societal challenges, such as the Horizon Europe missions, which typically involve complex problems and require solutions from a multitude of different disciplines. One example is climate change, which is a truly multi-disciplinary research domain and can include botanists, climatologists, computational modellers, geochemists, mathematicians, meteorologists and oceanographers. While these researchers need to find one another and learn to work together, they also need the right tools to be able to collaborate effectively. EOSC will provide a catalogue of value-added services that will provide computation, storage and analysis as well as other data-related services and tools to help researchers collaborate in a multi-disciplinary environment.

Lifting science beyond the human scale

For science to really break boundaries, researchers need to think beyond what they currently know and make connections that they do not currently see. One barrier is the exponentially increasing amount of data being produced, which is already too much for a human to process. Another barrier is the lack of interoperability across datasets, resulting in a fragmented data landscape. A further barrier is that humans are not able to pinpoint statistical correlations across a diverse range of different disciplinary datasets in a reasonable amount of time.

EOSC will lift science to a new technological level and help researchers make discoveries that could never be made with conventional methods. The deployment of smart algorithms, machine learning and AI services onto the Web of FAIR Data will allow unexpected correlations to be made across all interconnected datasets in real time. It is then the researchers' task to investigate these new scientific avenues and determine causation from the correlations, and the innovators' task to convert this new knowledge into societal benefit. Imagine, for instance, running a search on 'malaria' in a research discovery portal that has access to the Web of FAIR Data. Within seconds, the search tool delivers a structured collection of results summarising all related articles and relevant data from both expected and unexpected sources (such as a climatological institute), industrial stakeholders (such as a pharmaceutical company), and public institutions (such as a hospital). And then, after a short interaction to understand the nature of the enquiry more precisely, the search tool suggests a specific treatment for specific patients in a specific region: an exciting potential discovery only made possible through an approach such as EOSC.

7.4. Critical success factors

The developments and expected impacts described above will not happen spontaneously. For these benefits to materialise a number of critical success factors (CSFs) must be in place. The following CSFs have been identified for EOSC:

- Researchers performing publicly funded research make relevant results available as openly as possible;
- Professional data stewards are available in research-performing organisations in Europe to help implement FAIR principles and support Open Science;
- Researchers are skilled and incentivised to perform Open Science;
- The scope of EOSC is widened to serve the public and private sectors;
- Research data produced by publicly funded research in Europe is FAIR by design;
- The EOSC Interoperability Framework supports a wide range of FAIR digital objects including data, software and other research artefacts;
- European research is increasingly discovered and reused across disciplines as a result of EOSC;
- EOSC is operational and provides a stable and trustworthy infrastructure, supporting researchers addressing societal challenges;
- EOSC has a sustainable funding and business model;
- EOSC is populated with a valuable corpus of interoperable data and services;
- The provision of data and services is available across borders;
- EOSC is a valuable and valued resource to a wide range of users from the research and education, public and private sectors.

To manage these prerequisites, some have been translated into activities, outcomes and key performance indicators in the EOSC roadmap; others will be the focus of a sustained stakeholder engagement and communications strategy.

The next section outlines the roadmap for delivering a fit-for-purpose EOSC that will in turn deliver the expected impacts described above.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 139 –
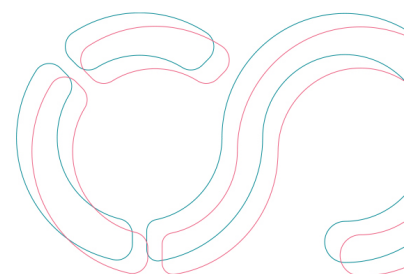
8 Roadmap

This Strategic Research and Innovation Agenda (SRIA) sets clear goals to develop the European Open Science Cloud (EOSC) and build a research environment that supports Open Science and increases trust in and reproducibility of research. The overall impact is a pan-European research landscape that will offer significantly improved discovery, access, interoperability, and exploitation of research outputs for researchers and for research and innovation stakeholders.

The funding periods for the Horizon Europe (HE) work programmes and the SRIA implementation stages (see below) both run from 2021 until and including 2027. Each stage defines a high-level objective for the period, aligned with the vision of enabling a trusted, virtual, federated environment in Europe to store, share and reuse research outputs across borders and scientific disciplines. For each stage, priority areas are defined, their prioritisation and scheduling reflecting community feedback from the open consultation process.

Stage 1 (2021−2022): Development towards added value from a functional federation of infrastructures
Enabling the European Open Science Cloud operations (the EOSC-Core) to provide necessary core functions of the Minimum Viable EOSC (MVE) that allow federation of existing and future infrastructures, according to associated Rules of Participation and governance that provision growth and expansion in the following stages.

Stage 2 (2023−2024): Expansion to production that generates added value
Expanding and building the core data infrastructure to support the full lifecycle of scientific research in key thematic areas. During this period, activities will build on pilots / demonstrators and work towards linking EOSC beyond the research communities to the wider public sector and the private sector.

Stage 3 (2025−2027 and beyond): Expansion to develop impact from Open Science
Deployment of federated research infrastructures with functionality that allows multiple European research communities to deliver impactful Open Science. In addition to European infrastructures, the national research infrastructures delivered from the Member States and Associated Countries in particular will help in this expansion phase.

The Multi-Annual Roadmap (MAR) 2025-2027 builds on the earlier roadmaps (MAR 2021-2022 and MAR 2023-2024) and the work programmes of the EC following from those. The MAR 2025 was detailed in SRIA v1.2; information about it can be found in Appendix B of the present document. In the following we present the process to develop the MAR for the period 2026-2027 and its outcomes. This document (SRIA 1.3) contains a revised and more up-to-date list of priorities for the years 2026-2027 than was included in SRIA 1.2, reflecting recent and current progress in the development of EOSC. Structuring the MAR into priorities for 2025 and for 2026-2027 aligns with the planning of the Horizon Europe programme, which will publish a Work Programme for 2025 followed by a Work Programme for 2026-2027.

Please note: The SRIA 1.3 builds further on SRIA 1.2, which builds on SRIA 1.1, which in turn builds on SRIA 1.0. In SRIA 1.3 (the present document), the priorities for the earlier periods, Stage 1 (2021−2022) and Stage 2 (2023−2024) are described in Appendix B and the priorities for 2025 are described in Appendix C. The remainder of this chapter focusses on 2026-27. First below a summary is given of the priorities for 2026-27.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 140 –

8.1. Establishing the MAR 2026-2027

From autumn 2022 to spring 2023, the MAR 2025-27 was produced, developing a set of proposed actions for all three years, but with a focus on the envisaged priorities for 2025.

In the autumn of 2024, the EOSC Association Board began the process of updating the MAR 2025-2027 by drafting new proposed priorities for 2026-2027, and structuring them into four new Strategic Pillars rather than under the three overarching objectives of EOSC as was the case for previous versions of the MAR. The Strategic Pillars correspond to developments in the EOSC landscape and allow a clearer presentation of the proposed actions. The Strategic Pillars are:

Strategic Pillar 1: Sustaining and enhancing the EOSC Federation
Strategic Pillar 2: Contributing to the web of FAIR data and the uptake of AI
Strategic Pillar 3: Ensuring research security and sovereignty
Strategic Pillar 4: Linking with other Common European Data Spaces and beyond.

Each Strategic Pillar structured according to the three known levels: European, national and institutional, as was the case for earlier versions of the MAR. The overall number of actions proposed in the updated MAR 2026-2027 is smaller than in previous versions, reflecting the emerging clearer focus and definition of EOSC.

This was used as the basis for a public consultation which ran from 3-24 September 2024 and the proposed set of priorities was updated as a result. In October 2024, this was shared with the European Commission services, who provided their input on the European-level priorities.

The EOSC Association Board thanks all those who contributed to the process of updating the MAR for 2026-2027, including its Task Forces, the EOSC Focus consortium[33] the EOSC Association secretariat and the community at large.

8.2. Results of the MAR 2026-2027 Consultation

Responses:

Mandated Member / Organisation:	12	Research Performing:	25
Member:	23	Research Funding:	3
Observer:	6	Service Providing:	19
Not a Member/Observer:	10	Other	4
Number of responses:	**51**		
Member or Observer EOSC-A		Not a Member or Observer EOSC-A	
On behalf of my organisation:	22	On behalf of my organisation:	4
On behalf of an EOSC-A Task Force:	2	On behalf of an Opportunity Area Expert Group:	1
On behalf of an EOSC-related project under Horizon Europe:	3	On behalf of an EOSC-related project under Horizon Europe:	2
In my own name:	14	In my own name:	3
Number of responses:	**41**	Number of responses:	**10**

---

[33] [EOSC Focus](#) a HORIZON Europe INFRAEOSC project
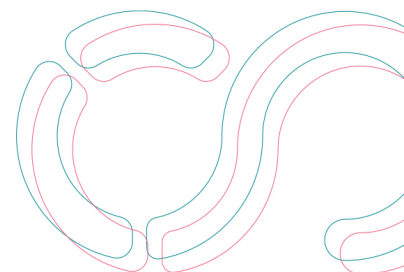
The community consultation received 51 responses, with 41 coming from EOSC-A members, representing roughly 20% of the total. Around 66% of the responses were submitted on behalf of organizations (51%), the EOSC Task Force (3.9%), HE EOSC-related project (9.8%), and EOSC Opportunity Area Expert Group (2%), while 33,3% came from individual participants.

As expected, most of the responses came from Research Performing Organisations (RPOs) and service providers (see table above).

The consultation included several structured questions to gauge the level of agreement with the Strategic Pillars regarding developments in the EOSC landscape. A sixth option, 'No opinion,' was added to the matrix (highly irrelevant - irrelevant - neutral - relevant - highly relevant) to differentiate a neutral position from having no opinion. Respondents also had the opportunity to provide specific comments and suggestions for changes. Of the 32 activities analysed from 51 respondents, an average of 3.7% indicated not having an opinion, a low percentage.

Overall, there was a very high-level of agreement with the priorities suggested for each of the Strategic Pillars (Figure 8.1see Figure 8.1, Figure 8.2, Figure 8.3).
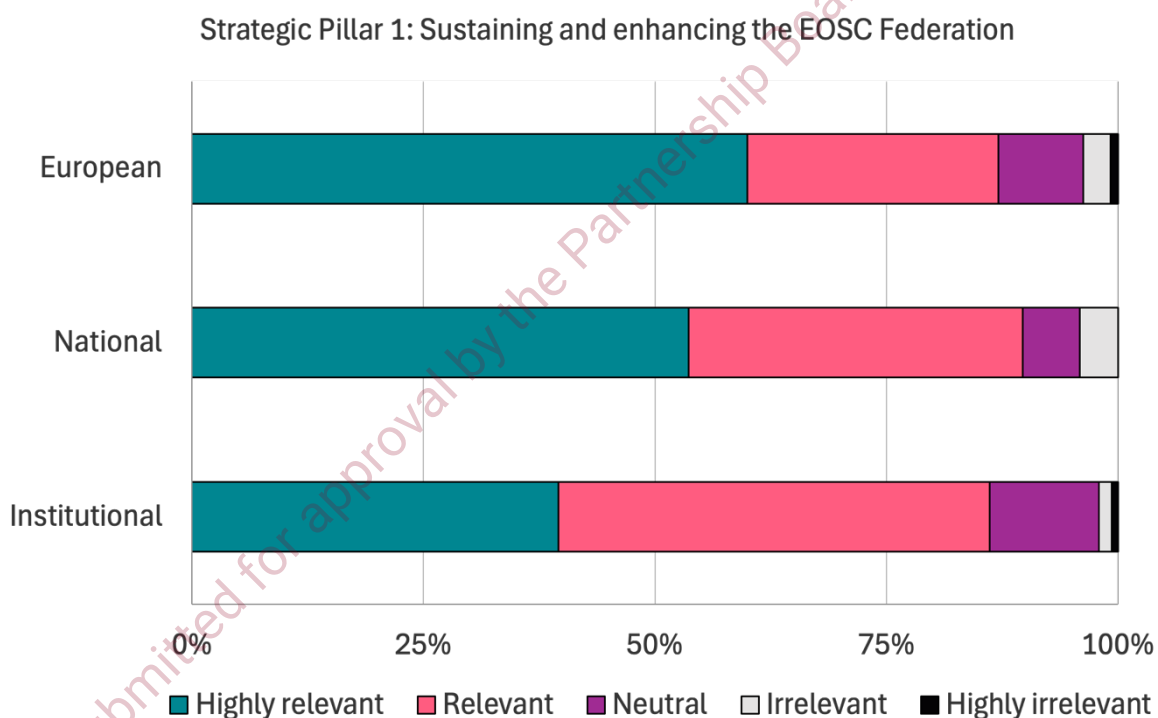
Figure 8.1: Level of agreement with the suggested priorities under Strategic Pillar 1: Sustaining and enhancing the EOSC Federation

Strategic Pillar 2: Contributing to the web of FAIR data and the uptake of AI

Highly relevant | Relevant | Neutral | Irrelevant | Highly irrelevant
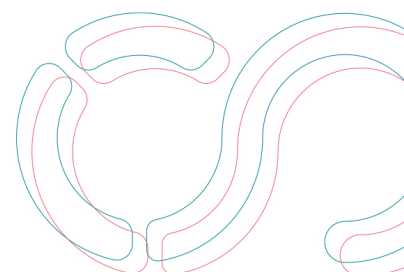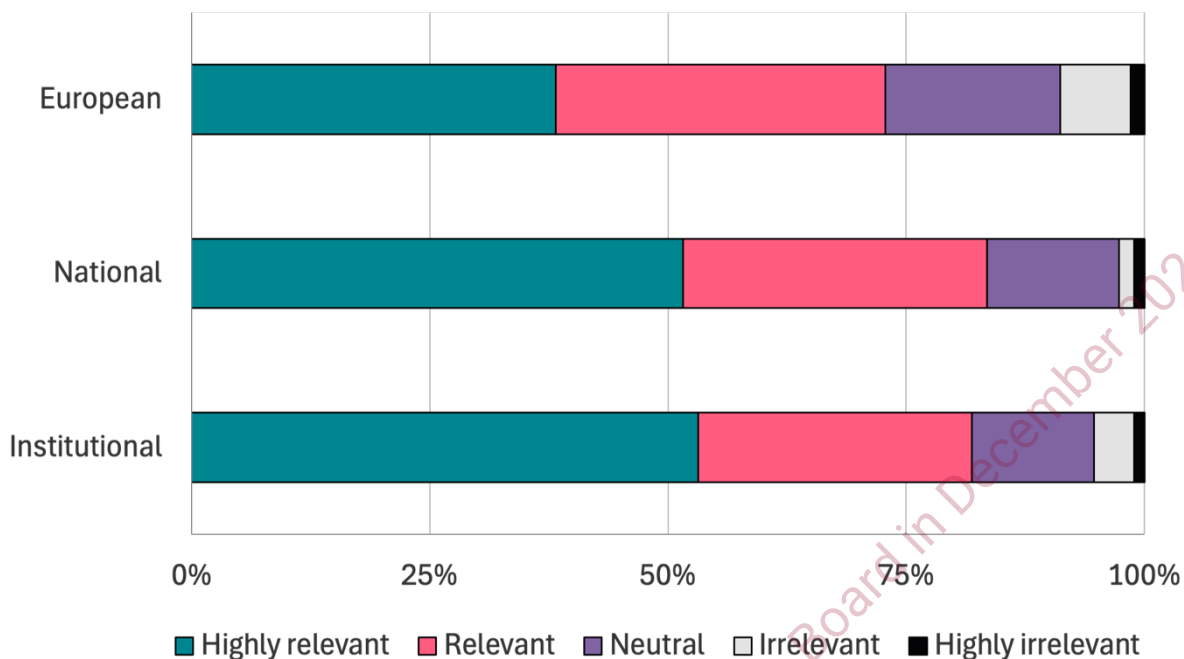
Figure 8.2: Level of agreement with the suggested priorities under Strategic Pillar 2: Contributing to the web of FAIR data and the uptake of AI

Strategic Pillar3: Ensuring research security and sovereignty

Highly relevant | Relevant | Neutral | Irrelevant | Highly irrelevant
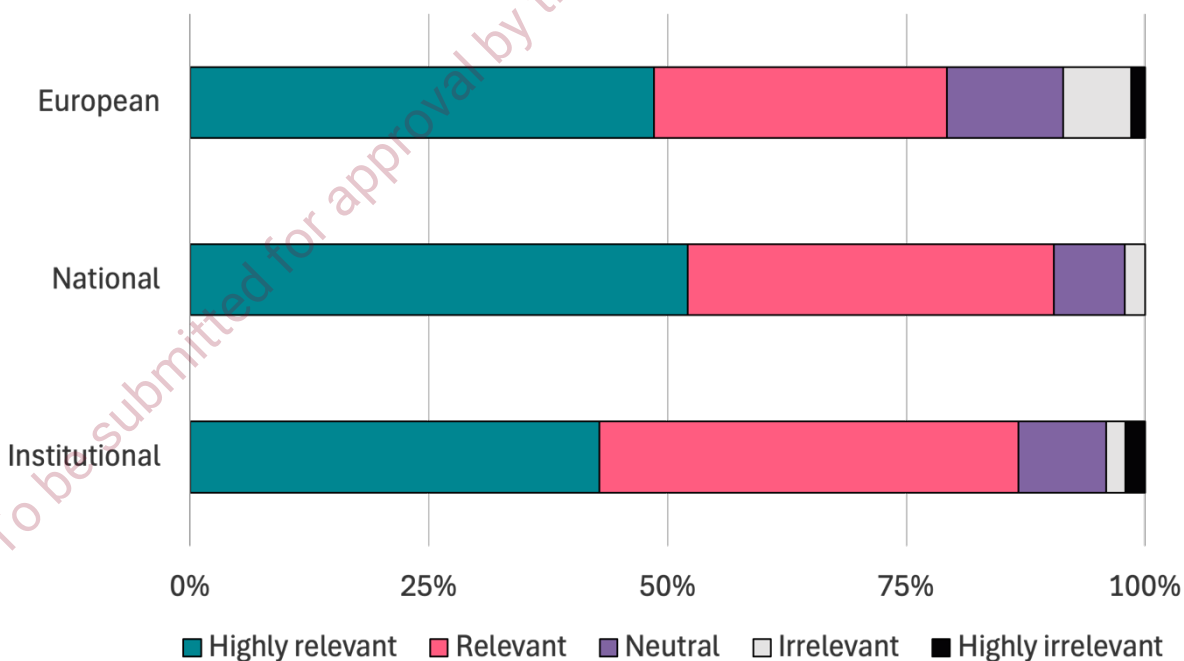
Figure 8.3: Level of agreement with the suggested priorities under Strategic Pillar 3: Ensuring research security and sovereignty
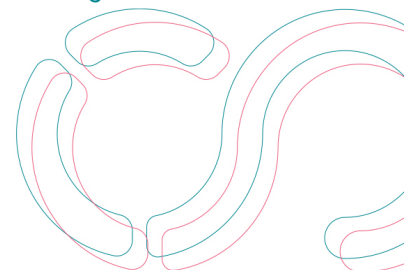
Strategic Pillar 4: Linking with other Common European data spaces and beyond
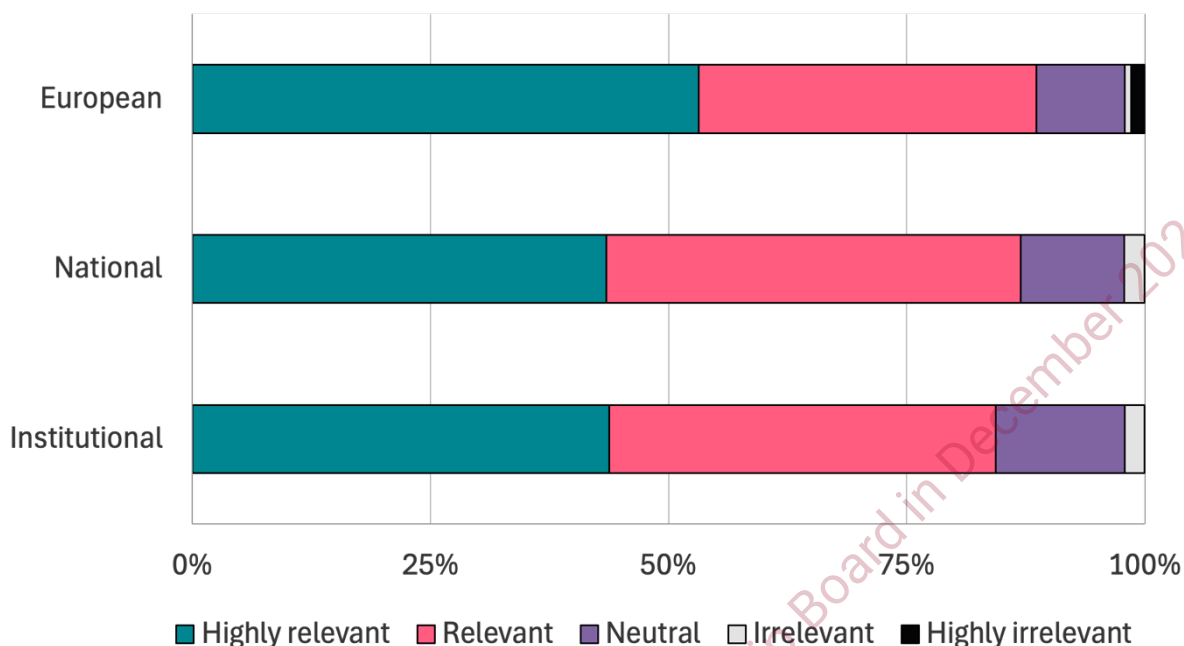
Figure 8.4: Level of agreement with the suggested priorities under Strategic Pillar 4: Linking with other Common European data spaces and beyond

The 2024 survey received 51 responses in total, contributing more than 200 comments. These were analysed by a team from EOSC Focus, divided into pairs dedicated to analysis of one Strategic Pillar each. All comments, including those submitted in the general "additional comments" section at the end of the survey, were analysed and assessed for the appropriate action to take to address them. The prior structuring of the MAR into four Strategic Pillars significantly reduced the amount of work involved in analysing the survey results compared to the previous consultation exercise.

Most of the comments related to specific MAR actions and processing them resulted in numerous suggested edits or additions to the proposed MAR actions. In addition, there was a small number of more general comments, which were processed by EOSC Association Board members. The updated set of actions for each Strategic Pillar was reviewed by a member of the EOSC Association Board. The complete set of updated actions was then subject to a final review by the Association Board, to check for overall coherence and consistency, resulting in the updated MAR 2026-2027 presented below.

8.3. MAR 2026-2027

Strategic Pillar 1: Sustaining and enhancing the EOSC Federation

European level:

A. Further develop the EOSC Federation model: a diverse set of national, regional, thematic and other EOSC Nodes integrating and sharing resources (Objective 3).
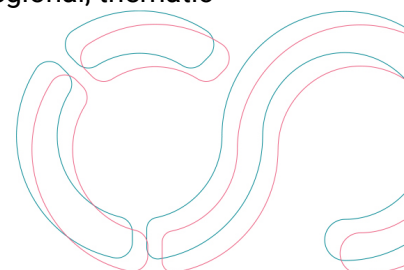
B. Improve and upgrade service provision based on users' needs and preferences to ensure a strong uptake by researchers (Objective 3).

C. Develop and adopt funding and governance models that will ensure sustainable long-term co-investment at European and national levels, including the additional costs for delivering resources beyond a Node's normal geographical or thematic user base (Objective 3).

D. Support concertation actions directed towards greater and more in-depth stakeholder engagement, coordination, governance and monitoring (Objectives 1 and 3).

E. Continue to stimulate the development and maintenance of open interfaces, alignments, guidelines, crosswalks and APIs that enable interoperability (Objectives 2 and 3).

National level:

F. National policy makers should support Research Infrastructures, e-Infrastructures and national service providers to align activities and establish EOSC support structures (Objectives 1 and 3).

G. Support and incentivize the use, maintenance and adoption of open standards and APIs to enable resource composability and to achieve interoperability across communities (Objectives 2 and 3).

Institutional level:

H. Support professional development programmes to ensure research support staff have the required data stewardship and information security management skills (Objective 1).

I. Research institutions should encourage and support researchers to adopt the data and services federated by EOSC to scale up usage (Objectives 1 and 3).

J. Encourage sharing of quality-assured software through institutional or thematic repositories (Objective 3).

Strategic Pillar 2: Contributing to the web of FAIR data and the uptake of AI

European level:

A. Support the establishment of a lightweight, widely representative governance body to incentivise more FAIR artefacts and elaborate more coherent FAIR assessment (Objectives 2 and 3).

B. Accelerate the adoption of interoperability and semantic artefact catalogues. Support AI-enabled services, engaging discipline-specific groups, that facilitate interoperability and reuse in a more automated way by automatic annotation, data linkage, data homogenisation, and data transformation, or other similar approaches, including global alignment (Objective 2).

C. Accelerate the uptake of AI in science, while addressing relevant challenges and risks. Support the development of AI-ready FAIR research data as well as tools and services that enable the development of scientific AI models. (Objective 2).

D. Support the development of standardised guidelines (e.g. ethical, legal), to describe the potential uses of Open Data in AI research to ensure transparency and fairness in AI model development across Europe (Objective 2).

National level:
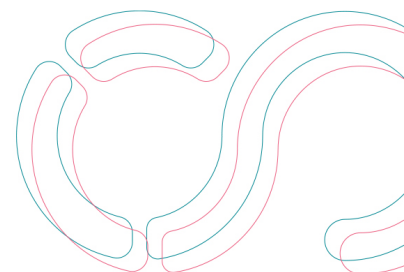
E. Use European and domain-specific semantic artefact catalogues in national infrastructures and guidelines, aligned with European standards and vocabularies (Objective 2).
F. Support adoption of both general and domain-specific standards to increase adoption of FAIR practices and develop plans to facilitate reuse (Objective 2).
G. Support the implementation of aligned European curricula for data stewardship and encourage their inclusion in research programmes, with a practical focus involving service providers (Objectives 1 and 2).
H. Establish protocols for dealing with the cost of data management, data stewardship, maintenance and preservation of research outputs (including software and semantic artefacts) and making them eligible within national funding schemes (Objective 1).
I. Promote collaborations between national infrastructures and cross-border initiatives to promote consistent FAIR data standards and enhanced international interoperability (Objective 1).
J. Further co-develop and evaluate research review mechanisms to ensure FAIR research outputs, use of PIDs and other Open Science practices are appropriately recognised and rewarded (Objective 1).

Institutional level:

K. Adjust, follow up, and evaluate research review mechanisms to ensure FAIR research outputs and Open Science are appropriately recognised and rewarded (Objective 1).
L. Integrate widely used and adopted PIDs into institutional services and incentivise usage of PID technologies being developed for EOSC (Objectives 1 and 2).
M. Encourage institutions to develop training modules on data stewardship, FAIR principles and assessment tools, PID usage and other Open Science practices to ensure long-term adoption and better integration with EOSC services (Objective 1).

Strategic Pillar 3: Ensuring research security and sovereignty

European level:

A. Further co-design, develop and deliver secure AAI solutions for the EOSC Federation, building on existing AAI and standards and drawing on the community's accumulated experience with developing and managing federations (Objectives 2 and 3).
B. Clarify the EOSC Rules of Participation and then develop machine-actionable means to monitor compliance with them (Objective 3).
C. Define a harmonised operational (including cybersecurity aspects) and legal framework to facilitate the secure sharing and governance of, and access to, data (including sensitive data) and services. (Objectives 2 and 3).
D. Encourage the development and coordination of national data sovereignty frameworks that align with European and international research initiatives (Objective 2).

National level:

E. Encourage national policy makers to review and adjust national policies, funding and regulations, and to develop training programs that enable services and data to be used in cross-border and cross-domain contexts, preserving data sovereignty and ensuring broader interoperability and accessibility (Objectives 2 and 3).
F. Support and embed appraisal, retention and preservation procedures for long-term data and other digital assets (e.g. software) and monitor the value and costs of the procedures (Objective 1).

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 146 –
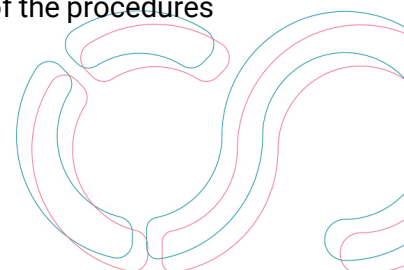
Institutional level:

G. Offer accredited and certified core Open Science, FAIR and CARE training to researchers and research support units at all levels and recognise the skills as part of the individual's professional development through certified means (Objectives 1 and 2).
H. Define and implement training and procedures to select and curate data, software and other research outputs that merit preservation (Objective 2).
I. Define and implement training and procedures to develop information security management skills especially relating to managing sensitive data in line with national and, where appropriate, European procedures (Objective 1).
J. Encourage the creation of institutional strategies and training programmes for data risk management that are aligned with European standards for research security (Objectives 1 and 2).

Strategic Pillar 4: Linking with other Common European Data Spaces and beyond

European level:

A. Engage and bridge with other Common European Data Spaces and relevant initiatives, including the EU Missions, EuroHPC and other relevant European Partnerships (Objectives 1 and 3).
B. Support and incentivize the use, maintenance and adoption of open standards and APIs to enable resource composability and to increase the interoperability between the research and other communities, including in the public administration and the private sector (Objective 2).
C. Encourage more focused collaboration with private sector data spaces to foster innovation while ensuring that data sovereignty, ethical standards and data privacy regulations are upheld (Objective 2).
D. Promote fora that will enable the EOSC Federation to contribute to the establishment of a global data commons (Objectives 1 and 3).

National level:

E. Leverage existing national Competence Centres and strengthen their participation in coordination networks at the European level (Objectives 1 and 3).
F. Support the integration of national-level Research Infrastructures into sectoral Common European Data Spaces. (Objectives 1 and 3).

Institutional level:

G. Ensure that researchers are aware of and can reference existing and domain-specific semantic artefact catalogues and other data services, including repositories and registries developed and used in EU initiatives (Objective 2).
H. Promote participation in existing initiatives, like Data Spaces, EU missions and European Partnerships at national and European level (Objective 1).

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 147 –
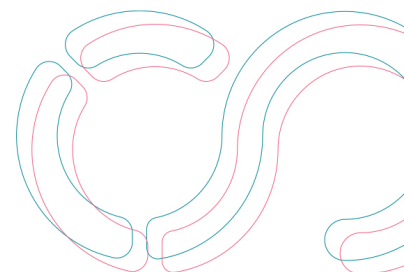
8.4. Objectives

The Co-programmed European Partnership on EOSC, as agreed between the EC and the EOSC Association, is to be implemented in an open, transparent, efficient, and flexible way. The EOSC Association demonstrates this openness and transparency via the participatory approach in which we consult with stakeholders when developing key documents like the SRIA, and our inclusiveness when defining membership for structures like our Task Forces. Moreover, the Memorandum of Understanding for the Co-programmed Partnership[34] and other key documents are publicly available. The intended cooperative relationship aims to achieve jointly defined objectives based on a long-term common vision and a clear commitment from the partners throughout the duration of the Partnership. More on the EOSC Partnership can be found in Section 2.7.1 of the SRIA.

As we have advanced almost three years since SRIA v1.0 was written, each of the objectives in SRIA v1.0 was carefully reviewed. As a result, we added an interpretation to some of the objectives, for the purpose of operationalising them. These interpretations, based on comments from EOSC-A stakeholders, are visible in the three tables below. The most significant item is the interpretation of Specific Objective (SO) 05. This relates to a new priority area of work, namely data quality, which was not reflected in the roadmap previously.

Some of the objectives in the tables below have a timeframe associated to them, while others have not. As the delivery is tracked via KPIs, the timeframes mentioned in the tables are seen more as the original idea and less as a strict deadline.

The **general objectives** (GOs) of the European Partnership, which are identical to the strategic objectives described in Section 3 of this SRIA, including their interpretation, are defined as follows (Table 8.1):

Objective#	SRIA v1.0 & MoU	Interpretation based on advancing insights
GO1	Ensure that Open Science practices and skills are rewarded and taught, becoming the 'new normal'	
GO2	Enable the definition of standards, and the development of tools and services, to allow researchers to find, access, reuse and combine results	We aim for not only defining these standards but also to have them adopted
GO3	Establish a sustainable and federated infrastructure enabling open sharing of scientific results	

Table 8.1: General Objectives (from SRIA v1.0 and interpretation)

The **specific objectives** (SOs), which are reflected in the critical success factors identified in Section 7.4 of this SRIA, including their interpretation, are the following (Table 8.2):

Objective#	SRIA v1.0 & MoU	Interpretation based on advancing insights
SO1	Increase in the number of relevant research results that are made available as open as possible by researchers performing publicly funded research	For the purpose of this objective "research results" include also all types of "research outputs"[35]
SO2	Professional data stewards are increasingly available in research performing organisations in Europe to support Open Science	
SO3	Development and adoption of incentives for researchers to perform Open Science	
SO4	Increasing amounts of research data produced by publicly funded research in Europe are FAIR by design	We aim to not only make research data from publicly funded research in Europe FAIR by design but we aim for this goal for all research outputs
SO5	The EOSC Interoperability Framework supports an increasing range and quantity of FAIR digital objects including data, software and other research artefacts	Common data quality indicators are agreed and implemented to ensure that research outputs within EOSC are ready for FAIR usage
SO6	Provide an increased number of services and resources to ensure that European research is discovered and reused within and across disciplines to extract new knowledge	
SO7	EOSC is operationalised and provides a stable and valuable infrastructure supporting researchers addressing societal challenges	
SO8	Essential additional functionalities for end users from the public and private sectors are implemented in EOSC (these developments are complementary to those of other European data spaces)	
SO9	EOSC increasingly establishes ties with related initiatives from regions around the world and becomes a partner in global cooperation frameworks for Open Science	

---

[35] The term "research output" is defined in the HE Framework programme and means the results generated by a given action to which access can be given in the form of scientific publications, data or other engineered results and processes such as software, algorithms, protocols and electronic notebooks.

Table 8.2: Specific Objectives (from SRIA v1.0 and interpretation)

The **operational objectives** (OOs), which are reflected in the action areas described in Sections 5 and 6 of this SRIA, including their interpretation, are the following (Table 8.3):

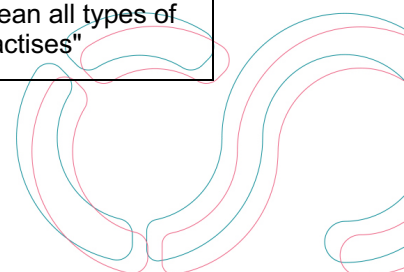| Objective#	SRIA v1.0 & MoU	Interpretation based on advancing insights
OO1	Deliver and operate all the necessary components of the Minimum Viable EOSC to share openly research data, publications, software, tools and services while attracting increasing numbers and categories of users (public and private) (based on a governance structure representative of the various stakeholders and including domain-specific user environments supporting Open Science) by 2025	
OO2	Make monitoring systems to gather data and evidence on best Open Science practices accessible through EOSC (including the development of a dashboard to monitor the evolving landscape of policies, infrastructures and open resources made accessible via EOSC by 2023)	This could utilise the EOSC Observatory dashboard to monitor the evolving landscape of policies and practices for EOSC and open science in Europe
OO3	Increasingly mainstream Open Science skills in European research-performing organisations (RPOs) including through the uptake of curricula and training frameworks related to data stewardship through the lifespan of the Partnership	Next to the uptake of curricula we strive to align them. This will be performed for the full lifecycle of data management
OO4	Co-develop domain-specific standards and adopt Open Science practices through the engagement with research communities during the lifespan of the Partnership	
OO5	Provide the technical components of a FAIR ecosystem for uptake and customisation by the communities by 2023 (including open specifications, standards, schemas, application programming interfaces (APIs), metadata frameworks supporting FAIR digital objects and their automated processing)	
OO6	Provide the metrics and tools to measure the adoption of the FAIR principles for research artefacts and provide frameworks to help in certifying that repository services enable FAIR in EOSC throughout the lifespan of the Partnership	For the avoidance of doubt "research artefacts" mean all types of "research outputs"
OO7	Co-develop a first generation of a robust pan-European network of infrastructures for software source code (including incentives for the effective documentation and sharing of research software) by 2025	
OO8	Co-design and adopt a Rewards and Recognition framework for FAIR and open data practices in research during the lifespan of the Partnership	For the avoidance of doubt "open data practices" mean all types of "open science practises"

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu   - 150 –
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Objective#	SRIA v1.0 & MoU	Interpretation based on advancing insights
OO9	Implement and evolve the EOSC Rules of Participation and onboarding process for EOSC providers and increase the number of service providers and services offered progressively over the course of the Partnership	
OO10	Deploy and operate an authentication and authorisation infrastructure (AAI) framework to manage user identity and access by 2024	
OO11	Implement the EOSC persistent identifier (PID) policy and architecture by 2025	This includes the development of a global PID resolver
OO12	Co-develop a minimum metadata framework and provide a common search and access mechanism to EOSC resources across the EOSC federation by 2025	The common search mechanism to EOSC resources will allow the user to find relevant resources, while access to them will be organized by the provider of the resources
OO13	Continuously monitor and promote the increased uptake of core services and EOSC resources, access to EOSC Exchange tools and services and ensure a feedback loop with the users	
OO14	Define models for availability and costing of services across borders by 2023	

Table 8.3: Operational Objectives (from SRIA v1.0 and interpretation)

8.5. Levels of implementation

The activities proposed for the different stages, as outlined in the sections below, should be carried out in concertation by the parties and levels involved. Indications of the levels at which to carry out these activities is given by three levels of implementation:

- **European level (L1)**: everything done at the European level either by or financed through the Horizon Europe programme or other sources, be it an effort by the EOSC Association or by a research infrastructure or service organisation operating at a European level. This could also be work done through the joint effort of countries, etc. In other words, every contribution to EOSC being delivered at the European level (irrespective of who pays).
- **National level (L2)**: the same as above but then at the country level, i.e., activities in one of the Member States or countries associated to the Horizon Europe programme (MS/AC), as long as the activities contribute positively to the development of the EOSC ecosystem as described in the SRIA.
- **Institutional level (L3)**: again, the same as above but then the activities at the level of the institutions within a country (e.g. the universities and the other RPOs), and again as long as the activities are aligned with the country strategy which should in turn be aligned with the European strategy as described in the SRIA.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 151 –

In the text describing the priorities in the sections below, the levels European, National and Institutional are used to indicate at which level the main (driving) force for implementation can be found.

8.6. Appraisal of the present and outlook to the future

The development of the MAR 2025-2027 coincides with a crucial juncture for the EOSC Partnership, marked by the delivery of the EOSC Future project and the successful launch of the EOSC EU Node delivered through the procurement done by the EC.

Simultaneously, discussions regarding EOSC beyond 2027 are gaining momentum within the tripartite collaboration, aiming to converge on a unified governance and financing model.

This juncture presents a unique opportunity, as the work programme for the final years of Horizon Europe will be divided into two distinct phases: a standalone programme for 2025 and another encompassing 2026−2027. This strategic division aligns nicely with addressing all pertinent topics during the concluding two years of Horizon Europe, ensuring a seamless transition into the period of Framework Programme 10 (FP 10).

Nonetheless, this endeavour also comes with its share of challenges that necessitate attention. One such challenge relates to coordination. The HE INFRAEOSC projects seek direction and guidance, particularly in the technical and policy realms. Full connection between their work and the discussion in the EOSC Association Task Force but also with the EOSC Tripartite reflections must be ensured. Neither EOSC-A nor the EOSC Focus project (funded by the European Commission to help addressing this coordination challenge) can fulfil all coordination needs due to limitations in information and mandate. This is partly inherent to the approach of implementing EOSC through a series of calls for proposals resulting in independent grants in a context of a co-programmed partnership (voluntary collaboration with low level of scientific, managerial and financing integration). The EOSC Association, the European Commission and its executive agencies are collaborating to address this coordination challenge at best including through projects' coordination events, EU investment in supporting activities (e.g. EOSC Focus), and a series of concrete tools and services gathered by the EOSC Association in the EOSC Handbook for Effective Collaboration ([Vademecum](#)).

In order to capitalize on the potential for significant exploitable outcomes from the HE INFRAEOSC projects, it will be imperative in the future to ensure a structured exchange of information and a symbiotic relationship between EOSC R&I activities and EOSC operations.

8.7. Building on the Horizon 2020 and Horizon Europe projects

A key contribution to the Partnership is the EU funding of Horizon Europe (HE) grants (managed by the EC and its executive agencies) supporting the development of EOSC. These grants result mainly from INFRAEOSC Calls under the Horizon Europe Work programmes on Research Infrastructures.

In turn, the EOSC-A and its members provide another key contribution to the Partnership in the form of additional in-kind activities advancing the EOSC SRIA.

Adequate complementarity and coordination of these contributions, and between EOSC projects funded by the Union and the EOSC-A is crucial for a successful implementation of EOSC through the Partnership.

The Horizon Europe Calls in 2021 and 2022 have resulted in nineteen EU-funded project consortia grants of direct relevance to the EOSC Partnership (fifteen funded under the INFRAEOSC destination and others under the WIDERA, INFRA-DEV and EMERGENCY calls

respectively). The goal of creating interaction between these projects, is to provide them with a clear overview of the Partnership context, to update them on relevant policy developments and to foster a spirit of cooperation between the various consortia. This also allows to illustrate the breadth of project's' contributions, to identify common interests and new opportunities for collaboration in the context of the Partnership and to jointly discuss the main challenges for the transition of EOSC from a development to an operational phase.

A series of EC funded projects are currently the main contributors to the implementation of EOSC, while also several national initiatives from Member States and Association Countries are expected to be offering resources to the EOSC ecosystem in due time, as reported under the MoU of the EOSC partnership.
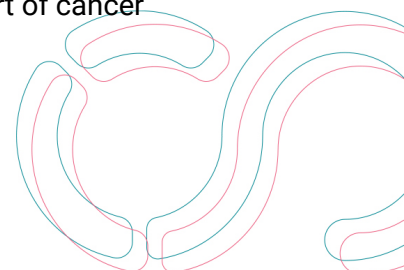
The wide range and diversity of projects illustrates the challenge faced by the EOSC Association in terms of coordination, as it is hard to build a coherent infrastructure from the results of a series of projects. While the EOSC-related projects under Horizon Europe do have an obligation to share results, the EOSC Association will monitor and where necessary request the sharing of results. A Handbook for Effective Collaboration or Vademecum has been drafted. This is intended to stimulate collaboration and coordination between the projects. Through this the EOSC Association tries to improve the concertation going forward. The Partnership aims to avoid overlapping activities or divergent approaches, and instead stimulates joint work towards an orchestrated result.

The INFRAEOSC calls that are of particular importance since they support the major investment in building EOSC and fostering the development of FAIR and Open Science practices:

- INFRAEOSC-01-2018: Access to commercial services through the EOSC hub
- INFRAEOSC-02-2019: Prototyping new innovative services
- INFRAEOSC-03-2020: Integration and consolidation of the existing pan-European access mechanism to public research infrastructures and commercial services through the EOSC Portal
- INFRAEOSC-04-2018: Connecting ESFRI infrastructures through Cluster projects
- INFRAEOSC-05-2018-2019: Support to the EOSC Governance
- INFRAEOSC-06-2019-2020: Enhancing the EOSC portal and connecting thematic clouds
- INFRAEOSC-07-2020: Increasing the service offer of the EOSC Portal

8.7.1. Calls Enabling an operational, open and FAIR EOSC ecosystem (2021)

- HORIZON-INFRA-2021-EOSC-01-01: Supporting an EOSC-ready digitally skilled workforce
- HORIZON-INFRA-2021-EOSC-01-02: Supporting the development and coordination of activities of the EOSC Partnership
- Supporting EOSC-Core: Enabling access to the Web of FAIR data and services
- HORIZON-INFRA-2021-EOSC-01-03: Deploying EOSC-Core components for FAIR
- HORIZON-INFRA-2021-EOSC-01-04: Innovative and customizable services for EOSC
- Building with the scientific community a Web of FAIR data for open science
- HORIZON-INFRA-2021-EOSC-01-05: Enabling discovery and interoperability of federated research objects across scientific communities
- HORIZON-INFRA-2021-EOSC-01-06: FAIR and open data sharing in support of European preparedness for infectious diseases
- HORIZON-INFRA-2021-EOSC-01-07: FAIR and open data sharing in support of cancer research

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 153 –

8.7.2. Calls Enabling an operational, open and FAIR EOSC ecosystem (2022)

- HORIZON-INFRA-2022-EOSC-01-01: Services and tools to underpin a research assessment system that incentivises open science practices
- HORIZON-INFRA-2022-EOSC-01-02: Improving and coordinating technical infrastructure for institutional open access publishing across Europe
- Building with the scientific community a Web of FAIR data for open science
- HORIZON-INFRA-2022-EOSC-01-03: FAIR and open data sharing in support of healthy oceans, seas, coastal and inland waters
- HORIZON-INFRA-2022-EOSC-01-04: Support for initiatives helping to generate global standards, specifications and recommendations for open sharing of FAIR research data, publications and software
- Public procurement: Delivering the EOSC core infrastructure and services

8.7.3. Calls Enabling an operational, open and FAIR EOSC ecosystem (2023)

HORIZON-INFRA-2023-EOSC-01-01: Build on the science cluster approach to ensure the uptake of EOSC by research infrastructures and research communities

HORIZON-INFRA-2023-EOSC-01-02: Development of community-based approaches for ensuring and improving the quality of scientific software and code

HORIZON-INFRA-2023-EOSC-01-03: Planning, tracking, and assessing scientific knowledge production

HORIZON-INFRA-2023-EOSC-01-04: Next generation services for operational and sustainable EOSC Core Infrastructure

HORIZON-INFRA-2023-EOSC-01-05: EOSC Architecture and Interoperability Framework

HORIZON-INFRA-2023-EOSC-01-06: Trusted environments for sensitive data management in EOSC

> The INFRAEOSC calls from 2023 resulted in seven new projects being funded for 36 months since January 2023. In addition, two projects, HORIZON-ZEN and EOSC-Track, were approved in 2023 focussing on the development of a FAIR enabling repository for projects funded by the European Commission as well as on monitoring EOSC, respectively.

8.7.4. Calls Enabling an operational, open and FAIR EOSC ecosystem (2024)

HORIZON-INFRA-2024-EOSC-01-01: FAIR and open data sharing in support of the mission adaptation to climate change

HORIZON-INFRA-2024-EOSC-01-02: Supporting the EOSC Partnership in further consolidating the coordination and sustainability of the EOSC ecosystem

HORIZON-INFRA-2024-EOSC-01-03: Enabling a network of EOSC federated and trustworthy repositories and enhancing the framework of generic and discipline specific services for data and other research digital objects

HORIZON-INFRA-2024-EOSC-01-04: Long-term access and preservation infrastructure development for EOSC, including data quality aspects

HORIZON-INFRA-2024-EOSC-01-05: Innovative and customizable services for EOSC Exchange

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 154 –
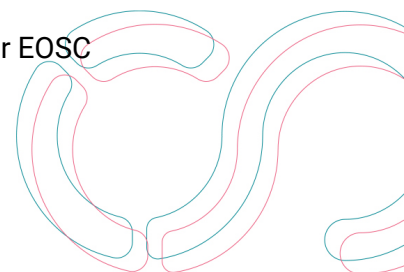
HORIZON-INFRA-2024-EOSC-02-01: Future Engagement Model for the EOSC Federation

Nine new projects that received a funding recommendation in the first round of calls for proposals in 2024 are expected to start work at the beginning of 2025.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931
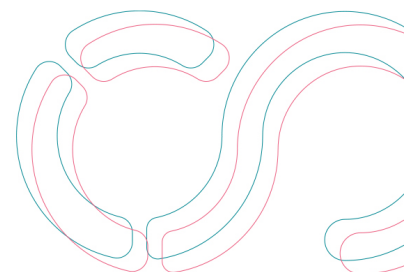
9 Conclusions

Digital age enables Open Science

Over the last few years, it has become clear that the rise of the digital age enables the evolution of research towards **Open Science**,[36] where knowledge is shared widely in multiple digital forms, including:

- **Documents**, such as publications;
- **Data**, in whatever format they are available; and
- **Software**, which has become a new way to represent knowledge.

As Open Science becomes the 'new normal', the world can expect:

- **Better science**, as researchers are able to optimise their work with more inputs coming more quickly from more sources;
- **Increased trust in science**, as the availability of related documents, data and software allow improved science reproducibility, leading to more reliable outputs and increased trust in results; and
- **Ability to meet global challenges**, as researchers from different disciplines are able to cooperate to address complex global challenges requiring diverse areas of expertise.

European leadership

Recognising the opportunity, Europe has taken the lead in building the European Open Science Cloud (EOSC). Its development falls into **three phases**:[37]

- **Preliminary period** (2015–2018), where a number of separate projects were undertaken that have provided the initial building blocks of the federation of infrastructures that will allow researchers to access the resources, they need to perform their research activities;
- **Transition period** (2019–2020), where a transition governance was established to increase the collaboration within and between research communities and projects, and to propose a sustainable model for EOSC organised around a new legal entity, the EOSC Association; and
- **Development period** (2021–2027), where, under **Horizon Europe,** the EOSC Association will act in partnership with the European Commission and the Member States and Associated Countries to orchestrate the research and innovation activities that will allow EOSC to become a reality at the service of the 2 million European researchers.

Future challenges and strategic objectives

Enabling Open Science in Europe involves three **key challenges** relating to people, knowledge and infrastructures that are addressed through corresponding **strategic objectives**.[38]

- Challenge: Convincing scientists that Open Science will allow them to do better and more rewarded research.
  Objective: **Open Science practices and skills are rewarded and taught, becoming the 'new normal'**. New roles and responsibilities will be created (e.g. data scientists, data stewards, etc.) and rewards and recognition schemes will evolve that acknowledge the value delivered by research in documents, data and software, extending the current rewards and recognition approach which is too heavily based on publications.

---

[36] The history of the digitisation of research in Europe, and the benefits of Open Science, are described in Section 1.

[37] The EC policy context for and stages of EOSC's development are described in Section 2.

[38] The challenges facing EOSC, and the overarching objectives that must be achieved in order to realise the potential benefits for science and society, are described in Section 3.

- Challenge: Enriching publications, data and software in order to make them usable by machines and scientists.
  Objective: **Standards, tools and services allow researchers to find, access, reuse and combine results**. All forms of knowledge need to be archived, referenced, described and credited by resource providers in order for users to retrieve information that is findable, accessible, interoperable and reusable.
- Challenge: Federating infrastructures in order to make them all available to scientists across borders and across disciplines.
  Objective: **Sustainable and federated infrastructures enable open sharing of scientific results**. Research communities have already engaged in developing infrastructures that provide digital resources to researchers. Those infrastructures will develop further while others will be created. The EOSC community has made the strategic decision to federate those infrastructures by providing three core federating elements:
    - a mechanism for naming and locating information and services;
    - a system for discovery of and access to information and services;
    - a common framework for managing identity and access.

Guiding principles and recommendations

As EOSC moves from the transition period towards a stable future, the following **guiding principles** have emerged that are now shared by the growing EOSC community.[39]

- **Multi-stakeholderism**: Open Science requires commitments from many stakeholders, from research-performing organisations to research-funding organisations and research infrastructures. EOSC's success will be driven by the cooperation of all these stakeholders adding their contributions towards the same goal.
- **Openness**: Open Science requires commitments towards openness from all stakeholders. While there are limits to openness with respect of privacy, security, intellectual property, and sovereignty, the community will have to strike the right balance in order to be 'as open as possible, as closed as necessary'[40].
- **FAIRness**: researchers will benefit from Open Science if knowledge is findable, accessible, interoperable and reusable. While these requirements are now accepted by the community, their implementation will still need continuous efforts.
- **Federation**: federating a wide diversity of infrastructures is a challenge at many levels, technical as well as social. EOSC's success will derive from the full participation and commitment of existing and future infrastructures.
- **Machine-actionable**: the volume and the diversity of knowledge available will require the use of machines at the service of researchers in order to reap the benefits of Open Science.

Sharing these principles, the transition governance organised Working Groups that issued a set of **recommendations** for EOSC to strengthen its foundations and deploy a Minimum Viable EOSC (MVE) that will convince researchers of the value added by EOSC:

- Evolve community governance;
- Define, design, implement and evolve the federating core of EOSC;
- Extend FAIR requirements to all digital objects;
- Design and deploy standards and tools;
- Widen promotion of and education on Open Science;
- Reward and recognise Open Science practices;
- Monitor policy deployments of Open Science.

---

[39] Guiding principles and recommendations are described in detail in Section 4.
[40] For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".

Action areas

Analysing the detailed scope of efforts that are necessary to make EOSC a reality, two sets of **action areas** have been identified. For each area, **status, gaps and priorities** have been established.

Implementation challenges:[41]

- Design, develop and deploy an interoperable environment for referencing using **identifiers**, describing using **metadata and ontologies**; and reusing all forms of knowledge;
- Design, develop and deploy **FAIR metrics and certification** procedures;
- Design, develop and deploy a shared **authentication and authorisation infrastructure**;
- Design, develop and deploy user-oriented services delivered by resource providers, with value-added **user environments** and **resource provider environments**.
- Develop and promote the **EOSC Interoperability Framework**.

Boundary conditions:[42]

- Agree on **Rules of Participation** for all EOSC stakeholders;
- Ensure continuous **landscape monitoring**, covering infrastructures, initiatives and policies;
- Identify viable **funding models**;
- Identify necessary **skills and** develop corresponding **training**;
- Promote new approaches for **rewards and recognition**;
- Develop and execute a **communication** strategy for EOSC;
- **Widen** EOSC **to public and private sectors**.

Roadmap

Focusing on **Stage 1**, the proposed priorities for each of the action areas have been assembled into a **roadmap**[43] for 2021–2022 in order to:

- Build or improve the key components of the **Minimal Viable EOSC** (MVE);
- **Enable Open Science** and deliver scientific impact;
- Enable **EOSC sustainability** and stakeholder engagement;
- Deploy the **Web of FAIR Data and Related Services** at the researcher level.

The following priorities were chosen to achieve their corresponding strategic objectives:

- Ensure that Open Science practices and skills are rewarded and taught, becoming the 'new normal':
  - Develop an effective **governance framework** that coordinates activities and directs development of the MVE;
  - Establish a **risk management structure** for EOSC;
  - Develop a cooperation framework to implement **Rules of Participation**;
  - Coordinate and align relevant **skills** curricula **and training** frameworks;
  - Create a **rewards and recognition** framework that incentivises FAIR data and Open Science;
  - **Inform** stakeholders and **engage** with communities;
  - **Promote** EOSC at all levels.
- Enable the definition of standards, and the development of tools and services, to allow researchers to find, access, reuse and combine results:

---

[41] These implementation challenges are described in detail in Section 5.
[42] These boundary conditions are described in detail in Section 6.
[43] This roadmap is described in detail in Section 8.
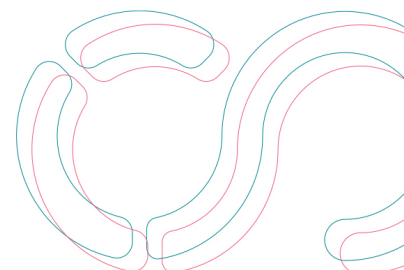
- Agree and adopt open specifications and **metadata frameworks** for improved discovery and interoperability;
- Develop common **standards for interoperability** and support development of schemas **and APIs**;
- Develop **standards and tools for** archiving, referencing, describing and citing **research software**;
- Encourage and incentivise effective **documentation and sharing of research software**;
- Establish a **common framework for managing AAI**;
- Implement the **EOSC PID policy and architecture**.
- Establish a sustainable and federated infrastructure enable open sharing of scientific results:
  - Ensure that **domain-specific user environments** are included in EOSC;
  - Support research communities to develop and **adopt domain-specific standards** where required;
  - Develop common standards and **framework for minimum metadata**;
  - Build a **common digital search** to enable discovery and exchange across federated repositories;
  - Build structures to ensure **effective metadata support and coordination** for research communities;
  - Liaise internationally to develop a **global cooperation framework** for Open Science infrastructures (L1+L2).
  - Define models for availability and **costing of services across borders**.

Levels of implementation (European, national, institutional) have been identified for each priority, together with key performance indicators for the expected outcomes.

At a later moment the priorities for **Stage 2**[44] (2023–2024) and **Stage 3**[45] (2025–2027) were developed. The most relevant topics according to the EOSC community for 2025 are:

1. Create an aligned core part of the curricula for data stewardship for Europe; stimulate the development of curricula in the countries, for the different types and levels of data stewards.
2. Collect ongoing national best practices and case studies contributing to the development of EOSC as described in the SRIA.
3. Advance the technical specifications to support digital objects to be FAIR-by-design and evaluate options to adopt the concept of FAIR digital objects for operational data services.
4. Stimulate community-specific FAIR metrics and data quality constraints in harmony with the FAIR assessment, especially on the interoperability aspect.
5. Develop, prototype and test machine-actionable research data repositories and link them to the EOSC EU Node, in line with the EOSC Interoperability Framework.
6. By the end of 2025 a sustainable initial EOSC Core should be up and running and thereafter further enhanced and extended to support emerging and challenging use cases. For this a user driven mechanism should be developed to determine a selection of most needed horizontal and thematic services, which can serve (as a pilot) for the centrally financed service portfolio for the period after Horizon Europe.
7. Define the requirements for national / local / thematic EOSC nodes. Stimulate their development and those that may come from Research Infrastructures or other parties.
8. A dedicated provision for supporting the outcome of the EOSC EU Node procurement; handing it over to the (interim) operating party; running and maintaining it; creating the

---

[44] See paragraph **Error! Reference source not found.**
[45] See paragraph **Error! Reference source not found.**
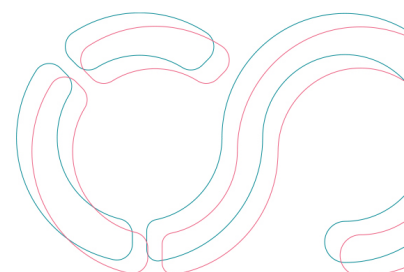
possibility to enhance it by 'onboarding' (pilot) nodes from countries / regions, domains, research infrastructures, etc.

EOSC Association takes the lead

Established as a legal entity under Belgian law on 29 July 2020 following a royal decree, the EOSC Association held its first official General Assembly on 17 December 2020. The 4 founding members – CESAER, CSIC, GARR, and GÉANT – formally admitted a total of 142 members, of which 21 are nationally mandated organisations, and 49 observers. These members and observers include research-performing organisations, research-funding organisations, research infrastructures, data service providers and others, which together represent research and innovation stakeholders across the EU and beyond.

It is now the mandate of the EOSC Association, guided by its elected president and board of 8 directors, to evolve this Strategic Research and Innovation Agenda (SRIA). Developed by the EOSC governing bodies during the transition period, the SRIA outlines the context and vision of the EOSC initiative, describes its strategic objectives, highlights its values and principles, offers recommendations, sets priorities and provides a roadmap for the next two to seven years.
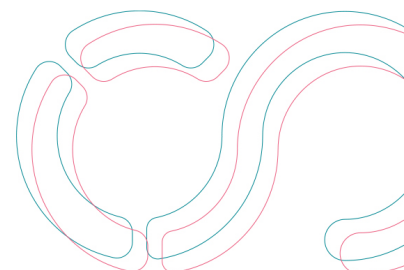
As described by the EOSC Interoperability Framework, in order to achieve that vision, the EOSC Association will have to address **technical, semantic, organisational and legal** challenges.

Given the representative nature, enthusiasm and commitment of its members and observers, the EOSC Association is well placed to engage all stakeholders and orchestrate and schedule their efforts, to enable EOSC to realise its full potential with the support of the EOSC Partnership between the European Union and the EOSC Association.

The EOSC Partnership drives the implementation of EOSC by a series of EC funded projects and several initiatives related to EOSC from the members of the EOSC Association. The wide range and diversity of projects forms an enormous challenge for the EOSC Association in terms of coordination. Expectations placed on new Horizon Europe projects by the EC to ensure collaboration and coordination with the EOSC Association, following the specific access conditions as defined for each EOSC call, will help to alleviate this position. The Partnership will strive to avoid overlapping activities or divergent approaches and favour joint work towards an orchestrated result.

Next to this, the technical and financial challenges that have to be overcome in trying to reach the objectives are very large. But even more significant challenges lie in combining local, national, and regional initiatives towards a true Open Science Commons, with global convergence on standards that are truly able to improve research. The federated approach requires alignment and convergence on standards and practices, which is not only technically challenging but necessitates lengthy consensus-building processes to ensure acceptance and adoption by research communities.

Fortunately, the EOSC Association is making progress and is rapidly becoming accepted as the voice of the community, with a General Assembly that now comprises 256 organisations. The Association is also well positioned to help coordinate across implementation projects and perform the necessary liaison with Member States to stimulate investments in EOSC. Ensuring political support, from the EC and the countries, will be key in the coming period to ensure national open science initiatives and RIs are incentivised to federate into EOSC, for it to gain critical mass and become a valuable resource for the research communities of the future.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 160 –

Appendix A   Related documents

During the transition period (2019–2020), the EOSC governing bodies have developed, discussed, reviewed and approved a series of documents which constitutes the corpus of knowledge describing EOSC's present and future. The documents are listed in Table A.1 below and summarised in the following sections.

Document title	Section
EOSC Partnership Proposal	A.1
EOSC Authentication and Authorisation Infrastructure	A.2
Persistent Identifier (PID) Architecture for EOSC	A.3
A Persistent Identifier (PID) policy for the European Open Science Cloud	A.4
Recommendations on FAIR Metrics for EOSC	A.5
Recommendations on certifying services required to enable FAIR within EOSC	A.6
Six Recommendations for Implementation of FAIR Practice	A.7
EOSC Interoperability Framework	A.8
Scholarly Infrastructures for Research Software	A.9
Landscape of EOSC-Related Infrastructures and Initiatives (the Landscape Report)	A.10
Country Sheets Analysis	A.11
Rules of Participation	A.12
Solutions for a Sustainable EOSC (the FAIR Lady Report)	A.13
Digital skills for FAIR and open science	A.14
Risk Management Study report	A.15
EOSC-Core Operational Costs Study report	A.16

Table A.1: List of documents related to the EOSC SRIA

A.1   European Open Science Cloud (EOSC) Partnership: Draft proposal for a European Partnership under Horizon Europe

The European Open Science Cloud (EOSC) Partnership will enable a trusted, virtual, federated environment in Europe to store, share and reuse research data across borders and scientific disciplines. The Partnership will bring together institutional, national and European initiatives and engage all relevant stakeholders to co-design and deploy a European Research Data Commons where data are findable, accessible, interoperable and reusable (FAIR). This European contribution to a 'Web of FAIR Data and Related Services for Science' will enhance the possibilities for researchers to find, share and reuse publications, data and software leading to new insights and innovations, higher research productivity and improved reproducibility in science.

The EOSC Partnership will be the European focal point to reduce the fragmentation of research information, harmonise access policies, and make the necessary links for eliminating silos and enhancing the circulation of knowledge in digital form. The Partnership will be instrumental to developing an ecosystem of research data and related services,
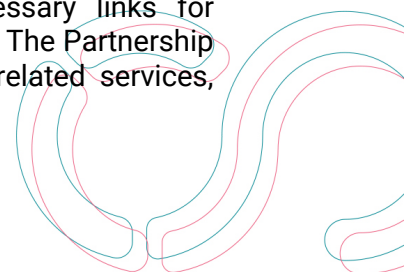
covering the whole data lifecycle, from discovery and mining to storage, management, analysis and reuse.

The Partnership will enable a new scale of cross-disciplinary and collaborative research that addresses major societal challenges and accelerates Europe's transition to open science, as presented in the Communication 'A European strategy for data' in 2020.

The Partnership proposal describes the context, objectives and impact of the Partnership. It outlines the necessity for a Co-programmed European Partnership on EOSC, its planned implementation, activities, resources and governance structure.

Link

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

A.2   EOSC Authentication and Authorisation Infrastructure

The purpose of the EOSC AAI report is to present the basic elements for building a common global ecosystem for identity and access control infrastructures for the European Open Science Cloud (EOSC).

Since EOSC is part of an international environment of research and education, the principles and architecture choices established by the EOSC AAI Task Force are globally viable.

The AAI Task Force has broken its outputs down into five parts:

- EOSC AAI First Principles and Requirements;
- EOSC AAI Baseline Architecture;
- EOSC AAI Federation Definition;
- EOSC AAI Federation Implementation;
- EOSC AAI Use Cases.

The current status of the EOSC AAI architecture is based on the Authentication and Authorisation for Research and Collaboration (AARC) Blueprint Architecture 2019 and identifies the challenges and the areas that require further work. The EOSC AAI Task Force has worked in collaboration with AEGIS, the AARC community and other stakeholders to address these gaps and feed the requirements from EOSC into the development of the AARC Blueprint Architecture 2020.

Link

https://op.europa.eu/en/publication-detail/-/publication/d1bc3702-61e5-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-188566729

A.3   Persistent Identifier (PID) Architecture for EOSC

The technical architecture document on persistent identifiers (PIDs) identifies opportunities for how interoperability between PID services can be achieved within the framework of the European Open Science Cloud (EOSC).

Drawing from the EOSC guidelines, it identifies stakeholders at different levels of PID namespaces, their roles and their ability to enforce policies.

The document is mainly targeted at PID and generic service providers and practitioners as guidelines on implementation of PIDs and related services compliant with the PID policy within EOSC. However, it may also be of interest for technically interested PID practitioners and for policy makers acting at an organisational level.

Link

https://op.europa.eu/en/publication-detail/-/publication/3136c3e6-4f07-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-183178379

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 162 –
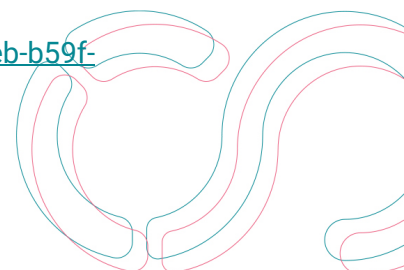
A.4 A Persistent Identifier (PID) policy for the European Open Science Cloud

The persistent identifier (PID) policy for EOSC sets out a series of principles, actions and associated roles and responsibilities to establish a viable, trusted PID infrastructure suitable for the long-term sustainability of EOSC. It is written for senior decision makers within potential EOSC service and infrastructure providers, and will be of interest to all EOSC stakeholders.

The policy defines a set of expectations about how, and for what purposes, persistent identifiers will be used to support a functioning environment of FAIR research. It acknowledges and builds on the wide variety of PIDs and associated services already in active use for a range of applications, and promotes interoperability across service providers and research infrastructures to create a functioning PID ecosystem. Requirements of providers and the basic services they offer are also outlined as well as the overarching governance and sustainability needed.

The associated 'PID Architecture for EOSC' document will guide the implementation of the policy. This phase of work will be overseen by the EOSC Association and delivered in collaboration with critical projects such as EOSC Future and FREYA.

Link

https://op.europa.eu/en/publication-detail/-/publication/35c5ca10-1417-11eb-b57e-01aa75ed71a1

A.5 Recommendations on FAIR Metrics for EOSC

This report contains an analysis of activities relevant to the definition of FAIR metrics in the EOSC context. It offers an analysis of gaps and potential opportunities, makes recommendations on the definition and implementation of metrics, and defines priorities for future work. The report proposes a set of metrics for FAIR data in EOSC, phasing these over the time periods used in the next work programme to incrementally enhance existing FAIR practices. These apply to data as well as other research objects such as software.

The report notes that these metrics need to be extensively tested to ensure they are appropriate to all research communities, and do not prevent engagement in EOSC or any other unintended consequences. As with other EOSC recommendations, work builds on existing international activity, notably, in this case, the RDA FAIR Data Maturity Model Working Group and FAIRsFAIR.

Link

https://op.europa.eu/en/publication-detail/-/publication/ced147c9-53c0-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-184008165

A.6 Recommendations on certifying services required to enable FAIR within EOSC

This report provides initial recommendations on certifying services required to enable FAIR research outputs. It has been developed by the Metrics and Certification Task Force within the EOSC FAIR Working Group. The Working Group has focused on defining the repository certification approach in EOSC, but also considered certification of other services such as PIDs required to support and enable FAIR.

The certification guidelines build on the findings of the relevant projects and the community input gathered. Certification is one of the main strands of work of the FAIRsFAIR project and was a key input to the recommendations. An adaptation of the CoreTrustSeal certification to reference the FAIR principles, as recommended by the 'Turning FAIR into reality' report, is

proposed, with requirements for certification only becoming mandatory in later stages of EOSC since the community is still in the early stages of maturity.

Link

https://op.europa.eu/en/publication-detail/-/publication/70aa74b5-53bf-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-184009543

A.7   Six Recommendations for Implementation of FAIR Practice

This report analyses the state of FAIR practices within diverse research communities and FAIR-related policies in different countries, concluding with six practical recommendations on how FAIR can be turned into practice. These recommendations are aimed primarily at decision-making entities of the European Open Science Cloud (EOSC), as well as research funders:

1. Fund awareness-raising, training, education and community-specific support.
2. Fund development, adoption and maintenance of community standards, tools and infrastructure.
3. Incentivise development of community governance.
4. Translate FAIR guidelines for other digital objects.
5. Reward and recognise improvements of FAIR practice.
6. Develop and monitor adequate policies for FAIR data and research objects.

In order to ensure widespread benefits of EOSC, improvements in FAIR practices are necessary. The report authors believe that the timing of the report, which coincides with the fully fledged launch of EOSC, could help EOSC, research funders and policy makers make crucial strategic decisions about investment needed to put FAIR principles into practice.

Link

https://op.europa.eu/en/publication-detail/-/publication/4630fa57-1348-11eb-9a54-01aa75ed71a1/language-en/format-PDF/source-166584930

A.8   EOSC Interoperability Framework

Document version: v1.0, Draft for community consultation, 3 May 2020

The EOSC Interoperability Framework identifies the general principles that should drive the implementation and use of EOSC, organising them into four layers of technical, semantic, organisational and legal interoperability. Achieving interoperability is essential in order to federate services, integrate data and enable interoperation with applications or workflows for analysis, storage and processing. Without a common framework of standards for the discovery and reuse of data and services beyond their existing contexts, EOSC will be unable to provide the anticipated added value for research communities.

For each aspect of interoperability, the EOSC Interoperability Framework identifies problems and needs, and makes a series of recommendations on what should be implemented within EOSC. A proposed framework for FAIR digital objects and an outline model which centres on the object layer and touches on service components is also provided. A proposal for a set of building blocks to make this a reality is also visualised. Recommendations on how this work should be taken forward in future years and priorities for investment such as defining a minimal common metadata framework are also made explicit.

Link

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

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 164 –
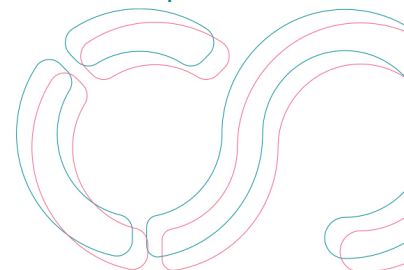
A.9   Scholarly Infrastructures for Research Software

The Task Force on Scholarly Infrastructures for Research Software, created by the EOSC Architecture Working Group, has established a set of recommendations to allow EOSC to include software, next to other research outputs such as publications and data, in the realm of its research artifacts.

This work is built upon a survey and documentation of a representative panel of current operational infrastructures across Europe, comparing their scopes and approaches.

The report summarises the state of the art, identifies best practices, as well as open problems, and paves the way for federating the different approaches in support of the software pillar of EOSC.

Link

https://op.europa.eu/en/publication-detail/-/publication/145fd0f3-3907-11eb-b27b-01aa75ed71a1/language-en/format-PDF/source-178301654

A.10   Landscape of EOSC-Related Infrastructures and Initiatives (the Landscape Report)

This report describes activities relating to EOSC in the Member States (MS) and Associated Countries (AC). It summarises existing policies and investments based on input from the MS and AC, and using the expert knowledge of the EOSC Landscape Working Group (WG) members and delegates to the EOSC Governing Board, complemented by information from Horizon 2020 research projects and from open sources. The work builds on existing surveys and information provided by national authorities, various stakeholder communities and the relevant H2020 projects in close collaboration with the MS and AC. Initial reference material includes the recent report of the e-Infrastructure Reflection Group (e-IRG), findings of the Commission Expert Group on National Points of Reference on Scientific Information, the surveys carried out by the OpenAIRE project, the EOSC-Pillar project, analysis of preliminary mapping of the UK's research and innovation infrastructure landscape, the experience of the European Strategy Forum on Research Infrastructures (ESFRI) on cross-disciplinary collaboration of ESFRI Landmarks, other relevant documents identified by the WG members, and outcomes of the survey (country sheets) performed by the Landscape WG itself. The WG has collated input from e-infrastructures, including data, HPC and distributed computing facilities, from European and National Research and Education Networks (NRENs), pan-European research infrastructures and ESFRI roadmap projects and clusters, and from interest groups such as the Research Data Alliance.

It was found that the majority (61%) of the MS and AC respondents have policies in place regarding open access to scholarly publications, but only 34% have a policy in place regarding FAIR data (although, encouragingly, 44% have one either planned or under development). Few countries seem ready to mandate that research data should automatically be made open. Relatively few countries (21%) mention EOSC in their policies, but 38% plan to do so in future; only three countries so far (Bulgaria, Denmark, Romania) include mention of EOSC as part of their funding criteria. More than half of responding countries have nominated contact points for open science (53%) and for EOSC (42%).

While it has proved difficult to obtain definitive and quantitative data on national levels of investment, it is already clear that the majority of the countries surveyed have made significant investment in national e-infrastructures of one kind or another that could, in principle, either be federated as part of EOSC, or made accessible to users through EOSC. The same applies to many of the data-intensive ESFRI landmark infrastructures. Such investment over the past decade now totals billions of euros, much more than the planned central investment into the

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 165 –

EOSC-Core. This shows that EOSC will only reach its true potential through effective federation of national and research infrastructure resources.

A virtual validation workshop was held online on 27 and 28 April 2020, to validate the provided contributions.

Link

https://op.europa.eu/en/publication-detail/-/publication/9c7f3c1e-2aea-11eb-9d7e-01aa75ed71a1/language-en/format-PDF/source-173604556

A.11 Country Sheets Analysis

The Landscape Working Group (WG) of the EOSC Executive Board undertook a survey of the landscape in late 2019 and revised the initial results in early 2020. Country questionnaire sheets were completed for 47 EU MS, AC and Other Countries (OC), which aimed to assess the existence of national-level policies in relation to Open Science (OS), EOSC and research evaluation, as well as to get a sense of the potential range and number of e-infrastructures that might be federated as part of or made available to users through EOSC. The Landscape Analysis reflects the rapid development in EOSC and delves deeper into the country sheet data and complements the work being carried out by the INFRAEOSC-5B projects, which are assessing the landscape at the local level.

In the majority of countries, many of the relevant policies are still in the planning stage (policies relating to OS, data/services, and referencing EOSC). With regard to available infrastructures that may be federated or made available, there are at least a few candidates in all of the countries reviewed. As the emerging EOSC-related indicators are further elaborated and agreed they will need to be refined to better reflect a more comparable and qualitative approach to EOSC readiness. The EOSC Partnership and countries must collectively decide which specific policy elements should be addressed in national policies and ensure that these are captured in future assessments. In addition, it must also be decided which sources of open data should be used for different fields of the country sheets moving forward. To enable its effective use by decision makers and other stakeholders, different use cases for the information provided by the EOSC-readiness profiles were defined, ensuring that they reflect the emerging indicators across the various EOSC Association, EOSC WGs, Task Forces, ESFRI, the EC and via the recently released Strategic Research and Innovation Agenda (SRIA) draft document. A virtual final validation workshop was held online on 28 and 29 September 2020, to validate the provided contributions to this analysis. The Landscape Analysis and the final validation workshop would not have been possible without the support of the EOSC Secretariat via co-creation budget.

Link

https://op.europa.eu/en/publication-detail/-/publication/95e4a900-2a21-11eb-9d7e-01aa75ed71a1/language-en/format-PDF/source-173316815

A.12 Rules of Participation

The EOSC Rules of Participation (RoP) underpin the policies, processes and procedures required to provide assurance of openness, quality and trust in the practices and services offered through voluntary participation in EOSC. The rules, which have been developed with input from EOSC implementation projects and using proposals published by other EOSC Working Groups, are at times aspirational in nature due to a number of policies and procedures that remain to be defined or agreed.

The RoP apply at the point of first interaction with EOSC and continue throughout involvement. They apply to all EOSC users including providers and consumers of EOSC resources. They apply to all resources provided or accessed via EOSC, including not only to data and services

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 166 –
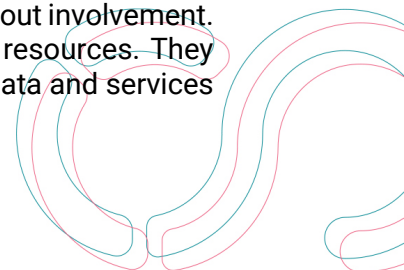
but also, for example, to software, tools, workflows, training and consultancy. They apply to publicly funded research users, including international users, and to publicly funded and commercial providers.

The RoP define rights and obligations, therefore EOSC will need to monitor compliance and apply sanctions for non-compliance as appropriate. They have been formulated primarily with the Minimum Viable EOSC and its initial operational phase in mind and are expected to evolve as EOSC, its user base, and usage scenarios grow beyond the currently foreseen landscape.

The Rules of Participation are referred to as 'minimal'. This is understood to mean that the RoP should be kept as simple, short and lightweight as possible whilst still fulfilling their purpose of defining the rights, obligations and accountability governing the activities of those using EOSC. The RoP are formulated as an overarching set of top-level rules that apply across the whole of EOSC, each with a commentary that discusses the issues relevant to that rule. In this way the Rules themselves should be relatively stable, while the commentary can be revised as EOSC evolves. The RoP will require further elaboration if they are to provide specific criteria that implementations must comply with. However, that level of detail will depend on the specificities of particular resource types and, accordingly, development of more detailed rules for specific resource types is devolved to those involved with those resources.

Link

https://op.europa.eu/en/publication-detail/-/publication/a96d6233-554e-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-184432576

A.13 Solutions for a Sustainable EOSC: A FAIR Lady report from the EOSC Sustainability Working Group

The Sustainability Working Group was tasked to explore possible means for ensuring the sustainability of the European Open Science Cloud as of 2021. The Working Group produced interim versions (known as the 'strawman' and 'tinman') of its report and collected feedback from the EC, Executive Board, Governance Board and the stakeholder community on each version. It also took into account the progress towards the EOSC goals and the outputs of commissioned studies. The final result is a standalone document (the 'FAIR Lady' report) that was submitted for publication on 13 November 2020. It considers the financing model, legal vehicle and governance structure under the planned Co-programmed European Partnership with the EC, as well as the regulatory and policy environment of EOSC. It recommends beginning with a first iteration to establish a Minimum Viable EOSC (MVE) addressing the needs of publicly funded researchers exploiting openly available data. Subsequent iterations expand EOSC to engage a wider user base, including the public and private sector, and promote its use beyond FAIR research data.

Link

https://op.europa.eu/en/publication-detail/-/publication/581d82a4-2ed6-11eb-b27b-01aa75ed71a1/language-en/format-PDF/source-175468053

A.14 Digital skills for FAIR and open science

A comprehensive skills and education strategy is needed to translate the vision of a strong EOSC research data ecosystem that exploits digital technologies into reality. The Skills and Training Working Group was tasked with providing a framework for building competence (skills) and capabilities (training) for EOSC. This work has involved consultation with a range of existing initiatives in Member States and Associated Countries, stakeholder communities and H2020 EOSC projects, and aimed to increase alignment on key components for skills development and training, identify what structures are needed to make EOSC sustainable, and determine how these can be embedded by different EOSC stakeholders.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 167 –

The final report documents outcomes in four key areas:

1. Minimal skillset. A framework describing the digital skills required to develop the next generation of FAIR and Open Science professionals, to maximise the uptake and utilisation of EOSC for all the actors (roles) in the EOSC ecosystem: researchers, research software engineers, data scientists and data analysts, data research infrastructure support professionals, data curators, data stewards, data librarians, EOSC enablers, EOSC educators, policy makers (including funders), and citizens. Recommendations are made regarding the use of this framework to define the skills and training necessary to support implementation of a Minimum Viable EOSC (MVE), which also constitutes the base set of digital skills for FAIR and Open Science.
2. Competence centres. An analysis of approaches that organisations at different levels use to implement their skills and training programmes for FAIR, Open Science and EOSC. Commonalities are identified in the range of services offered; focus of skills and training programmes; governance and business models utilised to increase sustainability; and alignment with other initiatives. Recommendations focus on areas where EOSC could support further development of competence centres to facilitate EOSC's growth.
3. Training catalogues. Review of current initiatives to demonstrate the need for the development of a trusted, long-lasting and federated EOSC catalogue for training resources. Recommendations detail the process for its realisation, and an interoperability and harmonisation framework is provided for current catalogues.
4. National strategies for digital skills. A study on national strategies for the development of digital skills that support EOSC implementation. Recommendations made include an EOSC Leadership Programme to foster advancement of national strategies to maximise development of digital skills relevant to EOSC.

This work identifies the next steps to overcome existing gaps and barriers to vital skills and training development, and further work will be needed to continue to advance this area.
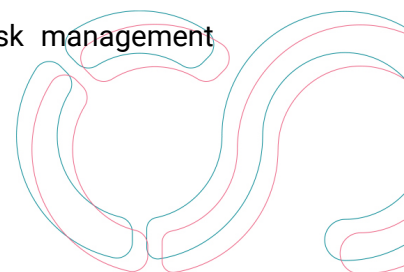
Link

https://op.europa.eu/en/publication-detail/-/publication/af7f7807-6ce1-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-190694287

A.15 Risk Management Study report – Support the strengthening of the EOSC Risk Governance through the implementation of an effective Risk Management System

A targeted study has been conducted by AON Hewitt on behalf of the EOSC Sustainability Working Group in order to provide clear and structured guidance on how to incorporate risk management into the governance of EOSC. Its aim was to understand the complex context in which EOSC operates, the variety of stakeholders involved and the maturity of the current and future EOSC governance systems, as well as the assessment and management of risks. The study identified 48 gaps in risk governance and made 32 recommendations to address these. The recommendations are to be considered by the EOSC Association, the EOSC contributing projects and the overall EOSC Partnership to develop a comprehensive Enterprise Risk Management system (ERM).

The analysis assessment highlighted that EOSC operates within a multiple factor environment with a high degree of complexity affecting the governance structure. A number of key drivers that act on multiple risks and may jeopardise participation in the EOSC Association and adoption of its services have been identified. In particular:

- The absence of a clear and formalised risk governance and risk management structure;

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 168 –

- The incomplete and not yet formalised definition of the value proposition for each stakeholder (internal and external to the Association);
- The incomplete definition of the economic-financial plan and a budget to ensure the long-term financial sustainability of EOSC;
- The absence of the definitions of roles, responsibilities, process and procedure to intervene in case of system failure.

However, the study also found a human capital very rich in multi-disciplinary technical skills, sensitivity to governance issues, and passion for the activities to be carried out and for the belief in EOSC itself. Moreover, the presence of almost all the essential aspects for the construction of effective risk management was also found. The study considered the EOSC major players to have a high interest in and awareness of the importance of the subject and that between them the stakeholders also have the relevant skills and experience in the different areas of risk management, even if currently they are not systematically introduced in a complete and effective risk governance.

The main recommendations are:

- Launch a comprehensive plan for setting up a risk management;
- Establish a governance structure for risk management;
- Define the EOSC ERM policies;
- Design the risk assessment and reporting process;
- Map the skills and competences of management bodies to assure independence in decision making;
- Establish a risk awareness programme;
- Set up an infrastructure and data security team.

Implementing these recommendations will significantly increase the value of EOSC and benefit its stakeholders by supporting its objectives and allowing a more effective use and allocation of resources. The ERM will also help to protect the assets, the corporate brand, the know-how of the key people, and optimise the operational efficiency.
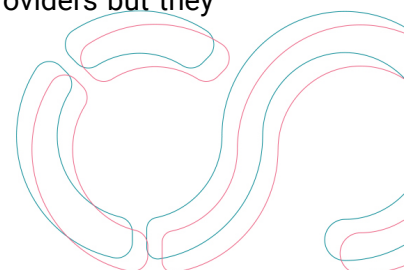
Link

https://doi.org/10.5281/zenodo.4423167

(Please note that the link is to the Executive Summary rather than the full report)

A.16 EOSC-Core Operational Costs Study report – The Vivus Study: Ensuring that EOSC-Core & Minimum Viable EOSC are sustainable through a study on their costings, potential business models and funding schemes

A targeted study has been conducted by AcrossLimits and Boundaryless on behalf of the EOSC Sustainability Working Group. Its initial aim was to examine EOSC components in relation to the mapping between the EOSC-Core services and the technical e-infrastructure services that would deliver them, as well as to explore business models that would support the EOSC-Core. The scope of the study later grew to include the Minimum Viable EOSC (MVE), service pricing and funding sources.

On investigating the EOSC-Core costs, a total cost of approximately €7M was revealed, which, however, seems insufficient to adequately deliver a sustainable service, given the complexity of EOSC. On top of this it was calculated that service pricing will be found to be in the region of an additional 25% to 50% on top of service costs. Another finding related to the EOSC-Core costs study was that the existing e-infrastructures offer an advantage to EOSC, in that they not only simplify the relationship that EOSC will maintain with its service providers but they also embody a great deal of knowledge important to EOSC.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 169 –

The study also discovered that cross-border delivery of research services is constrained by a number of non-technical obstacles that inhibit EOSC from achieving its multi-disciplinary and multinational objectives. EOSC stakeholders will need to persuade their national governments to allow EOSC participants to provide services to other Member States as part of the agreed spectrum of joint EOSC activities.

EOSC is not a technology and has a role that is strategic as well as operational. In order to build the needed trust in its user communities, a viable business model is required to migrate towards longer-term business-like income streams. The business-modelling activity considered EOSC from an ecosystem perspective and investigated relationships and dependencies. Two models were identified as being suitable, the Learning Engine and the Transaction Engine, with a final conclusion being that a hybrid model would be required initially.

Continued public funding for the EOSC-Core is felt to be acceptable in the short term, but in the medium-term actions should be taken to gain sustainability independent from the public funding. There are many examples of strategic platforms that have become independently sustainable; best practices should be collected and investigated for resources useful to EOSC.

Link

https://doi.org/10.5281/zenodo.4395885
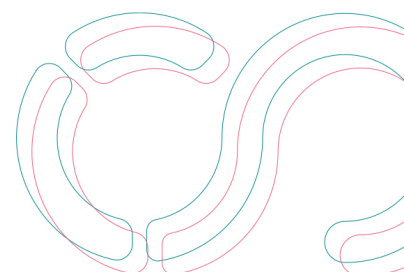
Appendix B Priorities at earlier stages

B.1 Priorities Stage 1 (2021–2022)

The SRIA vision and open consultation results have been used together with the three high-level EOSC objectives to develop a set of prioritised activities that should be initiated in Stage 1. The activity areas and expected outcomes achievable during the HE funding programme are described below. In all cases, coordination with existing H2020 activities is required to deliver the expected outcomes. As stated above, the level of delivery considered most appropriate for each activity and outcome is also indicated.

The activities and outcomes of Stage 1 will facilitate reaching the required level of readiness for the subsequent execution in Stage 2 and Stage 3 of the HE Work Programmes for EOSC. They can be seen as the foundational aspects of a federated infrastructure and they form the basis for developing the required coordination and governance of EOSC.

### B.1.1 Stage 1/Objective 1

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

Priority activities:

- Develop an effective governance framework that coordinates activities and directs development of the MVE (L1).
- Establish a risk management structure for EOSC (L1).
- Develop a cooperation framework to implement Rules of Participation (L1+L2).
- Coordinate and align relevant skills curricula and training frameworks (L1+L2+L3).
- Create a rewards and recognition framework that incentivises FAIR data and Open Science (L2+L3).
- Inform stakeholders and engage with communities (L1+L2+L3).
- Promote EOSC at all levels (L1+L2+L3).

Expected outcomes:

- A well-established governance structure, representative of the various stakeholders, that is successfully coordinating activities relating to the MVE.
- Well-documented use cases and related services.
- MVE technical developments are supported by coordinated efforts across borders and scientific communities, ensuring that relevant needs and practices from different stakeholders are taken into account.
- EOSC has clear engagement from the community with an increasing number of users and service providers.
- Workable business models are defined to allow / guarantee future sustainability and development of EOSC and Open Science.
- Improved availability of highly and appropriately skilled professionals enabling the practice of FAIR data and Open Science through their knowledge of standards, applications and tools, and best practices.
- Competences related to Open Science that allow researchers to transform the way they exploit research outputs are embedded into the research ecosystem, leading to better-quality research and increased productivity.
- Design of a coordinated rewards and recognition framework based on multiple dimensions, including Open Science.
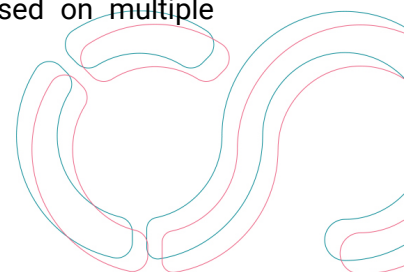
- Increased awareness and uptake of EOSC as a result of stakeholder engagement activities.
- An EOSC brand strategy with clear guidelines on reuse of the logo and assets.

Considerations based on the recommendations from the EOSC Executive Board Working Groups are as follows:

Coordination and provisioning of clear policy and governance Rules of Participation for users and providers of EOSC are essential in Stage 1. Explicit coordination between projects is necessary when multiple independent projects move towards a common end point. Clarifying the role of the EOSC Association and the development of a workable governance and sustainability model must be started and is considered a high priority. Business models that give tangible benefits for data producers and service providers need to be developed. Clarity on delivery for and engagement with EOSC should be included in all relevant HE calls.

Rules of Participation need to be bi-directional incentives for data producers to make their resources available and provisioned to EOSC. An EOSC without the required data is a poor investment. The heterogeneity in the data-producer landscape (in all areas including volume and theme) makes defining a single set of rules particularly challenging. Coordination between national agencies and between specific research infrastructures will be needed. Specific aspects of policy and the legalities of data sharing indicate the need for a carefully considered EOSC data policy framework, which should take into account the existing policies at the institutional, national and European level.

Priority should be given to training, outreach and thematic specificities. The success of EOSC will to some degree be tied to enabling the research communities to work with and within a landscape that has Open Science as an overall vision. FAIR data and Open Science are steadily becoming the new normal. EOSC should be used to drive that migration and actively promote and train researchers at all career levels to follow and deliver best practice Open Science. The heterogeneity of the research landscape does not allow a 'one size fits all' approach and in many cases a top-down approach could also result in a lack of understanding of nuance and complexity in the research workflow. Prioritising training and outreach, as well as policies for rewards and recognition for the professionals involved, is essential.

B.1.2   Stage 1/Objective 2

Enable the definition of standards, and the development of tools and services, to allow researchers to find, access, reuse and combine results

Priority activities:

- Agree and adopt open specifications and metadata frameworks for improved discovery and interoperability (L1).
- Develop common standards for interoperability and support development of schemas and APIs (L1+L2).
- Develop standards and tools for archiving, referencing, describing and citing research software (L1).
- Encourage and incentivise effective documentation and sharing of research software (L2+L3).
- Establish a common framework for managing AAI (L1).
- Implement the EOSC PID policy and architecture (L1+L2+L3).

Expected outcomes:

- Researchers have access to a Web of FAIR Data and Related Services based on theme and impact area.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu   – 172 –
Reg. number: 0755 723 931 | VAT number: BE0755 723 931
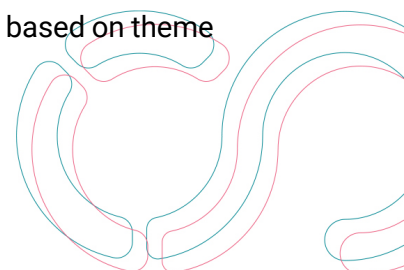
- Digital objects follow relevant community standards and are interoperable.
- Standards, recommendations and methodologies essential for putting FAIR principles into practice and implementing the EOSC ecosystem are developed in alignment, where possible, with international efforts and practices.
- Uptake of and compliance with FAIR principles and practices by national and European research-resource providers and repositories through EOSC is standard.
- European researchers can find, access and reuse an increasing amount of research outputs across borders and disciplines by enhancing the technologies underpinning discoverability and interoperability.
- Research software becomes increasingly FAIR.
- A robust pan-European network of infrastructures for software source code.

B.1.3   Stage 1/Objective 3

Establish a sustainable and federated infrastructure enabling open sharing of scientific results

Priority activities:

- Ensure that domain-specific user environments are included in EOSC (L1+L2+L3).
- Support research communities to develop and adopt domain-specific standards where required (L1+L2).
- Develop common standards and framework for minimum metadata (L1).
- Build a common digital search to enable discovery and exchange across federated repositories (L1).
- Build structures to ensure effective metadata support and coordination for research communities (L1+L2).
- Liaise internationally to develop a global cooperation framework for Open Science infrastructures (L1+L2).
- Define models for availability and costing of services across borders (L1+L2).

Expected outcomes:

- A fully operational, secure cloud-based EOSC infrastructure technology layer.
- A robust pan-European network of infrastructures for sharing and reusing research results.
- A performant and functional MVE, enabling Open Science and adding value for EOSC stakeholders.
- A comprehensive toolbox of common technical infrastructure standards and technical service provision standards available within the network and beyond.
- Institutional and national repositories are federated with common search and access.
- Use of federated research infrastructures is incentivised across all parts of the research landscape, from university researchers to large central facilities.
- Exploitation of a web of open scientific data and services enables efficient impactful science and new ways of working.
- Researchers have improved access to research outputs from across the world through a global web of (open) scientific data and services.
- Publications, data and software repositories are sustainable and trusted through certification.

Considerations based on the recommendations from the EOSC Executive Board Working Groups are as follows:

Persistent identifiers and common AAI are areas that require a clear federated solution and that, together with metadata and ontologies, enable the technologies of the essential common
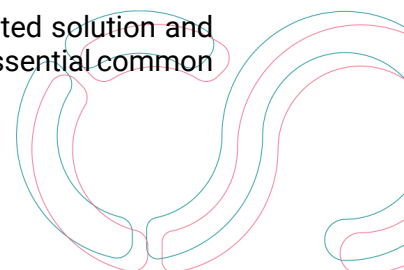
search. For EOSC to be successful, a data infrastructure is needed that utilises performant technology layers to deliver both persistence to research data and FAIR digital objects, and services that allow researchers to benefit from FAIR data and Open Science.

The ongoing development of infrastructure and technology (such as AAI) should build on the activities undertaken during H2020. Specific attention to software and APIs is essential to deliver on the principle of machine actionability. The consultation responses have prioritised machine interoperability for a later stage. Essential for Stage 1 is a low-level development of standards and sharing to enable future exploitation of research results.

Coordination and a multi-faceted approach to research-resource persistence for different providers is needed to ensure that research results are available in the long term. This is especially important for research-data generators that are not resourced for long-term storage of large volumes of data. In many cases, this multi-faceted approach is mirrored by the multi-faceted nature of research and research infrastructures in Europe.
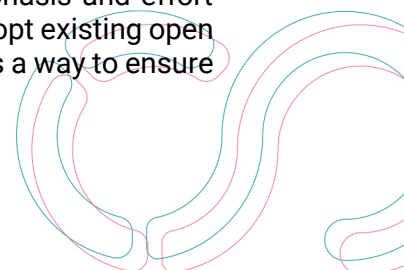
B.2   Priorities Stage 2 (2023-2024)

The priorities for 2023-2024 are drawn from results from projects (being) executed on the basis of the 2021-2022 roadmap and consolidated responses from the public consultation. They focus on embedding as much as possible the foundational governance structures and EOSC technical infrastructure achieved in 2021-2022. The consultation suggested a number of priority areas of work which have been agreed by the EOSC Association Board and are explained in the paragraphs below, namely:

- Business / resourcing models
- EOSC-Core development and onboarding procedures
- Interoperability and data search
- Data quality
- Member State engagement
- Recognition and rewards

The key priority in the 2023-2024 period is to identify and test a number of funding models to ensure the EOSC-Core and EOSC-Exchange are viable over the long-term. Each component will be funded differently; the Core will be supported in a common manner by public funds, while the large variety of services in the Exchange will be resourced using different mechanisms. The EOSC-Exchange must accommodate the range of funding models in use by service providers who wish to federate into EOSC. Such funding models could include subscriptions, licence payments or other pay-per-use models, as well as services that are made freely available. In addition, the Exchange should enable publicly-funded infrastructures such as data repositories and computational resources, to be used in a cross-border context with the possibility of recovering costs. This area of work is complex and it is critical to prioritise efforts here in order to clarify the value proposition for any service provider to federate into EOSC.

Associated with the business and funding models, emphasis is placed on solidifying the EOSC-Core and onboarding procedures. It must be clear how to federate into EOSC and to have efficient processes to enable RIs and other service providers to expose their data and services via EOSC. There are parallel developments here in the 2023-2024 period since the EOSC Future project is scheduled to end in September 2023 and the public tender for the EOSC procurement call will likely be awarded in the first half of 2023.

The EOSC-Core addresses interoperability as one of its four key components and since this is the glue that enables the federation of data and services, much more emphasis and effort need to be directed to these activities. The implementation projects must adopt existing open standards, open interfaces, open APIs, and promote the use of crosswalks as a way to ensure

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 174 –

an interoperable and functional EOSC-Core. The EOSC-Core functionality also includes a common search for content (i.e. data, software and other resources) held within EOSC. This must be improved to ensure value is demonstrated to research communities and that greater reuse of research outputs is enabled. The search functionality should adopt accepted standards (e.g. DDI-CDI, DCAT, schema.org) and support crosswalks to discipline-specific metadata standards that are in use by the RIs and the scientific communities. In order for this to succeed, community engagement on agreeing implementation approaches and the role for the EOSC Task Forces in steering what is happening within projects is paramount.

The EC emphasised data quality as a new area in the forthcoming work programme, in line with EOSC Association suggestions. While policy drivers push for research outputs to be Open and FAIR, content that is readily available and reusable is useless if it is of poor quality. Work is needed to define common data quality indicators for EOSC and to implement these.

Many questions were asked about the resourcing of EOSC. Much investment has been made by the EC into core infrastructure and services, but given that so much research infrastructure is funded nationally, greater engagement of the Member States and Associated Countries is also required. Workable business models are a prerequisite however for Member States to co-invest and encourage or require national infrastructure and services to federate into EOSC.

Another area of priority is that of recognition and rewards. The way we evaluate researchers has been changing towards rewarding Open Science, but this needs to intensify. Incentivising open science practices needs to be made the norm. Beyond the EC, other funders, and research institutes will have to play a huge and important role in this.

In the following three sections, priorities are structured according to the objectives listed in this SRIA. For each objective a series of priority activities and expected outcomes is specified, as well as contextual information to explain plans for implementation or potential issues that need to be highlighted. The last section of the chapter elaborates on how to build on the Horizon Europe 2021-2022 call.

B.2.1 Stage 2/objective 1 – Open Science as the new normal

Objective 1 states: Ensure that Open Science practises and skills are rewarded and taught, becoming the 'new normal'.

Below are the priority activities needed to enact this change in the 2023-2024 period at European, National and Institutional level.

B.2.1.1 European level priorities

- Include Open Science principles in EU data policy and legislation, as has been done by incorporating access to data in the Open Data Directive. The Open Data Directive shall be transposed into Member State law. This is an important step towards broader legal interoperability of resources.
- Liaise with Horizon Europe projects and other key EOSC projects to coordinate activities jointly leading to an MVE.
- Manage and evolve the EOSC Observatory to record statistics on investments, policies, research outputs, open science skills, and infrastructure capacities being planned and delivered via Member States, Associated Countries, and EOSC Association members.
- Clarify the overall rules for onboarding into the Minimum Viable EOSC and produce simple manuals for service providers to follow. Also align the existing procedures for onboarding with the EOSC Rules of Participation.

- Provide support for users from different target groups to federate their digital environments with those of other national providers, horizontal and thematic RIs, and e-Infrastructures.
- Provide ample opportunities for the community to learn about, be consulted on, and engage in EOSC developments under the leadership and coordination of the EOSC Association.
- Widen and deepen the possible EOSC user base via collaboration with various relevant stakeholders.
- Focus on engaging research communities to increase their participation in EOSC.
- Continue to support the development of curricula for Open Science and FAIR skills and align these at a European level to ensure a base-level standard is reached across the European research community.
- Support the development of networks for data stewards, research software engineers, and Open Science communities at European level.
- Stimulate the definition of measures for researcher assessment to help create better rewards and recognition systems for quality Open Science. Although it has a broader remit than recognition and rewards for Open Science, the EC's research assessment group is a useful vehicle for progressing this activity.[46]
- Support the creation and dissemination of shared domain specific standards and evaluation frameworks and how they can be recognised, e.g. via open badges.
- Develop the EOSC visual identity and define the conditions for its reuse by stakeholders, based on the new EOSC logo.
- Systematically review and evaluate the effect of new Open Science metrics after their introduction, checking they are incentivising good practice and mitigating any unintended effects.

B.2.1.2   National level priorities

- Demonstrate national commitment to EOSC by actively contributing to activities of the EOSC Steering Board, the Tripartite Collaboration, and the Council of National Open Science Coordination, including the sharing of best-practices.
- Discuss long-term funding commitments and the time after Horizon Europe (2028 onwards) on a national level.
- Harmonise national Open Science monitoring and KPIs to feed into the EOSC Observatory activities.
- Focus on engaging research communities to increase their participation in EOSC.
- Coordinate awareness raising and uptake of EOSC in collaboration with the designated EOSC Association Mandated Organisation or other EOSC national structures, national funders, and RIs to interlink the national Open Science ecosystems to EOSC.
- Develop national policies and funding streams to support the development of EOSC and incentivise national RIs to federate into EOSC.
- Support regional cooperation across countries (e.g. the Nordics) learning from regional projects to strengthen national engagement.
- Incentivise and encourage smaller community and institutional infrastructures in their growth towards ensuring FAIR data and providing long term access.
- Establish the cost of data management, data stewardship and maintenance of research outputs (including software) as eligible within national funding schemes.

---

46 See: https://ec.europa.eu/info/news/process-towards-agreement-reforming-research-assessment-2022-jan-18_en
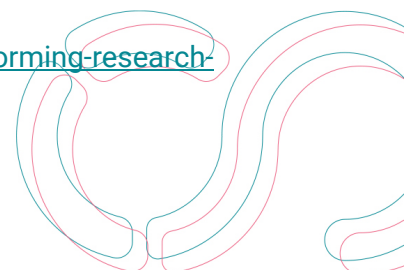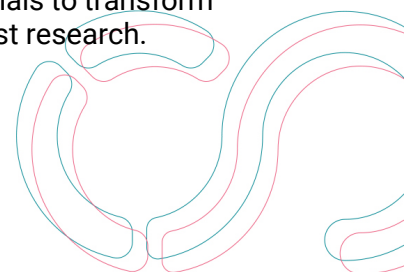
- Develop curricula for Open Science and FAIR data skills in line with aligned European curricula and encourage the inclusion of these as a core element in research programmes.
- Support training of data stewards and research software engineers, increasing their numbers in the European scientific context.
- Offer Open Science and FAIR skills training to policy makers, funders and other relevant stakeholders, like evaluators to promote alignment in evaluation criteria.
- Establish policy and coordinate implementation frameworks that measure and reward FAIR and Open Science practises.
- Recognise research data and software properly as outputs for research in national evaluation of science.

B.2.1.3 Institutional level priorities

- Focus on engaging research communities to increase their participation in EOSC through, e.g. Rectors' conferences, EOSC reference points, and national and regional tripartite events.
- Incentivise and encourage institutional infrastructures to share best practises on how to channel a broad span of researcher needs into the further development of EOSC.
- Offer core Open Science and FAIR training to researchers at all levels and recognise these skills as important for professional development. Such training could be incorporated into doctoral training programmes together with research integrity.
- Embed EOSC information into institutional research programmes to support uptake by research communities.
- Foster awareness of local FAIR-enabling practices and the value of long-term preservation of data and metadata amongst research communities.
- Support professional development programmes to ensure research support staff have the required Open Science and FAIR skills.
- Engage in networks for data stewards, research software engineers, and Open Science communities to share models and harmonise best practises.
- Provide meaningful career pathways that allow research software engineers and data stewards to progress in academia from postdoc level and beyond.
- Support the development of an open science policy and associated implementation groups to create an open institutional culture.
- Adjust research review mechanisms and researcher incentives to ensure FAIR research outputs and Open Science are appropriately recognised and rewarded.

B.2.1.4 Expected outcomes

- A growing number of stakeholders actively engaged in EOSC, representative of the various groups and disciplines at European, national, and institutional level, that channels EOSC information to the community, supports alignment, and helps to ensure EOSC developments meet real-world use cases.
- FAIR, persistent and trusted research outputs and services from a growing range of research communities are available in EOSC.
- Validated statistics on national and institutional contributions to EOSC are available, supporting future strategy development.
- Researchers, including doctoral candidates, are equipped with the appropriate Open Science and FAIR skills and are supported by highly skilled professionals to transform the way they use research outputs, leading to better-quality and robust research.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 177 –

- Researchers are supported by incentives for career advancement that include open science principles.
- Data Stewards, Data Scientists, Research Software Engineers and other associated professions are established as new job profiles in the science system.
- A higher proportion of research software engineers and other highly-sought staff remain in academia as their skillset is appropriately recognised and rewarded.
- Data stewards and researchers can network at a European and national level to access peers for support, consolidating good practices.
- After having established the new EOSC visual identity, the rules for using the EOSC brand are codified, communicated, and made available for use by the EOSC Community and Projects.
- Open Science becomes a national priority and the public national and regional funding agencies use Open Science principles as a funding condition for research activities.

B.2.1.5   Clarification text

Coordination across EOSC implementation projects is paramount to ensure developments result in a coherent, operational platform. The EOSC Association will play a key role here. Hence, all new Horizon Europe projects should be encouraged to collaborate and engage with the EOSC Association and to contribute to this SRIA. As coordinator of the EOSC Focus project, the Association will oversee these activities by facilitating liaison and alignment among the EOSC projects and the relevant stakeholders acting at European, national and institutional level by setting up an EOSC community platform and hosting regular coordination fora. These activities will provide the foundations for the planned liaison, necessary to effectively co-create EOSC, ensuring that the needs of research communities are appropriately addressed and to widen EOSC engagement. Through the EOSC Focus project, the EOSC Partnership monitoring function will also be continued. It is expected that the EOSC Association will become owner of the monitoring data.

The role of national structures is also key to incentivise RIs to federate into EOSC and support smaller communities on their journey towards FAIR and Open Science practices. While much work has already been done to align EU and national policies with Open Science principles, actions are needed to incentivise the adoption of these practices. This will include support for development of standards and a network of FAIR-enabling trusted repositories.

Following on from the definition of the Rules of Participation by the EOSC Executive Board RoP Working Group in 2021, a review should be undertaken to ensure the onboarding procedures implemented by the projects align with these Rules and are clear and fit for purpose. This activity should evaluate whether the expected number of national RIs and community services have been federated into EOSC to provide a solid platform from which to build. There should also be clear added value for service providers (e.g. integration into EOSC AAI, access to monitoring, etc.) and those offering data and code (e.g. increased discovery, ability to combine data from different domains) to federate into EOSC. Actions at European level and from Member States are needed to continue to incentivize and encourage participation amongst service providers and research communities to build the EOSC federation.

Since a growing number of courses are emerging for FAIR and Open Science, priorities in this phase of implementation focus on aligning curricula to ensure a base-level competency across the European research community. While the European level is responsible for coordinating activities and providing networks to enable data stewards, research software engineers, and Open Science communities to engage with peers, much activity focuses on the national and institutional level. It is within these bounds that core competencies can be included in research programmes to ensure a base level of skills. Moreover, national funders

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 178 –

and research organisations operate the recognition and rewards programmes and set grant conditions which will be critical to incentivising researchers and support staff and to providing appropriate career paths to retain their expertise. Systematic reviews of the new metrics and incentives applied are also needed to ensure they have the desired effect. European level coordination is also needed, since fragmentation is still a serious problem when it comes to collecting and using new metrics.

Data stewards and research software engineers are largely new funded roles at the RPOs. Researchers may transition into these new roles; hence it is important to determine possible career paths for them. There is also a need to increase the number of professional support staff, hence the need to specify those job roles and offer aligned curricula and support for a to ensure they are recognised appropriately across Europe.
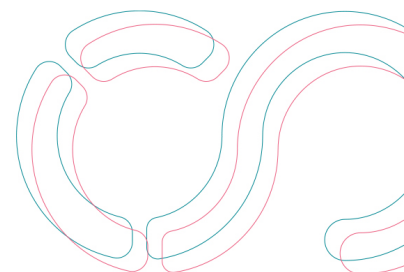
B.2.2   Stage 2/objective 2 – Definition and development of standards and tools

Objective 2 states: Enable the definition of standards, and the development of tools and services, to allow researchers to find, access, reuse, and combine results.

Below are the priority activities needed to enact this change in the 2023-2024 period at European, National and Institutional level.

B.2.2.1   European level priorities

- Embed the process by which community standards are endorsed for use within EOSC, developing multiple registries to address metadata standards and other semantic artefacts.
- Further enhance the EOSC Interoperability Framework to foster composability between services and resources, facilitating the creation of innovative solutions that enable research use cases and foster collaboration between diverse research communities.
- Coordinate access to resources from public authorities within EOSC, supporting work conducted at a national level to promote appropriate sharing of sensitive data and enabling European alignment and cross-border access.
- Develop and maintain open interfaces, alignments, crosswalks, and APIs that enable interoperability and foster adoption of EOSC. Research communities must be supported financially to undertake such work to facilitate interoperability.
- Establish clear interoperability guidelines towards all Data Spaces, foreseen in the European Data Strategy.
- Identify research communities and disciplines that lack standards and promote their development.
- Define, disseminate and implement common measures of quality to be applied to data, software, and other semantic artefacts, so the content within EOSC can be evaluated by users.
- Monitor technology developments and adopt new standards and software languages, where appropriate, to avoid that components of the EOSC-Core become outdated.
- Integrate EOSC services into EOSC-wide AAI systems. Core service integration should be mandatory whilst integration of EOSC-Exchange services should be strongly encouraged.
- Collaborate with national and thematic competence centres to support adoption of the EOSC-Exchange amongst service providers and researchers.

- Ensure the quality, long-term availability, and usability of the services in EOSC-Exchange by means of detailed specifications, standards adoption, assessment, certifications, and monitoring.
- Evaluate and, as necessary, develop trusted environments for managing, preserving, and sharing sensitive data, recognising work already in place at national level.
- Develop a global scalable PID-resolving infrastructure to support the broad ecosystem of PIDs used. The PID-resolving infrastructure will re-use existing mechanisms as much as possible.
- Support a network of FAIR-enabling trusted repositories for EOSC, including code repositories, that provide and adopt interoperable tools and aligned services. This network should act as a community platform for shared experience and practices that engages with and uplifts wider data services.
- Develop multilingual tools and search to facilitate uptake by researchers across Europe using their native languages.
- Promote the alignment of vocabularies not only in terms of disciplines but also in terms of cultural and linguistic coverage.

B.2.2.2   National level priorities

- Encourage and incentivise effective documentation and sharing of all outputs relevant to research, including software and source code by defining dedicated funding mechanisms to support it.
- Provide national RIs and RPOs with sustainable funding streams related to EOSC, domain, and cross-domain initiatives.
- Create access to resources from public authorities within EOSC, engaging with national legal and ethical experts to promote appropriate sharing of sensitive data at a European level and alignment to facilitate cross-border access.
- Support and incentivize the development, maintenance and adoption of open standards and APIs to enable resource composability and to achieve interoperability across communities including in the private sector.
- Base EOSC interoperability guidelines on existing community-accepted standards and encourage research communities and national RIs to, where relevant, develop crosswalks (i.e. mappings of relationships between fields of different metadata schemas) and alignments with others.
- Incentivise assessment and open certification of data repositories to ensure they enable FAIR outputs, are trustworthy and sustainable, and provide an interoperability layer across the European-wide network.
- Encourage and promote the different cultural dimensions and native languages in the development of tools and standards to maintain the diversity of European communities.
- Contribute to the alignment of national vocabularies in all the scientific disciplines with other languages and prepare their sustainability through regular updates.

B.2.2.3   Institutional level priorities

- Raise awareness of infrastructure to support the creation, management, and sharing of research data and software, encouraging adoption.
- Support research communities to adopt both general and domain-specific standards to increase adoption of FAIR practices and reuse.
- Provide guidance and assistance to small repositories to enable sustainability and interoperability through standards compliance, assessment, and certification.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 180 –

- Advocate for research communities to support and test the crosswalks and APIs being developed to support interoperability across and inside research disciplines and repositories.
- Support the adoption of common quality requirements for FAIR research outputs to be implemented at EOSC level, including through institutional, disciplinary, and other data repositories.
- Implement tools to plan, track, and assess scientific knowledge production (e.g. by connecting science knowledge graphs, Research Information Management Systems, Data Management Plans (DMPs), and metrics) based on open definitions, standards, and models.

B.2.2.4 Expected outcomes

- A wide range of disciplinary standards are endorsed and adopted in EOSC and the associated interoperability developments support a growing range of use cases on multi-disciplinary reuse of data and services.
- The adoption of metadata standards increases the proportion of FAIR data and the range of disciplinary resources that can be discovered via an EOSC search interface.
- Services offered within the EOSC-Exchange are reliable and of high quality, growing trust and adoption of EOSC.
- Infrastructure for sensitive data sharing is provided, via which a comprehensive collection of data from public authorities is made available for research reuse.
- A network of trusted data repositories is available and repository managers can identify clear benefits (e.g. efficiency and streamlined operations) of being part of EOSC.
- FAIR-enabling trustworthy repositories act as key nodes in the full data lifecycle network of institutional, national, and European practice.
- Transparency of (meta)data curation and preservation practice across FAIR-enabling data services including trustworthy repositories are mandated and incentivised.
- PIDs are used systematically and links between them are accessible via knowledge graphs to visualise and track research network activity.
- Common requirements for the quality of research software and data are agreed and these measures are used to define quality indicators and standards accepted by research communities.
- Infrastructures for research software and source code are promoted at national and institutional level, increasing the amount of software and code accessible via EOSC.
- Sustainability is secured by national (state and public) institutions and resources, not based mainly on participation in EU funding instruments (e.g. the HE Programmes).
- Shared best practices on the alignment of vocabularies and their sustainability are available and promoted.

B.2.2.5 Clarification text

One area of specific interest in the 2023-2024 period is extending the content available within EOSC to sensitive data from public authorities. Trusted environments for managing and sharing such data will be offered at the European level and will be supported by national level actions to facilitate uptake. A standard set of methods should be developed that can effectively enable data sharing and processing of sensitive content. These should be general enough to be applicable within the same country and in cross-border scenarios. Activities should be cross-articulated with the Data Spaces foreseen in the European Data Strategy to strengthen integration with them.
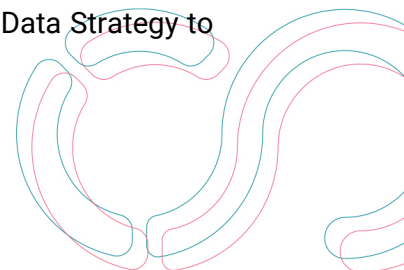
FAIR-enabling repositories for data and code should be brought together in a network to support the development and adoption of standards, tools, and practices that promote interoperability and consistency of services. By sharing tools and practices across the network of repositories, less developed services can be supported and advanced. Such activities could facilitate automated metadata uplift and offer multilingual knowledge management tools and community curation mechanisms. Certification of repositories is also critical to ensure the services are trustworthy and can be relied upon to maintain the content available via EOSC. Work should also be undertaken to clarify how assessment frameworks for FAIR research outputs and FAIR-enabling trusted repositories work together. This could take into account disciplinary and generic initiatives, but also cultural and native language practices.

Mechanisms to assess the quality of data, code, and other semantic artefacts within the frame of EOSC are also needed to ensure that all stakeholders have a clear understanding of the adequacy of the information they are accessing for their scientific application. A data quality framework needs to be defined, which includes specifying quality indicators, exploring ways to disseminate quality information, providing clear standards to follow and considering potential certification schemes. Guidelines initiated in the EOSC Association Task Forces will be used to steer projects implementing these activities. Institutions will also play a key role to support the adoption of common quality requirements to be implemented within EOSC.

This second phase of work will also focus on growing the RIs and communities connected to EOSC and enhancing the interoperability across them. Collaboration with national and thematic competence centres is envisaged to increase the range of services within EOSC. Inputs from Member States and institutions are needed here to incentivise communities to engage and embed EOSC within a wider user group.

In terms of interoperability, the focus is on endorsing and, where necessary, extending existing community standards rather than inventing new ones for EOSC. By doing so, the different communities will have an abstraction layer which will make everything easier to interoperate with the EOSC-Core, and to share and compose resources with each other. The standards adopted should be broadly accepted by research communities, mature, well-maintained, and ideally in use globally. It is also critical to develop crosswalks and open APIs to allow different community standards to be integrated. A prioritisation is required for this integration work based on actual scientific use cases, for example the combination of social science and public health data to address the COVID-19 pandemic. It should also be noted that interoperability goes beyond semantics and addresses legal, organisational, and technical issues. For example, machine-actionable licences are needed to clarify rights management, addressing not only who has the permission to reuse content, but also who is responsible for storage, curation and preservation.

B.2.3   Stage 2/objective 3 – Establish a sustainable and federated infrastructure

Objective 3 states: Establish a sustainable and federated infrastructure enabling open sharing of scientific results.

Below are the priority activities needed to enact this change in the 2023-2024 period at European, National and Institutional level.
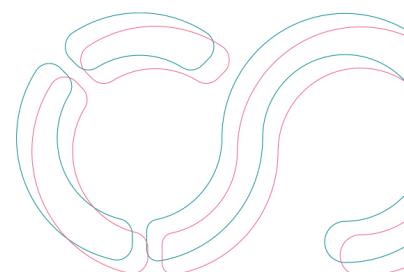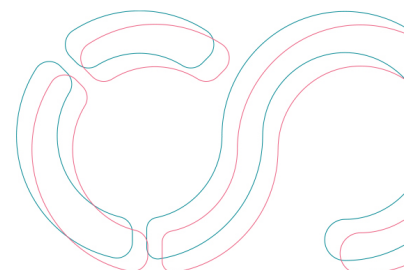
B.2.3.1 European level priorities

- Examine how the EOSC governance could further evolve to more effectively coordinate activities and sustain and grow the MVE.
- Widen the resources accessible for EOSC to cover the full spectrum of research outputs (e.g. models, software, and workflows) and reliable common services (e.g. AAI, accounting, and monitoring).
- Enable researchers across all disciplines and organisations (small and large research infrastructures and less developed organisations) easy access to the federated infrastructure.
- Evolve and further establish the architecture blueprint for the EOSC-Core to enhance the model of the EOSC federated architecture, and as a toolkit to federate research data and enable thematic data spaces in the EOSC-Exchange.
- Ensure the EOSC-Core has well-defined technical, governance, and sustainability plans, supported by assessments of critical elements that need high availability and reliability.
- Continue to evolve an EOSC AAI Federation to enable more communities to connect their AARC Blueprint compatible AAIs for seamless access to resources and services.
- Enhance the federated search functionality for EOSC to improve discovery, tailored recommendations, and reuse of content.
- Ensure a continuous security and personal data protection framework as required by the GDPR and other relevant regulatory frameworks, addressing licensing and IPR.
- Investigate the issues around sustainability models for long-term preservation services, including topics such as the costs of data storage, curation, and ensuring long-term usability.
- Provide federated long-term access and preservation infrastructure for EOSC, which can be utilised by repositories and RIs. This should include provision of services that support reproducibility of preserved results (e.g. workflows and data processing container images) as well as the data itself.
- Identify resourcing models that enable cross-border access and use of existing national infrastructure capacity.
- Trial resourcing models identified in Phase 1 to provide the availability of resources in the MVE, including use and contribution by the wider public and private sectors.
- Continue to liaise internationally to develop a global cooperation framework for Open Science infrastructures also using EC partnerships with other regions such as Latin America and Africa.

B.2.3.2 National level priorities

- Align research and e-Infrastructure strategies at national level and also between the national and EU levels.
- Introduce measures to incentivise RIs and RPOs to adopt appropriate community standards to be federated into EOSC.
- Ensure national investments are directed to national RIs that adopt appropriate community and interoperability standards and can be federated into EOSC.
- Gather information on the financial models implemented by national and European RIs, national research and education networks (NRENs), and communities to understand any constraints on pooling resources and making services available at a European level to develop cost-recovery models that incentivise the participation of institutions.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 183 –

- Continue to harmonise requirements for DMPs (which should address software and other outputs) and encourage the use of machine-actionable DMPs, so that content from DMPs can be more easily aggregated and exchanged with other systems.

B.2.3.3   Institutional level priorities

- Encourage the adoption and use of the federated infrastructure to be provided by EOSC, by sharing examples of how research use cases have been addressed and signposting relevant resources and services when engaging with researchers.
- Encourage the use of domain specific and existing RIs.
- Define and implement procedures to select data, software, and other research outputs of long-term value to be preserved via EOSC.
- Adopt machine-actionable DMP tools and openly share institutional DMP data, where possible, in pursuit of establishing best practices and increasing FAIR alignment. DMPs should be interlinked to trusted repositories and other institutional internal systems for managing projects.

B.2.3.4   Expected outcomes

- Options for a future potential legal structure for EOSC such that the MVE can be sustained and further developed, by implementing adequate Rules of Participation.
- Sustainability of EOSC is secured by co-investment at European and national levels with cost-recovery models that incentivize participation of institutions and cross-border activity.
- An increasingly broad range of content is available via EOSC and usage statistics demonstrate increased discovery and reuse,
- The operation of EOSC-Core services (such as AAI) and federated infrastructure model enable a wide range of communities to connect to EOSC.
- Consolidation and further enhancement of the EOSC-Core, strengthening the federation aspects and integration.
- Long-term preservation infrastructure is offered within EOSC, enabling selected research outputs to be retained and actively preserved for at least 10 years.
- Researchers are supported to select data, code, and other outputs of long-term value.
- Resourcing models to sustain the MVE are trialled and advanced, and a core set of storage and compute resources are offered cross-border.
- Data from DMPs is reused to share good practices, increasing adoption of FAIR, and making the management of research projects more effective.

B.2.3.5   Clarification text

Within the area of long-term preservation, it is recommended to undertake further studies that can elaborate on sustainability models that go beyond individual organisations. This should investigate the costs and facilities needed, review which file and data formats are appropriate to support, consider different models of implementing preservation services and assess the relative cost/benefit ratio of each, and make recommendations on the minimum retention and preservation periods for different types of data and research software. Decision-making mechanisms to select which data to preserve are also critical and should be elaborated and implemented in conjunction with data services, research communities, and RPOs. Recommendations on long-term preservation services should be implemented by providing effective and cost-efficient long-term storage infrastructure at European level to which different repository types can connect.

![eosc logo]

A critical challenge in developing EOSC is to conclude whether it would be better to move beyond project-based partnerships and collaboration models to more sustainable long-term operations, such that resource providers can offer resources to any researcher in Europe and are assured that the resources consumed by researchers from outside their targeted user community are consumed in a financially sustainable way. The EOSC Association Task Force on Financial Sustainability, together with the EOSC Focus project, will explore resourcing models that enable cross-border access and use of services. There is not expected to be a single business model that can support all of EOSC, and the financing for the EOSC-Core will differ from that of the services made available via EOSC-Exchange. The financing models of EOSC-Core should ensure its wide availability and should not be unduly influenced by any single supplier.

The EOSC Partnership must work towards financial models that support federated, cross-border access to existing national capacity to avoid fragmentation and maximise the investments made across Europe. It is likely that the financing model of EOSC will evolve over time and will also be linked to the evolution of the EOSC Association. In the long term, basic infrastructure such as that required to operate the EOSC-Core would be financed by public funds, whilst the EOSC-Exchange would implement a broader range of financial models which are yet to be defined. However, in the initial setup phase, stimulation in both the EOSC-Core and EOSC-Exchange should come not only from European but also from national funding.

There is an overarching focus under this objective on further developing the federated infrastructure and interoperability as an enabler of Open Science practices. Work should continue to enhance and embed the EOSC-Core, enabling a greater range of services and research communities to connect. The common digital search functionality should continue to be enhanced to improve search options and filters. Actions on national and institutional levels will have to support this by ensuring Research Infrastructures and communities adopt the appropriate standards to be federated into EOSC and make resources discoverable and reusable. Where communities lack standards to contribute to EOSC, initiatives should be supported to promote their development. The role of bodies such as the Research Data Alliance are important here to link up with international stakeholders and form the European component of the global cooperation frameworks for Open Science.
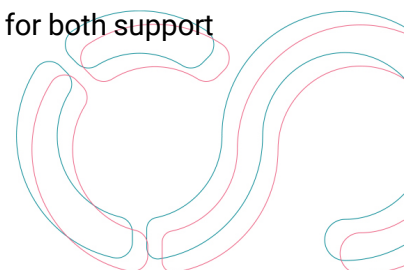
B.3 Priorities Stage 3 (2025-2027)

Please note:

1) Only the priorities for 2025 have been indicated with 2025 the ones for 2026 and 2027 have no indication of the year.
2) The priorities under each objective are not placed in order of value. They are thematically clustered on top. They have been labelled purely for ease of referencing in the consultation.

B.3.1   Stage 3 /Objective 1 – Ensure that Open Science practices and skills are rewarded and taught, becoming the 'new normal'

B.3.1.1   European level priorities

A. Align on core curricula for data stewardship, Open Science and FAIR for both support staff and researchers. (2025)

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 185 –

B. Support the development and sustainability of networks for data stewards, research software engineers, semantic artefact curators and Open Science communities, leveraging on existing national networks and coordinating these at European level to promote the added value of EOSC. (2025)

C. Collect, connect and scale up ongoing good practices and case studies for the research and career assessment of open science practices in collaboration with the Coalition for Advancing Research Assessment (CoARA). (2025)

D. Support and complement the CoARA movement with developing tools and processes for the research and career assessment of open science practices.

E. Pilot indicators/metrics at the European level for the research and career assessment of open science practices and disseminate them Europe-wide.

F. Support and embed appraisal and preservation procedures for long-term data preservation and monitor their value and costs. Practices should be transparent, aligned were possible and interoperable (i.e. through machine actionable rights management).

G. A dedicated provision for supporting the EOSC Partnership in its concertation actions. This covers stakeholder engagement, coordination, governance and monitoring. A cascading grant mechanism should support widening activities with under-represented countries and national competence centres, amongst other things.

B.3.1.2 National level priorities

H. Support the implementation of aligned European curricula for data stewardship and encourage the inclusion of these as a core element in research programmes. The establishment of proper career paths for research support professionals is also needed.

I. Establish the cost of data management, data stewardship, maintenance and preservation of research outputs (including software and semantic artefacts) as eligible within national funding schemes.

J. Leverage existing national Competence Centres and strengthen their participation in coordination networks.

K. Use cases from the public sector, citizen science and industry should be encouraged in the next period in parallel to the research sector, whilst ensuring that services developed remain under the control of academics. Also use cases running at national or regional level are interesting. Emphasis should be to support uptake (2025).

B.3.1.3 Institutional level priorities

L. Offer codified and certificated core Open Science, FAIR and CARE training to researchers and research support units at all levels and recognise the skills as part for their professional development through different certified means.

M. Support professional development programmes to ensure research support staff have the required data stewardship skills.

N. Engage in competence centres and networks for data stewards, research software engineers, semantic artifact curators and Open Science communities to share models and harmonise best practises.

O. Adjust, follow up, and evaluate research review mechanisms to ensure FAIR research outputs and Open Science are appropriately recognised and rewarded.
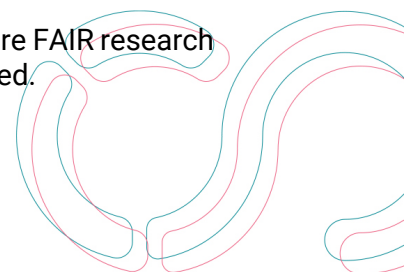
P. Define and implement training and procedures to select data, software, and other research outputs that retain the value necessary to be preserved.

B.3.1.4   Expected outcomes

Q. Research assessment in Europe is more harmonised, supporting Open Science as the new normal.

R. Researchers and support staff have appropriate data stewardship skills and access to peer networks, semantic artefacts and research software so a base-level of knowledge is reached across the European academic community, including countries where skill gaps exist.

S. Research decision makers recognise the value of data stewardship and routinely include this as an element in grant proposals and the research process.

T. Researchers are supported by assessment procedures and metrics (fully recognized by their organisations) that incentivise open science practices.

U. Research communities proactively engage in decisions about which data, software and other outputs to preserve for the long-term and guidelines are available per community.

V. A dedicated provision for supporting the EOSC Partnership and its transition into the new phase of governance under the next Framework Programme. This also covers stakeholder engagement, coordination, governance and monitoring. A cascading grant mechanism should support widening activities with under-represented countries/communities and national competence centres, amongst other things.

W. EOSC Federation is widened resulting in access to content from key data spaces, broader uptake from research performing organisations and new initiatives in previously under-represented countries and communities.

X. The value of EOSC is demonstrated via different (sectoral) use cases, increasing the range of applications supported and growing the user base.
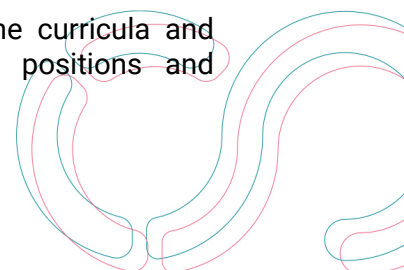
B.3.1.5   Clarification text

Use case - Case study – good/best practice
The description of what is meant by these terms is being discussed in the frame of the tripartite collaboration, which consists of the European Commission, the EOSC Steering Board (Expert Group of the EC with members from the Member States and countries Associated to Horizon Europe), and the EOSC Association (having Research performing, Research funding and Research service providing members). For now, we suggest using the term **good practices** for all of the examples that might be valuable for other parties in the ecosystem.

Data stewardship
Codified and certificated Data stewardship curricula for both support staff and researchers should be aligned as much as possible at the European level to ensure a base standard of Open Science and FAIR practice is reached. The curricula should build on existing work conducted by EOSC projects, institutions and the global community, for example in the context of RDA Interest Groups. In particular, we recommend an emphasis on semantic artefact curation to bridge between domain specific and domain agnostic skills. There is a lack of awareness and skills when it comes to the curation of semantic artefacts that future data stewardship curricula should address.

Both Member States and institutions have a responsibility to roll out the curricula and recognise the importance of these skills, particularly by supporting positions and

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 187 –

opportunities for career development for professional data stewards. Outreach programmes should target young researchers by connecting at the regional level. In addition to the curricula, professional networks for data stewards and a recognition of the associated costs as eligible in national grant schemes are needed to incentivise the change of culture.

Competence Centres

The Centres of Competence (CC) represent one possible point of reference in a specific Country/Region/Theme/Science Cluster. They strengthen their infrastructures and these in turn strengthen the European structure. Like the already existing national HPC competence centres[2]. In any case, ongoing projects should be linked to these CC's. Existing (data) competence centres play an enabling role in advocating Open Science for various actors, in accordance with the Digital Skills Agenda and the EOSC. As well to advance the offer of support, counselling and training locally. Building such CC's are key to knowledge transfer on Open Science, FAIR and CARE in the EOSC context. While the structure, operational mode and organisation of Competence Centres may vary very widely, they are usually associated with excellence, training, interdisciplinarity, standardisation and a collaborative approach that ensures harmonisation and alignment at European level.

Research assessment

In order to support the transition to Open Science and FAIR practice, the work done by researchers relevant for this transition need to be recognized and rewarded. New measures need to be defined and tested to assess uptake and ensure they lead to the desired culture change. A recent major development in the landscape is the coalition for advancing research assessment (CoARA). This initiative will entail a period of experimentation by coalition members for several years. This experimentation needs to be supported by further EOSC projects aiming at development, piloting, implementation, and medium-to-long-term evaluation of new methods for research assessment. Research funding and performing organisations, whether CoARA members or not, should include Open Science activities in research review mechanisms.

Data preservation

The community needs to be made more aware about the roles and responsibilities for long-term preservation as part of the overall goal to make this the 'new normal'. It is needed to embed preservation systems and appraisal practices and to make them machine actionable for efficiency, interoperability and transparency. Disciplinary expectations for data and metadata preservation must be combined for cross-disciplinary needs. Transparency of appraisal ensures broad input on decisions to changing levels of preservation or transfer to alternate repositories. Retention and preservation should be transparent in digital object metadata as well as at the repository level. Semantic artefacts - as key knowledge organisation tools - also require special attention concerning long-term preservation.

Engagement and widening

Expanding on use cases should be a key emphasis in this period to support uptake and widening of EOSC. The Minimum Viable EOSC (MVE) should be tested by different communities, including the data spaces and industry to validate applicability beyond academic researchers. Member State involvement is key to incentivise other sectors to trial EOSC.

A dedicated provision is also recommended to assist the EOSC Partnership in evolving its work with different stakeholders, specifically research performing organisations, the data spaces and countries less engaged in EOSC. The actions could be supported by cascading grants to facilitate uptake where it is lacking, or dedicated funding lines to support the countries and regions. Dedicated calls like WIDERA-CSA or other types of instruments are expected. Many countries have existing national open science strategies or initiatives, and these should be cooperated with to align with EOSC. Activities in the dedicated provision
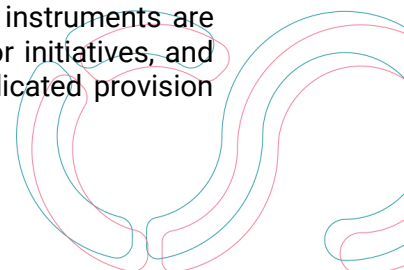
should address topics such as competence centres, the transition to a new governance structure, service delivery models and monitoring of Open Science activities via the EOSC Observatory. The EOSC Association should be properly financed over time as part of sustainability plans when adopting responsibility for project outputs.

B.3.2 Stage 3 /Objective 2 – Enable the definition of standards, and the development of tools and services, to allow researchers to find, access, reuse and combine results

B.3.2.1 European level priorities

A. Promote and sustain the use of Persistent Identifiers (PIDs) that are more and more common practice. Support activities where PID usage is not yet a common practice.

B. Advance technical specifications to support digital objects to be FAIR-by-design and evaluate viable options to adopt the concept of FAIR digital objects for operational data services. (2025)

C. Continue to invest in the creation, adoption and governance of community-based metadata and data standards in adequacy with latest EOSC developments towards a minimum common metadata schema - to support the discovery, accessibility, interoperability and reuse of research data and software.

D. The EOSC Interoperability Framework should facilitate the composition of EOSC data and services on the basis of (rich) metadata, ensuring responsible data exchange. For this it is needed to continue developing and maintaining open interfaces, alignments, guidelines, crosswalks, and APIs that enable interoperability and foster adoption of EOSC.

E. Evaluate the viable options to provide European and international researchers with a platform for research software through development, preservation, citation and reuse, most likely as a federated network of research software repositories, and other research software artefacts following the FAIR principles, with EOSC-level guidance on good practice use and quality standards.

F. Focus on domain-specific FAIR metrics and data quality constraints (specifically for the semantic interoperability of these) by engaging with research clusters to apply FAIR in their contexts. (2025)

G. Support the establishment of a widely representative governance body to incentivise more FAIR artefacts and elaborate more coherent FAIR assessment by interacting with global initiatives.

B.3.2.2 National level priorities

H. Promote and sustain the use of PIDs that are already common practice. Support activities where PID usage is not yet a common practice.

I. Integrate widely used and adopted PIDs into national services. Incentivise usage of PID technologies being developed in EOSC.

J. Support and incentivize the use, maintenance and adoption of open standards and APIs to enable resource composability and to achieve interoperability across communities, including in the private sector.
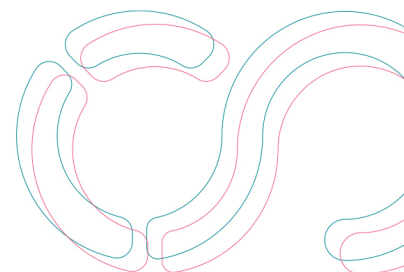
B.3.2.3 Institutional level priorities

K. Integrate widely used and adopted PIDs into institutional services and incentivise usage of PID technologies being developed in EOSC.

L. Support research communities to adopt both general and domain-specific standards to increase adoption of FAIR practices and develop plans to facilitate reuse.

M. Develop procedures and support for publishing semantic artefacts through institutional repositories or specific thematic repositories.

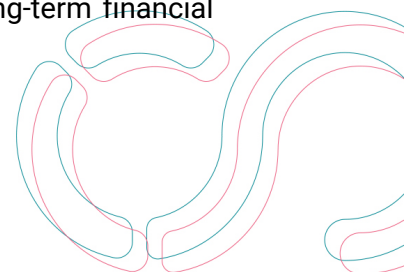N. Encourage sharing of quality assured software through institutional or thematic repositories.

B.3.2.4 Expected outcomes

O. A clear increase in the usage of PIDs and exploitation of links between them via typing, contextualisation and scientific knowledge graphs to visualise and track research activity.

P. The integration of PIDs into data services is enabled by supporting the interoperability of PID systems. This reduces the current overhead and confusion caused by re-registering the same research outcome at different PID providers.

Q. Existing Infrastructures for research software and source code are promoted at national and institutional level, increasing the amount of software and code accessible via EOSC.

R. Metrics are defined to measure FAIR on a community level and evaluator tools enable reliable and trusted comparisons between these.

S. Ensure the use of existing or community driven recommendations via assessments regarding standardisation and data quality.

T. Cross-domain reuse of data increases due to higher levels of standards adoption and alignments/crosswalks between these.

U. Effective quality control mechanisms for semantic artefacts are in place.

B.3.2.5 Clarification text

PIDs

Persistent Identifiers (PIDs) are a critical component of the underlying infrastructure to allow resources to be accurately identified, accessed and linked to other objects. Many types of PID are already in common usage and this should be sustained via support at the European and national level, encouraging adoption by communities that are not already active. The integration of different actors in Europe supporting PID services should also be coordinated for sustainability. Support should be directed towards national and institutional services, enabling their adoption of PIDs to ensure content federated into EOSC is well contextualised. Services being supported at the European level such as PID Meta-Resolver, Data Type Registries, PID graph technologies and the PID Compliance Assessment Toolkit should also be incentivised by Member States and institutions to promote uptake and researchers should be encouraged to use PIDs for the whole research lifecycle, as well as for their research activities. The interoperability of PID systems should also be encouraged as objects often get assigned two or more different PIDs since few interactions are possible between the providers. As noted under objective 3, a long-term sustainability plan should also be developed within EOSC for PIDs to help the research community understand their long-term financial commitment related with PID adoption.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- 190 –

Interoperability

The absence of quality metadata continues to be a major obstacle to the discovery and reuse of research outputs. As such, continued investment in developing, evaluating and maintaining community-based metadata standards is recommended at the European level and work to encourage adoption is suggested within institutions. The adoption of standards is also critical to support interoperability and accessibility. Work related to an EOSC Interoperability Framework is ongoing in EOSC Future and relevant parallel activities are planned in the SIMPL procurement for the Common European Data Spaces. The EOSC Interoperability Framework should be compatible with the SIMPL guidelines for the Data Spaces, requiring collaboration as both develop. To enable interoperability, work on the development of open interfaces, alignments, crosswalks and APIs must continue, and research communities need to be encouraged and supported to dedicate time to this important work. Interoperability should cover all the outputs of research as well as the services, enabling composability of resources so research workflows can run smoothly without manual intervention from the user. Research institutions also play a key role in encouraging researchers to adopt relevant standards and implement FAIR.
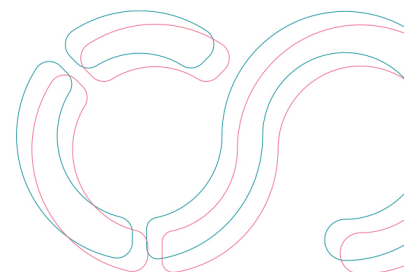
An emphasis has been placed on further support for FAIR digital objects since their implementation encourages the usage of PIDs and agreed metadata schemas, and as such facilitates interoperability. They provide a conceptual and implementation framework to develop scalable cross-disciplinary capabilities, deal with the increasing data volumes and their inherent complexity, build tools that help to increase trust in data, create mechanisms to efficiently operate in the domain of scientific assertions, and promote data interoperability. FAIR digital objects are identified via PIDs and can be regarded as contextualised sub-graphs of general scientific knowledge graphs (SKGs), which allow to account, describe and share everything about particular research.

Source-code repositories

Further support is needed to support the development, archiving, sharing and reuse of research software. At the European level, we need to evaluate the viable options to provide European and international researchers with a platform for software development, sharing and preservation, most likely as a federated network of software repositories. A study should be launched to build a Pan-European software catalogue. Stress should be put on reusability, and the use of standards that facilitate the deployment and uptake of software.

FAIR metrics

Building on the work conducted to define a canonical set of FAIR metrics based on a harmonized workflow for metadata harvesting, these should be tested in different communities to evolve community-specific metrics and evaluation procedures. The research clusters will play a pivotal role in this to help build community consensus. A sufficient degree of interoperability across community-specific metrics must also be achieved to be able to compare evaluations at an EOSC-wide level. Use of metadata schemas, standardised distribution formats, data models and controlled vocabulary enables data quality assessment. A community forum where tool providers could come together would help in this aim. To enhance community trust in FAIR Assessments, an emergent widely representative and transparent governance body around FAIR metrics should be supported and encouraged, such that the tools being applied to judge stakeholder's data holdings have undergone an open peer-review process.

B.3.3 Stage 3/Objective 3 – Establish a sustainable and federated infrastructure enabling open sharing of scientific results

B.3.3.1 *European level priorities*

A. By the end of 2025 a sustainable initial operational EOSC Core should be up and running, its operation monitored and evaluated to ensure EOSC is fit-for-purpose, and lessons distilled from the Core procurement exercise as input for the next round of Core funding.

A1 From 2025 onwards, the initial EOSC Core should be enhanced and extended to support emerging and challenging use cases.

B. Define and develop the federation model of EOSC, consisting of national, regional, and thematic nodes integrated with the EOSC core and horizontal services and with an initial network of European-level research data lakes, and enhance it by defining a harmonised operational (including cybersecurity aspects) and legal framework to facilitate the sharing of and access to data and services across European countries.

B1 Analyse and agree upon funding and cost-recovery models, and a governance model, which provide sustainable long-term co-investment at European and national levels.

C. Set up bottom-up, user-driven mechanisms to identify which horizontal and thematic services should be centrally financed (2025), which can serve as a pilot for the centrally financed service portfolio for the period after Horizon Europe.

D. EOSC Should continue to co-design, develop and deliver AAI solutions building on existing AAI and standards, as provided by national and regional AAI Federations, potentially addressing the European Researcher Wallet. From 2027 onwards, EOSC will remain a major stakeholder in the pan-European AAI for research and education and will actively contribute with requirements, use cases and participation in standardisation activities. EOSC should also engage and bridge with European Common Data Spaces and Gaia-X, as well as other initiatives in the EC missions if relevant. In addition, EOSC should promote common authentication and authorisation policies and guidelines. In general, EOSC should foster the uptake of policies and AAI services in countries and thematic communities that have not fully adopted such a federated approach.

E. Continue to maintain and enhance the EOSC common search and access engine for FAIR research objects, including multilingual functionality and Search Engine Optimisation.

F. Accelerate the adoption of interoperability and semantic artefact catalogues. Support AI-enabled services that facilitate interoperability in a more automated way.

G. Services for dependable semantic artefact catalogues should be part of the EOSC Core, akin to how vocabularies are published by the EC publications Office at EU Vocabularies.

H. Compliance monitoring of the EOSC Rules of Participation shall be machine actionable as much as possible to facilitate the scalability as well as reducing the cost of implementing and monitoring the rules.

I. A transition towards machine-actionable repositories should be encouraged by adding this as a requirement to the EOSC Rules of Participation. (2025)

J. Provide a PID sustainability plan for EOSC to help the research community understand their long-term financial commitment related with PID adoption.
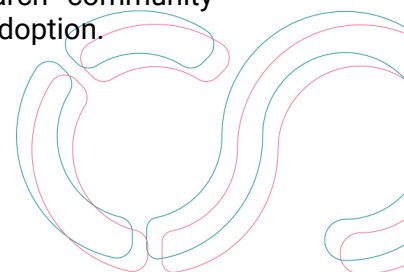
B.3.3.2 National level priorities

K. National policy makers should work with Research Infrastructures, e-Infrastructures and national service providers to establish and operate national or regional EOSC support structures to accelerate the integration with EOSC with regards to the provision as well as usage of data and services.

L. Encourage national policy makers to review and adjust national policies and regulations to enable services to be used in a cross-border context, and engage in a series of in-depth trials to evaluate the utility of proposed resourcing models to agree approaches for EOSC.

M. National policy makers should work with the National Research and Education Networks (NRENs) and National grid Infrastructures (NGIs) to ensure the AAI is available to national research infrastructures, resource providers and users and can be used within and across borders. (2025)

N. Acknowledge European and domain specific semantic artefact catalogues in national infrastructures and guidelines.

O. Support the establishment of last resort data repositories/stores to support European and national open access data policies.

P. Support repositories to become machine-actionable for greater reuse of content via funding programmes and technical support actions. (2025)

B.3.3.3 Institutional level priorities

Q. Research institutions should encourage and support researchers to adopt the data and services federated by EOSC to scale up usage.

R. By 2027 all relevant institutions should be integrated into the National AAIs, enabling their users to collaborate seamlessly at the national and European levels. (2025)

S. Make sure that researchers are aware of and reference existing European and domain specific semantic artefact catalogues.

T. Encourage institutional repositories to become machine-actionable and recommend that research communities deposit data in machine-actionable repositories. (2025)

B.3.3.4 Expected outcomes

U. Consolidation and further enhancement of the EOSC-Core, strengthening the federation aspects and integration leading to greater uptake by national, regional and institutional nodes, data spaces and other partnerships.

V. Sustainability of EOSC is secured by co-investment at European and national levels, focussed on community needs, with agreed cost-recovery models that incentivise participation of institutions and cross-border activity, and by consolidating and harmonising the network of EOSC nodes.

W. Levels of FAIRness increase and more content is accessed by machines due to federated repositories supporting such interactions.

X. Interoperability is enhanced thanks to a greater number of semantic artefact catalogues and maturity models.

Y. EOSC infrastructure is implemented as a federation of EOSC nodes, supported by community agreed Rules of Participation and an interoperability framework.

B.3.3.5   *Clarification text*
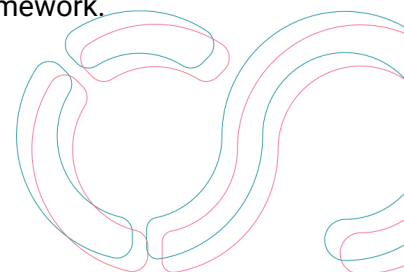
EOSC Federation and MVE
As noted in the *EOSC operations and evolution post-2027* reflection Paper, "EOSC should be a federation of existing and planned research data infrastructures, adding a soft overlay to connect them and making them operate as one seamless European research data infrastructure". There are several basic elements of the Federation, including AAI, PIDs, Interoperability Frameworks, Rules of Participation and common search, all of which are addressed in this MAR.

Work is planned via the forthcoming INFRAEOSC calls and EOSC procurement call, which intends to establish a European EOSC node (EOSC EU Node) to which national, regional and institutional nodes can connect. As such we expect to have an initial EOSC Core up and running by 2025. The emphasis in the 2025-27 work programme should therefore be to enhance and extend the other components of the MVE, namely the EOSC Exchange and the Data Federation, aiming for a completion of the MVE structure by 2026. Ongoing investment is needed to ensure continuity and sustainability. The EOSC Federation model should also be further enhanced by defining a harmonised operational and legal framework to facilitate the sharing and the access of data and services across European countries. Centres of expertise should also support stakeholders with onboarding.

Actions are highlighted at national and institutional level to encourage further integration by other nodes, research infrastructures and adoption by researchers. The sustainability question falls to the European and National level. It is foreseen that the EOSC Core needs to be centrally supported, however contributions to this could transition from European funds to national inputs. Outcomes will be dependent on the future governance structure of EOSC and nature of the legal entity.

To enable the adoption of the EOSC platform we recommend an EOSC-wide approach that provides all European-authenticated users with adequate and easy access to high demand services which can serve as a pilot for the centrally financed service portfolio for the period after Horizon Europe. Alternative means to grants could be explored and trialled to determine feasible options for the EOSC Exchange in the future. The timing needs to be aligned with the outcomes of the EOSC procurement call and forthcoming INFRAEOSC-01-04 call which will develop next generation services for the EOSC Core.
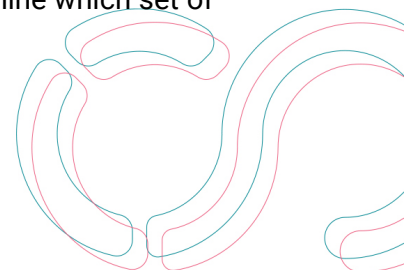
EU nodes, national nodes & regional nodes
The EOSC Federation is a "system of systems" which will be realised by so-called nodes on different levels that are connected horizontally and vertically with each other. In addition to the EU node, which is the subject of an on-going procurement action, national, regional, institutional, thematic and generic nodes may also exist. "Regional" does not necessarily mean a region within one country, but can also mean the aggregation of countries resp. different regions of several countries (= community level). To realise the (connection of) nodes, the commitment from the national policy makers will be essential.

EOSC resourcing models
Enabling nationally funded facilities to be used in a cross-border context continues to be a challenge in the uptake of EOSC. While a variety of current projects and EOSC Association Task Forces are working on this issue to identify potential resourcing models to use, greater engagement is needed at the Member State level to review and where appropriate adjust

local policies and regulations. Support from Member States is needed to encourage inputs from national infrastructures and to permit facilities to be used in a cross-border context. The proposed resourcing models also need to be tested and evaluated to determine which set of approaches to implement in the context of EOSC.

Authentication and Authorisation Infrastructure (AAI)

In the last two decades, research collaboration has been a driving force for the evolution of AAI in Europe and across the globe. With the work in the national academic federations, eduGAIN, and then AARC, the foundation has been established for building a robust EOSC AAI that can meet the requirements of scientific collaboration. With the development and implementation of AAIs across Europe and with the new building block provided by the European Commission and the Member States in eIDAS 2.0 and the European ID wallet, digital identities are becoming fundamental for the digital life of the European citizens.

In the period 2025 - 2027, EOSC should expect that there is no further need to develop, operate and maintain its own AAI, but rather to be able to use pan-European AAI that will be provided by the National and Regional AAI Federations and which will meet all its requirements, enabling EOSC to focus on the seamless delivery of services. Efforts should be focused on activities at national and institutional level to ensure AAI is available to national infrastructures and resource providers and that all institutions are integrated into national AAIs.

EOSC search

Although a cross-search has been developed for EOSC enabling data, software and other research resources from a variety of repositories to be collectively queried, more work is still needed to enhance search capabilities and expand the number of resources and types of research outputs included. In particular, Search Engine Optimisation, multilingual tools, and cross-domain mappings of search terms to facilitate uptake by researchers across Europe using their native languages are recommended.

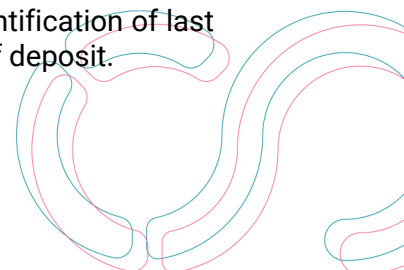## Semantic interoperability and semantic artefacts

Projects such as WorldFAIR and FAIR Impact continue to work on domain standards and developing models for cross-domain interoperability. Within the forthcoming work programme, emphasis should shift to a maturity model for cross-domain interoperability, enabling and encouraging communities to increase the potential for reuse of data. Catalogues of semantic artefacts are also needed in addition to domain standards. Semantic artefacts are machine readable models of knowledge such as controlled vocabularies, thesauri, and ontologies which facilitate the extraction and representation of knowledge within data sets using annotations or assertions. Dependable semantic artefact catalogues should be part of the EOSC core and acknowledged in national infrastructure and guidelines to ensure researchers are aware of and using existing catalogues.

Rules of Participation

A first iteration of a compliance monitoring framework should be developed for the EOSC Rules of Participation. This should include the preparation of a certification framework, training materials and toolsets necessary for the experts to perform the machine-aided certification of EOSC resources and catalogues. Activities are also needed to increase awareness of the RoP among the community. This should include early engagement with the EOSC stakeholders to facilitate adoption of the RoP and to avoid them becoming a barrier to federation.

Machine-actionable repositories

In line with FAIR principles and data quality which are key fundamentals of EOSC, repositories should be machine-actionable so the data, software and publications they hold can be queried and reused at scale. Actions are needed at all levels to support this. At the European level, policies should encourage the transition to machine-actionable repositories by including requirements in the EOSC Rules of Participation. Support programmes are also needed at Member State level to enable national, thematic and institutional repositories to make the transition. Institutions also play a key role in encouraging research communities to choose machine-actionable repositories. It is also recommended to support the identification of last resort data repositories/stores to ensure researchers always have a place of deposit.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 195 –

Widening

In conjunction with the widening actions emphasized under objective 1, actions are needed at the European level by the operators of EOSC to target other data spaces and partnerships to be part of the federation and integrate into EOSC. The policy and governance actions led by the EOSC Partnership to enable collaboration with these stakeholders can support activity here.
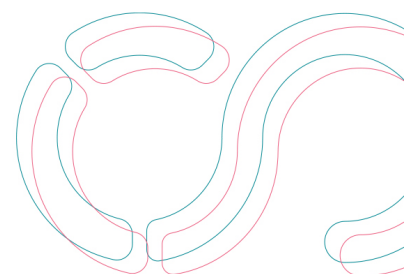
Appendix C   Summary of priorities for 2025

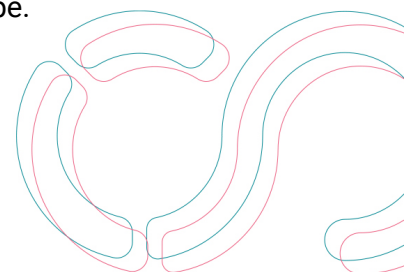## C.1   For OS to become the 'new' normal (objective 1)

1. Create an aligned core part of the curricula for data stewardship for Europe; stimulate the development of curricula in the countries, for the different types and levels of data stewards and develop part of these curricula as core common denominators, helping both support research supporting staff and researchers. Develop networks for data stewards, as well as research software engineers, semantic artefact curators and Open Science communities, leveraging on existing national networks and coordinating these at European level (Activities under this priority should take into account and build on results of relevant projects such as Skills4EOSC and their approach to competence centres).

2. Collect ongoing national best practices and case studies contributing to the development of EOSC as described in the SRIA. Two examples: 1) Citizen Science developments; 2) the research and career assessment part of open science practices. Activities under this priority should take into account and build on results of relevant projects such as GraspOS, and align with related initiatives such as the Coalition for Advancing Research Assessment (CoARA).

C.2   To develop standards and tools (objective 2)

3. Advance the technical specifications to support digital objects to be FAIR-by-design and evaluate viable options to adopt the concept of FAIR digital objects for operational data services.

4. Stimulate community-specific FAIR metrics and data quality constraints in harmony with the FAIR assessment model at international level (RDA) (especially on the interoperability aspect). Engage with research infrastructures (clusters) helping them to apply FAIR in their contexts. Compare and evaluate / test the different tools for measuring 'FAIRness' and further develop them for different disciplines / domains. Guidelines should be provided to navigate the various tools.

## C.3   To establish a federated structure (objective 3)

5. Develop, prototype and test machine-actionable research data repositories and link them to the EOSC EU Node. Building on the work programme for 2024: HORIZON-INFRA-2024-EOSC-01-03: Enabling a network of EOSC federated and trustworthy repositories and enhancing the framework of generic and discipline specific services for data and other research digital objects. By creating blueprints on how to transition towards machine-actionable repositories the amount of data accessible by machines will increase, leading to cross-domain reuse of data and enhanced interoperability of data and services, including AI based services. Developing and maintaining open interfaces, alignments, crosswalks, and APIs will enable this interoperability and will facilitate the automated composition of EOSC data and services. This should be done in full synergy with the EOSC Interoperability framework and in conformity with relevant legislation including the European Data and AI Acts.

6. By the end of 2025 a sustainable initial EOSC Core should be up and running and thereafter further enhanced and extended to support emerging and challenging use cases. For this a user driven mechanism should be developed to determine a selection of most needed horizontal and thematic services, which can serve (as a pilot) for the centrally financed service portfolio for the period after Horizon Europe.

7. Define the requirements for national / local / thematic EOSC nodes. Stimulate their development and those that may come from Research Infrastructures or other parties.
8. A dedicated provision for supporting the outcome of the EOSC EU Node procurement; handing it over to the (interim) operating party; running and maintaining it; creating the possibility to enhance it by 'onboarding' (pilot) nodes from countries / regions, domains, research infrastructures, etc.

[1] See: https://ec.europa.eu/info/news/process-towards-agreement-reforming-research-assessment-2022-jan-18_en
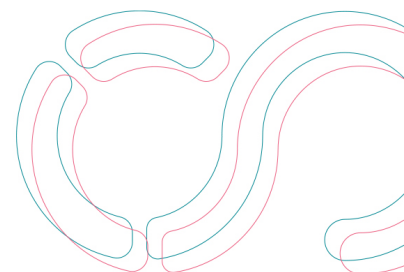
Appendix D   Clarification Text

Data stewardship
Codified and certificated Data stewardship curricula, both for support staff and researchers, should be aligned as much as possible at the European level to ensure a base standard of Open Science and FAIR practice is reached. The curricula should build on existing work conducted by EOSC projects, institutions and the global community, for example in the context of RDA Interest Groups. In particular, we recommend an emphasis on semantic artefact curation to bridge between domain specific and domain agnostic skills. There is a lack of awareness and skills when it comes to the curation of semantic artefacts, which should be addressed by future data stewardship curricula.

Both Member States and institutions have a responsibility to roll out the curricula and recognise the importance of these skills, particularly by supporting positions and opportunities for career development for professional data stewards. Outreach programmes should target young researchers by connecting at the regional level. In addition to the curricula, professional networks for data stewards and a recognition of the associated costs as eligible in national grant schemes are needed to incentivise the change of culture.

Competence Centres and Support
Competence Centres (CC) provide a possible point of reference in a specific Country/Region/Theme/Science Cluster. They strengthen their infrastructures and these in turn strengthen the European structure, like the already existing national HPC competence centres[47]. Ongoing projects should be linked to these CC's. Existing (data) competence centres play an enabling role in advocating Open Science for various actors, in accordance with the Digital Skills Agenda and the EOSC, as well as advancing the local offer of support, counselling and training. Building such CC's is key to knowledge transfer on Open Science, FAIR and CARE in the EOSC context, and more broadly in the context of Data Spaces. Technical support to researchers in the use of research services is also key to successful research outcomes and full exploitation of research resources. While the structure, operational mode and organisation of Competence Centres may vary very widely, they are usually associated with excellence, training, interdisciplinarity, standardisation and a collaborative approach that ensures harmonisation and alignment at European level.

FAIR assessment
To support the transition to Open Science and FAIR practice, the work done by researchers relevant for this transition needs to be recognised and rewarded. New measures need to be defined and tested to assess uptake and ensure they lead to the desired culture change. A recent major development in the landscape is the coalition for advancing research assessment (CoARA). This initiative will entail a period of experimentation by coalition members for several years. This experimentation needs to be supported by further EOSC projects aiming at development, piloting, implementation, and medium-to-long-term evaluation of new methods for research assessment. Research funding and performing organisations, whether CoARA members or not, should include Open Science activities in research review mechanisms.

Preservation of research outputs
The community needs to be made more aware about the roles and responsibilities for long-term preservation of research outputs (including data, metadata, software and semantic artefacts) as part of the overall goal to make this the 'new normal'. Preservation procedures, systems and appraisal practices must be embedded and made machine actionable for

---

[47] EuroCC ACCESS (eurocc-access.eu)
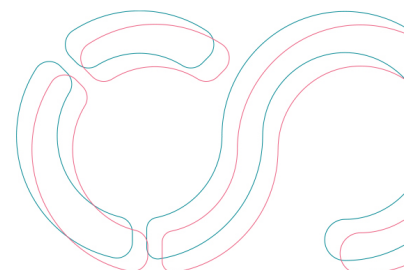
efficiency, interoperability and transparency, but the cost of this needs to be recognised and funded. Disciplinary expectations for preservation of research outputs must be combined for cross-disciplinary needs. Transparency of appraisal ensures broad input on decisions about changing levels of preservation or transfer to alternative repositories. Retention and preservation should be transparent in digital object metadata as well as at the repository level. Semantic artefacts - as key knowledge organisation tools - require special attention concerning long-term preservation.

Engagement and uptake
Expanding on use cases should be a key emphasis in this period to support uptake and widening of EOSC. The EOSC Nodes concept should be tested by different communities, including the data spaces and industry to validate applicability beyond academic researchers. Member State involvement is key to incentivise other sectors to trial EOSC.

A dedicated provision is also recommended to assist the EOSC Partnership in evolving its work with different stakeholders, specifically research performing organisations, the data spaces and countries less engaged in EOSC. The actions could be supported by cascading grants to facilitate uptake where it is lacking, or dedicated funding lines to support the countries and regions. Many countries have existing national open science strategies or initiatives, and these should be cooperated with to align with EOSC. Activities in the dedicated provision should address topics such as competence centres, the transition to a new governance structure, service delivery models and monitoring of Open Science activities via the EOSC Observatory. The EOSC Association should be properly financed over time as part of sustainability plans when adopting responsibility for project outputs.
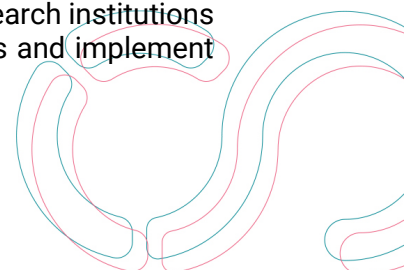
PIDs
Persistent Identifiers (PIDs) are a critical component of the underlying infrastructure to allow resources to be accurately identified, accessed and linked to other objects. Many types of PID are already in common usage and this should be sustained via support at the European and national level, encouraging adoption by communities that are not already active. Support should be directed towards national and institutional services, enabling their adoption of PIDs to ensure content federated into EOSC is well contextualised. Services being supported at the European level such as PID Meta-Resolver, Data Type Registries, PID graph technologies and the PID Compliance Assessment Toolkit should also be incentivised by Member States and institutions to promote uptake, and researchers should be encouraged to use PIDs for the whole research lifecycle as well as for their research activities. The interoperability of PID systems should also be encouraged as objects often get assigned two or more different PIDs since few interactions are possible between the providers. A long-term sustainability plan should also be developed within EOSC for PIDs to help the research community understand their long-term financial commitment related with PID adoption.

Interoperability
The absence of quality metadata continues to be a major obstacle to the discovery and reuse of research outputs. As such, continued investment in developing, evaluating and maintaining community-based metadata standards is recommended at the European level and work to encourage adoption is suggested within institutions. The adoption of standards is also critical to support interoperability and accessibility. The EOSC Interoperability Framework should be compatible with the SIMPL guidelines for the Data Spaces, requiring collaboration as both develop. To enable interoperability, work on the development of open interfaces, alignments, crosswalks and APIs must continue, and research communities need to be encouraged and supported to dedicate time to this important work. Interoperability should cover all the outputs of research as well as the services, enabling composability of resources so research workflows can run smoothly without manual intervention from the user. Research institutions also play a key role in encouraging researchers to adopt relevant standards and implement FAIR.

FAIR digital objects and FAIR metrics

An emphasis has been placed on further support for FAIR digital objects since their implementation encourages the usage of PIDs and agreed metadata schemas, and as such facilitates interoperability. They provide a conceptual and implementation framework to develop scalable cross-disciplinary capabilities, deal with the increasing data volumes and their inherent complexity, build tools that help to increase trust in data, create mechanisms to efficiently operate in the domain of scientific assertions, and promote data interoperability. FAIR digital objects are identified via PIDs and can be regarded as contextualised sub-graphs of general scientific knowledge graphs (SKGs), which allow to account, describe and share everything about particular research. Production and use of FAIR artefacts should be encouraged by developing FAIR assessment, AI-ready FAIR research data, adoption of consistent FAIR data standards and FAIR practices including recognition and rewards and training.

FAIR metrics should be tested in different communities to evolve community-specific metrics and evaluation procedures. A sufficient degree of interoperability across community-specific metrics must be achieved to be able to compare evaluations at an EOSC-wide level. Use of metadata schemas, standardised distribution formats, data models and controlled vocabulary enables data quality assessment. To enhance community trust in FAIR Assessments, an emergent widely representative and transparent governance body around FAIR metrics should be supported and encouraged, such that the tools being applied to judge stakeholders' data holdings have undergone an open peer-review process.

Source-code repositories

Further support is needed to support the development, archiving, quality assurance, sharing and reuse of research software. At the European level, we need to evaluate the viable options to provide European and international researchers with a platform for software development, sharing and preservation, most likely as a federated network of software repositories. A study should be launched to build a Pan-European software catalogue. Stress should be put on reusability, and the use of standards that facilitate the deployment and uptake of software.

EOSC Federation

The EOSC Federation aims to provide Europe's researchers with the necessary digital resources to conduct research within and across disciplines and borders according to the principles of FAIR data and open science, in a trustworthy and secure environment driven by the scientific communities.

This is done by putting in place a 'system of systems' with appropriate organisational and operational structure, between institutional, regional, national, or European (thematic) data repositories, research infrastructures, e-infrastructures and other scientific service providers. For the EOSC Federation, these "institutional, regional, national, or European data repositories, research infrastructures, e-infrastructures and other scientific service providers" – i.e. the "systems" comprising the "system of systems" – are considered "EOSC Nodes".

The target end-users of the resources of the EOSC Federation are primarily researchers of research performing organisations and research infrastructures across Europe and beyond. Potential users of the EOSC Federation also include citizen scientists and researchers in industry.

The organisations that participate in the EOSC Federation will work together to better understand and address the needs of researchers, providing them with appropriate support and equipping them with the right tools.

The EOSC Federation will seek to provide a seamless experience to its users, aiming for a high degree of integration and interoperability of the provided resources. It will count on the

EOSC Association AISBL
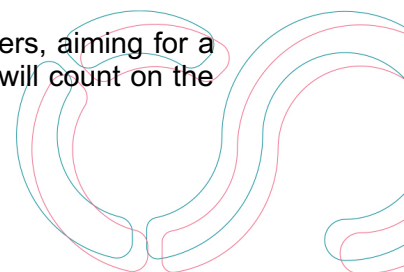
latest community know-how, adopting and building on well-established and well-functioning structures and frameworks.

It will increase adoption of the FAIR principles through standardised FAIR practices and improved data and service interoperability and will foster a more collaborative and integrated scientific environment.

The EOSC Federation will act to secure long-term preservation of selected data and will contribute to the alignment of national guidelines facilitating openness wherever possible. It will monitor the activity of the Federation, the included resources and opportunities for consolidation.

To fully succeed in enabling Open Science and achieving global impact, the EOSC Federation will seek to establish links with similar initiatives in other regions of the world.

EU nodes, national nodes & regional nodes

The EOSC Federation is a "system of systems" which will be realised by so-called nodes on different levels that are connected horizontally and vertically with each other. In addition to the EU node, created in 2024, national, regional, institutional, thematic and generic nodes will come in 2025. "Regional" does not necessarily mean a region within one country but can also mean the aggregation of countries resp. different regions of several countries (= community level). To realise the (connection of) nodes, the commitment from the national policy makers will be essential.

The EOSC Nodes will be the entry points for end users to the entire EOSC Federation, with each node offering its own services (including data reposing and accessing services) and possibly services of other providers ('third-party' services onboarded to the EOSC Node), in compliance with the EOSC Federation's common rules and requirements as well as the own policies of the EOSC Node, An organisation may therefore enrol its activities as an EOSC Node in the EOSC Federation or onboard its services on an existing EOSC Node as a provider. These nodes will thus offer (scientific) services and/or data that adhere to the FAIR principles, including curated research outputs (such as publications, datasets, software, etc.), specialised knowledge, applications, tools, infrastructure and/or platform services, and/or data processing and storage capabilities.
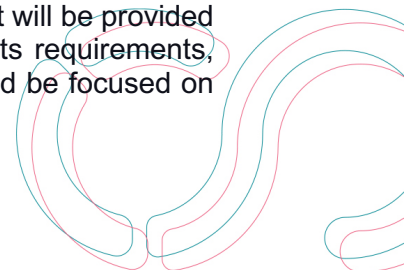
EOSC resourcing models

Enabling nationally funded facilities to be used in a cross-border or cross-domain context continues to be a challenge in the uptake of EOSC. Some current projects are working on this issue, to identify potential resourcing models to use, but engagement is needed at the Member State level to review and where appropriate adjust local policies and regulations. Support from Member States is needed to encourage alignment of national infrastructures and to permit facilities to be used in a cross-border or cross-domain context. The proposed resourcing models also need to be tested and evaluated to determine which set of approaches to implement in the context of EOSC.

Authentication and Authorisation Infrastructure (AAI)

In the last two decades, research collaboration has been a driving force for the evolution of AAI in Europe and across the globe. With the work in the national academic federations, eduGAIN, and then AARC, the foundation has been established for building a robust EOSC AAI that can meet the requirements of scientific collaboration. With the development and implementation of AAIs across Europe and with the new building blocks provided by the European Commission and the Member States in eIDAS 2.0 and the European ID wallet, digital identities are becoming fundamental for the digital life of European citizens.

In the period 2026-2027, EOSC should expect that there is no further need to develop, operate and maintain its own AAI, but rather to be able to use pan-European AAI that will be provided by the National and Regional AAI Federations and which will meet all of its requirements, enabling EOSC to focus on the seamless delivery of services. Efforts should be focused on

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 202 –

activities at national and institutional level to ensure AAI is available to national infrastructures and resource providers and that all institutions are integrated into national AAIs.

Semantic interoperability and semantic artefacts

Projects such as WorldFAIR and FAIR Impact work on domain standards and developing models for cross-domain interoperability. Emphasis needs to shift to a maturity model for cross-domain interoperability, enabling and encouraging communities to increase the potential for reuse of data. Catalogues of semantic artefacts are also needed in addition to domain standards. Semantic artefacts are machine readable models of knowledge such as controlled vocabularies, thesauri, and ontologies which facilitate the extraction and representation of knowledge within data sets using annotations or assertions. Dependable semantic artefact catalogues should be part of the EOSC Federation and acknowledged in national infrastructure and guidelines to ensure researchers are aware of and using existing catalogues.

Machine-actionable repositories

In line with FAIR principles and data quality, which are key fundamentals of EOSC, repositories should be machine-actionable so that the data, software and publications they hold can be queried and reused at scale. Actions are needed at all levels to support this. At the European level, policies should encourage the transition to machine-actionable repositories by including requirements in the EOSC Federation policies. Support programmes are also needed at Member State level to enable national, thematic and institutional repositories to make the transition. Institutions also play a key role in encouraging research communities to choose machine-actionable repositories.

Widening

Actions will be needed at the European level by the operators of EOSC to target other data spaces and partnerships to be part of the federation and integrate into EOSC, and to contribute to the establishment of a global data commons. The policy and governance actions led by the EOSC Tripartite should enable collaboration with these stakeholders.

Security and Risk Management

Services, data and other resources available as part of the EOSC Federation must adhere to recognised cybersecurity standards, enforced by the operators of EOSC according to an agreed framework. Node operators will need to take responsibility for ensuring the security of the resources they offer. A risk assessment and management framework will also need to be developed at federation level and implemented across the network of nodes.

Artificial Intelligence

Artificial intelligence (AI) and machine actionability will become indispensable as reliance increases on computational support to deal with the growing volume and complexity of data. Metadata plays an increasingly important role in promoting semantic and technical interoperability and establishing the FAIR principles for digital objects, but machine actionability will be required to exploit resources to their full advantage.

EOSC will need to deliver infrastructure and data services which support the development of AI models, machine learning workflow management tools, and AI-based assistance tools for the discovery and composition of open science resources.

Rules of Participation

The EOSC Rules of Participation should continue to develop as part of the creation of the EOSC Federation. It needs to include a certification framework, training materials and toolsets necessary for machine-aided certification of EOSC resources and catalogues, and monitoring of their compliance. Activities are also needed to increase awareness of the RoP among the

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
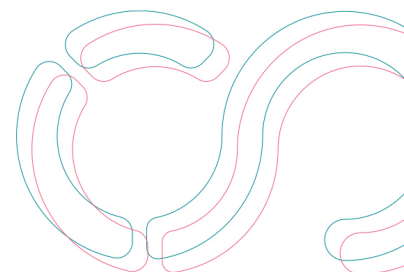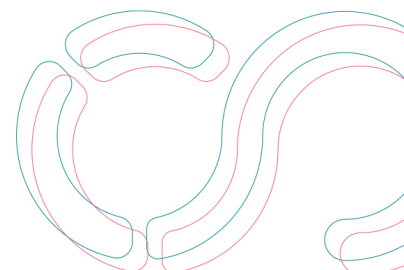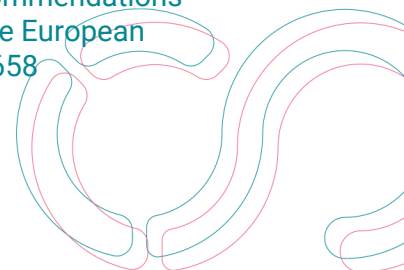Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 203 –

community, including engagement with EOSC stakeholders to facilitate adoption of the RoP and to avoid them becoming a barrier to federation.

References

[AARC]	https://aarc-project.eu/
[AARC_BPA]	https://aarc-project.eu/architecture/
[AEGIS]	https://aarc-project.eu/about/aegis/
[AGU_IN41A]	'FAIR Data Is Not Enough: Communicating Data Quality and Making Analytical Code FAIR I' https://agu.confex.com/agu/fm18/meetingapp.cgi/Session/60523
[AGU_TH43B]	'Data Fair: Sharing Your Software — What Is FAIR?' https://agu.confex.com/agu/fm18/meetingapp.cgi/Session/56228
[AGU_U41A]	'How Safe and Persistent Is Your Research?' https://agu.confex.com/agu/fm17/meetingapp.cgi/Session/25700
[AI4EOSC]	https://ai4eosc.eu/
[ARCHIVER]	https://www.archiver-project.eu/
[Beckman_2019]	P. Beckman, J. Dongarra, N. Ferrier et al., 'Harnessing the Computing Continuum for Programming Our World' https://www.researchgate.net/publication/332246123_Harnessing_the_Computing_Continuum_for_Programming_Our_World
[Budapest_OAI]	https://www.budapestopenaccessinitiative.org/read
[C_SCALE]	https://c-scale.eu/
[C19_Portal]	https://www.covid19dataportal.org
[CARE]	https://www.gida-global.org/care
[CESAER_WP_BCESR]	'Boost the Careers of Early-Stage Researchers', White Paper dated 10 February 2020 https://www.cesaer.org/news/boost-careers-of-early-stage-researchers-451/
[CESAER_WP_HRS4R]	'Sharing Experiences with the Human Resources Strategy for Researchers', White Paper dated 23 May 2018 https://www.cesaer.org/news/sharing-experiences-with-human-resources-strategy-for-researchers-122/
[CESAER_WP_NGM]	'Next Generation Metrics', White Paper dated 10 June 2020 https://www.cesaer.org/news/next-generation-metrics-545/
[CESAER_WP_RDM]	'Advancing Research Data Management in Universities of Science and Technology', White Paper dated 13 February 2020 https://www.cesaer.org/news/advancing-research-data-management-456/
[CERIF]	https://www.eurocris.org/eurocris_archive/cerifsupport.org/cerif-in-brief/index.html
[COAR]	https://www.coar-repositories.org/
[CODATA]	https://codata.org/
[CodeMeta]	https://codemeta.github.io/
[Cohen_2019]	J. B. Cohen et al. (2019) 'Towards Responsible Research Career Assessment' http://doi.org/10.5281/zenodo.3560479
[CoreTrustSeal]	https://www.coretrustseal.org/
[Crossref]	https://www.crossref.org/

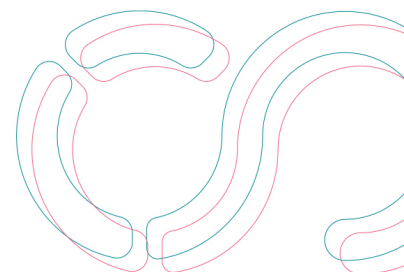|	
[DataCite]	https://datacite.org/
[DICE]	https://www.dice-eosc.eu/
[DICOM]	https://www.dicomstandard.org/
[Digital_Science_2019]	Digital Science; Fane, Briony; Ayris, Paul; Hahnel, Mark; Hrynaszkiewicz, Iain; Baynes, Grace; et al., 'The State of Open Data Report 2019'. Digital Science. Report https://doi.org/10.6084/m9.figshare.9980783.v1
[Dillo_2020]	Ingrid Dillo, Marjan Grootveld, Simon Hodson & Sara Pittonet Gaiarin, 'Second Report of the FAIRsFAIR Synchronisation Force (D5.5) (Version 1.0)', 2020. Zenodo. https://doi.org/10.5281/zenodo.3953978
[DIN_31644]	https://global.ihs.com/doc_detail.cfm?document_name=DIN%2031644&item_s_key=00585595
[DOI]	https://www.doi.org/index.html
[DONA]	https://www.dona.net/
[Donoho_1995]	D. Donoho et al. (1995). 'WaveLab and Reproducible Research', in A. Antoniadis & G. Oppenheim (eds.), *Wavelets and Statistics* 103, 55–81 https://doi.org/10.1007/978-1-4612-2544-7_5
[DORA]	https://sfdora.org/
[DORA_1]	DORA (2020), 'Rethinking research assessment – Ideas for action' https://sfdora.org/wp-content/uploads/2020/05/DORA_IdeasForAction.pdf
[DORA_2]	DORA (2020), 'Good Practices – Research Institutes' https://sfdora.org/resource-library/?_resource_type=good-practices&_intended_audience=research-institutes
[DS_Roles]	https://doi.org/10.5281/zenodo.4321264
[Dzale_2017]	Yeumo E. Dzale, M. Alaux, E. Arnaud et al., 'Developing data interoperability using standards: A wheat community use case' [version 2; peer review: 2 approved]. F1000Research 2017, 6:1843 https://doi.org/10.12688/f1000research.12234.2
[EC_Cloud]	'European Cloud Initiative – Building a competitive data and knowledge economy in Europe', (COM(2016) 178 final https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=15266
[EC_COM_New_ERA]	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:628:FIN
[EC_Data_Strategy]	'A European strategy for data', 19.2.2020 COM(2020) 66 final https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf
[EC_EG_FAIR]	'Turning FAIR into reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data', 2018, ISBN 978-92-79-96546-3, doi 10.2777/1524 KI-06-18-206-EN-N https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1
[EC_EG1_EOSC]	https://op.europa.eu/en/publication-detail/-/publication/2ec2eced-9ac5-11e6-868c-01aa75ed71a1
[EC_EG2_EOSC]	'Prompting an EOSC in practice: Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC)', 2018, doi:10.2777/112658

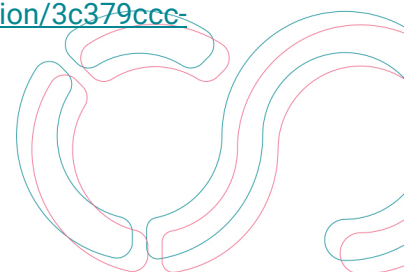|	https://op.europa.eu/en/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1
[EC_ERA_Actions]	https://ec.europa.eu/info/files/new-european-research-area_en
[EC_ERAvsCoronaAP]	https://ec.europa.eu/info/files/first-eravscorona-action-plan-short-term-coordinated-research-and-innovation-actions_en
[EC_FP1]	http://www.jeupiste.eu/horizon-2020-and-around/historical-timeline-framework-programme
[EC_HE_Missions]	https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme/missions-horizon-europe_en
[EC_ICT_Standard]	https://ec.europa.eu/digital-single-market/en/rolling-plan-ict-standardisation
[EC_ICT_TechSpec]	https://ec.europa.eu/growth/single-market/european-standards/ict-standardisation/ict-technical-specifications_en
[EC_Interoperability]	European Interoperability Framework, European Union, 2017, doi 10.2799/78681 https://op.europa.eu/en/publication-detail/-/publication/bca40dde-deee-11e7-9749-01aa75ed71a1/language-en
[EC_NG-OS-Metrics]	'Next-generation metrics: Responsible metrics and evaluation for open science', 2017, ISBN 978-92-79-66130-3 doi:10.2777/337729 KI-01-17-130-EN-N https://ec.europa.eu/research/openscience/pdf/report.pdf
[EC_Open_Vision]	https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe
[EC_OS]	Shaping Europe's digital future	Open Science. Last updated 5 November 2019. https://ec.europa.eu/digital-single-market/en/open-science.
[EC_OSPP]	https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform
[EC_Procurement]	https://op.europa.eu/en/publication-detail/-/publication/d922fcc0-b292-4f70-96c2-3c18fceea06b/language-en
[EC_PSI]	Directive (EU) 2019/1024 on open data and the re-use of public sector information https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024
[EC_Rec_2018/790]	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018H0790
[EC_Rec_C(2012)4890]	https://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf
[ECTS]	https://ec.europa.eu/education/resources-and-tools/european-credit-transfer-and-accumulation-system-ects_en
[eduGAIN]	https://edugain.org/
[EGI_ACE]	https://www.egi.eu/project/egi-ace/
[ELIXIR]	https://elixir-europe.org/
[eLTER]	https://www.lter-europe.net/
[EMBRC]	https://www.embrc.eu/
[Enabling_FAIR]	https://osf.io/jy4d9/

![EOSC logo]

[ENVRI-FAIR]	https://envri.eu/home-envri-fair/
[EOSC]	https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/eosc_en
[EOSC4CANCER]	https://eosc4cancer.eu/
[EOSC_Brand_Guidelines]	https://www.eoscsecretariat.eu/sites/default/files/guideline_eosc_secretariat_2020.pdf
[EOSC_CCF]	https://www.eoscsecretariat.eu/funding-opportunities/co-creation-requests
[EOSC_Comms_Toolkit]	https://www.eoscsecretariat.eu/about/communications-kit
[EOSC-Core_Costs]	'The Vivus Study: Ensuring that EOSC-Core & Minimum Viable EOSC are sustainable through a study on their costings, potential business models and funding schemes', 2020, DOI 10.5281/zenodo.4395885 https://doi.org/10.5281/zenodo.4395885
[EOSC_EAP]	https://eosc-hub.eu/eosc-early-adopter-programme
[EOSC_EB]	https://www.eoscsecretariat.eu/eosc-governance/eosc-executive-board
[EOSC_Focus]	https://www.eosc.eu/eosc-focus-project
[EOSC_Future]	https://eoscfuture.eu/
[EOSC_GB]	https://www.eoscsecretariat.eu/eosc-governance/eosc-governance-board
[EOSC_Glossary]	https://www.eosc-portal.eu/glossary
[EOSC_Gov]	https://www.eoscsecretariat.eu/eosc-governance
[EOSC-hub]	https://www.eosc-hub.eu/
[EOSC-hub_CBSvcs]	EOSC-hub project, 'EOSC-hub Briefing Paper – Provision of Cross-Border Services', Version 1.2, 11 September 2020 https://www.eosc-hub.eu/sites/default/files/EOSC-hub%20Briefing%20Paper%20-%20Provision%20of%20Cross-Border%20Services%20-%20For%20Consultation.pdf
[EOSC_IF]	'EOSC Interoperability Framework', (v1.0) 3 May 2020 Draft for community consultation https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf
[EOSC_Landscape]	'Landscape of EOSC-Related Infrastructures and Initiatives: Report from the EOSC Executive Board Working Group (WG) Landscape', 2020, ISBN 978-92-76-25380-8, doi 10.2777/132181, KI-02-20-971-EN-N https://op.europa.eu/en/publication-detail/-/publication/9c7f3c1e-2aea-11eb-9d7e-01aa75ed71a1/language-en/format-PDF/source-173604556
[EOSC_Landscape_CS]	'Country Sheets Analysis: Report from the EOSC Executive Board Working Group (WG) Landscape', 2020, ISBN 978-92-76-25379-2 doi 10.2777/568900, KI-02-20-969-EN-N https://op.europa.eu/en/publication-detail/-/publication/95e4a900-2a21-11eb-9d7e-01aa75ed71a1/language-en/format-PDF/source-173316815
[EOSC-Life]	https://www.eosc-life.eu/
[EOSC_Marketplace]	https://marketplace.eosc-portal.eu/
[EOSC-Nordic]	https://www.eosc-nordic.eu/

eosc

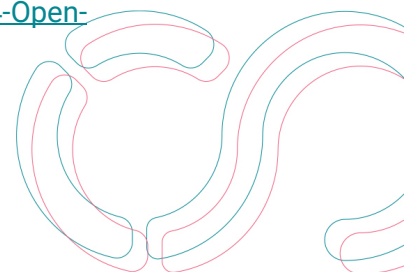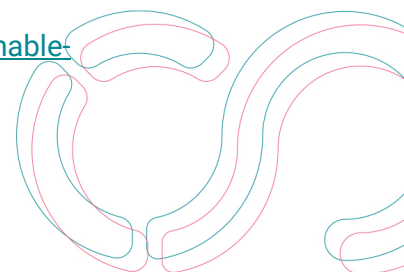|	
[EOSC_PID_Policy]	'A Persistent Identifier (PID) policy for the European Open Science Cloud: Report from the European Open Science Cloud FAIR and Architecture Working Groups', 2020, ISBN 978-92-76-22780-9, doi: 10.2777/926037, KI-04-20-576-EN-N https://op.europa.eu/en/publication-detail/-/publication/35c5ca10-1417-11eb-b57e-01aa75ed71a1
[EOSC-Pillar]	https://www.eosc-pillar.eu/
[EOSCpilot]	https://eoscpilot.eu/
[EOSCpilot_D7.3]	Angus Whyte, Jerry de Vries, Rahul Thorat et al., D7.3: Skills and Capability Framework, 2018 https://www.eoscpilot.eu/sites/default/files/eoscpilot-d7.3.pdf
[EOSC_Portal]	https://www.eosc-portal.eu/
[EOSC_PP]	https://ec.europa.eu/info/files/european-open-science-cloud-eosc_en
[EOSC_Providers]	https://eosc-portal.eu/for-providers
[EOSC_Roadmap]	https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf
[EOSC_RoP]	'EOSC Rules of Participation: Report from the EOSC Executive Board Rules of Participation (RoP) Working Group (WG)', 2021, ISBN 978-92-76-28591-5, doi: 10.2777/30541, KI-01-21-007-EN-N https://op.europa.eu/en/publication-detail/-/publication/a96d6233-554e-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-184432576
[EOSC_Sec]	https://www.eoscsecretariat.eu/
[EOSC_SF]	https://www.eoscsecretariat.eu/eosc-governance/eosc-stakeholder-forum
[EOSC_SIP]	'European Open Science Cloud (EOSC) Strategic Implementation Plan', 2019, ISBN 978-92-76-09175-2, doi 10.2777/202370, KI-03-19-507-EN-N https://op.europa.eu/en/publication-detail/-/publication/78ae5276-ae8e-11e9-9d01-01aa75ed71a1
[EOSC_Svcs&Res]	https://www.eosc-portal.eu/services-resources
[EOSC-Synergy]	https://www.eosc-synergy.eu/
[EOSC_WG_Arch]	https://www.eoscsecretariat.eu/working-groups/architecture-working-group
[EOSC_WG_FAIR]	https://www.eoscsecretariat.eu/working-groups/fair-working-group
[EOSC_WG_Land]	https://www.eoscsecretariat.eu/working-groups/landscape-working-group
[EOSC_WG_RoP]	https://www.eoscsecretariat.eu/working-groups/rules-participation-working-group
[EOSC_WG_Skills]	https://www.eoscsecretariat.eu/working-groups/skills-training-working-group
[EOSC_WG_Sustain]	https://www.eoscsecretariat.eu/working-groups/sustainability-working-group
[EOSC_WGs]	https://www.eoscsecretariat.eu/eosc-working-groups
[EOSC_Work_Plan]	https://op.europa.eu/en/publication-detail/-/publication/3c379ccc-ee2c-11e9-a32c-01aa75ed71a1

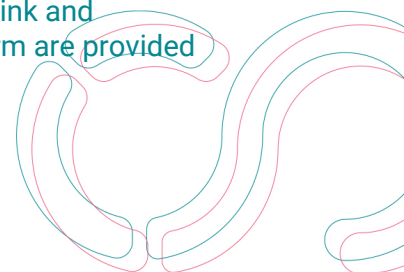| [EP_AID&R]	https://ec.europa.eu/info/sites/info/files/research_and_innovation/funding/documents/ec_rtd_he-partnerships-european-partnership-on-artificial-intelligence-data-and-robotics.pdf
[ePIC]	https://www.pidconsortium.net/
[ESCAPE]	https://projectescape.eu/
[ESRC_Data_Policy]	'ESRC Research Data Policy' https://esrc.ukri.org/files/about-us/policies-and-standards/esrc-research-data-policy/
[ETP4HPC_SRA]	https://www.etp4hpc.eu/pujades/files/ETP4HPC_SRA4_2020_web(1).pdf
[EUA_RATOS]	B. Saenen, R. Morais, V. Gaillard and L. Borrell-Damián, '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
[EUDAT_B2FIND]	https://www.eudat.eu/services/b2find
[EuroHPC_JU]	https://ec.europa.eu/digital-single-market/en/eurohpc-joint-undertaking
[EXIF]	http://cipa.jp/std/documents/download_e.html?DC-008-Translation-2019-E
[ExPaNDS]	https://expands.eu/
[FAIR4RS_WG]	https://www.rd-alliance.org/groups/fair-4-research-software-fair4rs-wg
[FAIR_Principles]	M. Wilkinson, M. Dumontier, I. Aalbersberg et al., 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data* volume 3, Article number: 160018 (2016) https://www.nature.com/articles/sdata201618
[FAIRCORE4EOSC]	https://faircore4eosc.eu/
[FAIR_IMPACT]	https://fair-impact.eu/
[FAIRsFAIR]	https://www.fairsfair.eu/
[FAIRsharing]	https://fairsharing.org/
[FAO]	http://www.fao.org/home/en/
[FIM4R]	https://fim4r.org/
[FI_OS_Coord]	https://avointiede.fi/en/coordination
[FI_OS_Decl]	'Declaration for Open Science and Research 2020–2025' https://doi.org/10.23847/isbn.9789525995251
[FI_OS_WGs]	https://avointiede.fi/en/open-science-expert-panels/open-data
[Forbes_TB-L]	https://www.forbes.com/sites/joemckendrick/2019/07/15/world-wide-web-inventor-tim-berners-lee-pushes-for-a-better-web/#4393878c762d
[FOSTER]	https://www.fosteropenscience.eu/
[FOSTER_OS]	https://www.fosteropenscience.eu/content/what-open-science-introduction/
[FREYA_PG]	https://www.project-freya.eu/en/blogs/blogs/the-pid-graph
[G7_OS]	http://www.g8.utoronto.ca/science/2017-G7-Annex4-Open-Science.pdf

[GAIA-X]	https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html
[Genova_2017]	F. Genova, C. Arviset, B. M. Almas et al., 'Building a Disciplinary, World-Wide Data Infrastructure', *Data Science Journal*, 16, p. 16. DOI: http://doi.org/10.5334/dsj-2017-016
[GIDA]	https://www.gida-global.org/
[GODAN]	https://www.godan.info/
[GO FAIR]	https://www.go-fair.org/
[GO FAIR_Principles]	https://www.go-fair.org/fair-principles/
[Gruenpeter & Thornton]	M. Gruenpeter and K. Thornton, Pathways for Discovery of Free Software (slide deck from LibrePlanet 2018) https://en.wikipedia.org/wiki/File:Pathways-discovery-free.pdf
[Handle]	http://www.handle.net/
[Hausman_2019]	J. Hausman, S. Stall, J. Gallagher and W. Mingfang, 'Software and Services Citation Guidelines and Examples', Figshare, 2019. https://doi.org/10.6084/M9.FIGSHARE.7640426
[HNSciCloud]	https://www.hnscicloud.eu/
[Horizon_Europe]	https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en
[IAU_FITS]	https://fits.gsfc.nasa.gov/
[Ifremer]	https://wwz.ifremer.fr/
[INORMS_RtR]	https://arma.ac.uk/rethinking-the-rankings/
[IPOL]	https://www.ipol.im/
[ISO_16363]	http://www.iso16363.org/
[Jetten_2021]	Jetten, M. et al. 'Professionalising data stewardship in the Netherlands: competences, training and education – Dutch roadmap towards national implementation of FAIR data stewardship' (2021) https://doi.org/10.5281/zenodo.4320504
[Leiden_Manifesto]	http://www.leidenmanifesto.org
[LERU_AP24_OS]	League of European Research Universities, 'Open Science and its role in universities: A roadmap for cultural change', Advice Paper no. 24 May 2018 https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf
[LIBER_Copyright]	https://libereurope.eu/blog/2020/04/14/copyright-coronavirus-statement/
[Lindau_Dec]	http://lindau-declaration.org/
[nestor_Seal]	https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html
[NI4OS-Europe]	https://ni4os.eu/
[OBM]	https://openbiomaps.org/
[OCRE]	https://www.ocre-project.eu/
[OECD_ARDPF]	OECD, OECD Principles and Guidelines for Access to Research Data from Public Funding https://www.oecd.org/sti/inno/38500813.pdf
[OECD_BDWC]	OECD, 'Building digital workforce capacity and skills for data-intensive science', OECD Publishing, Paris, July 10, 2020

	http://www.oecd.org/publications/building-digital-workforce-capacity-and-skills-for-data-intensive-science-e08aa3bb-en.htm
[OECD_DSTI]	OECD, 'The Digitalisation of Science, Technology and Innovation' https://doi.org/10.1787/b9e4a2c0-en
[OECD_SDG]	http://www.oecd.org/dac/sustainable-development-goals.htm
[OpenAIRE_Nexus]	https://www.openaire.eu/openaire-nexus-project
[OpenAIRE_RG]	https://www.openaire.eu/blogs/the-openaire-research-graph
[ORCID]	https://orcid.org/
[ORFG_IB]	'Incentivizing the sharing of research outputs through research assessment: a funder implementation blueprint' http://www.orfg.org/incentivization-blueprint
[OSPP_Report]	'Progress on Open Science: Towards a Shared Research Knowledge System – Final Report of the Open Science Policy Platform https://doi.org/10.2777/00139
[PaNOSC]	https://www.panosc.eu/
[Plan_S]	https://www.coalition-s.org/
[Plomp_2019]	Esther Plomp, Nicolas Dintzner, Marta Teperek and Alastair Dunning, 'Cultural Obstacles to Research Data Management and Sharing at TU Delft', *Insights* 32 (1): 29, 2019. DOI: http://doi.org/10.1629/uksg.484
[RDA]	https://www.rd-alliance.org/
[RDA_FAIR_DMMWG]	https://www.rd-alliance.org/groups/fair-data-maturity-model-wg
[RDA_PID_Kernel]	https://www.rd-alliance.org/group/pid-kernel-information-wg/wiki/pid-kernel-information-guiding-principles
[RDA_PID_Registry]	https://www.rd-alliance.org/group/pid-information-types-wg/outcomes/pid-information-types
[RDA_Recs]	https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs
[RELIANCE]	https://www.reliance-project.eu/
[Scholtens_2019]	S. Scholtens, M. Jetten, J. Böhmer, Ch. Staiger, I. Slouwerhof, M. Van der Geest and C. W. G. Van Gelder, (2019, October 3). 'Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise'. Zenodo. http://doi.org/10.5281/zenodo.3474789
[Sempreviva_2017]	Anna Maria Sempreviva, Allan Vesth, Christian Bak et al., 'Taxonomy and metadata for wind energy R&D', 2017, December 12. Zenodo. http://doi.org/10.5281/zenodo.1199489
[SSHOC]	https://www.sshopencloud.eu/
[Stoy_2020]	Lennart Stoy, Bregt Saenen, Joy Davidson, Claudia Engelhardt and Vinciane Gaillard, 'D7.1 FAIR in European Higher Education (Version v1.0_draft), 2020. Zenodo. https://doi.org/10.5281/zenodo.3629682
[SW_Heritage]	https://www.softwareheritage.org/
[SW_Heritage_IPOL]	https://www.softwareheritage.org/2020/06/11/ipol-and-swh/?lang=es
[TARA]	https://oceans.taraexpeditions.org/en/
[TB-L_Step]	https://inrupt.com/blog/one-small-step-for-the-web
[UN_SDG]	https://www.un.org/sustainabledevelopment/sustainable-development-goals/

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 212 –

[Von_der_Leyen_Agenda]   https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

[VSNU_PP]   'Room for everyone's talent: towards a new balance in the recognition and rewards of academics' https://www.vsnu.nl/recognitionandrewards/wp-content/uploads/2019/11/Position-paper-Room-for-everyone%E2%80%99s-talent.pdf

[VSNU_SEP]   'Strategy Evaluation Protocol 2021-2027' https://vsnu.nl/files/documenten/Domeinen/Onderzoek/SEP_2021-2027.pdf

[WCRI_HKP]   https://wcrif.org/guidance/hong-kong-principles

[WDS]   https://www.worlddatasystem.org/

[Wellcome_ORF]   https://wellcome.ac.uk/grant-funding/people-and-projects/grants-awarded?scheme_id=3580

[WG_Arch_SIRS]   'Scholarly Infrastructures for Research Software: Report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS', 2020, doi: 10.2777/28598 https://op.europa.eu/en/publication-detail/-/publication/145fd0f3-3907-11eb-b27b-01aa75ed71a1/language-en/format-PDF/source-178301654

[WG_FAIR_6Recs]   'Six recommendations for implementation of FAIR practice: By FAIR in Practice Task Force of the European Open Science Cloud', 2020, DOI: 10.2777/986252 https://op.europa.eu/en/publication-detail/-/publication/4630fa57-1348-11eb-9a54-01aa75ed71a1/language-en/format-PDF/source-166584930

[WG_FAIR_CSRecs]   EOSC FAIR Working Group, 'Recommendations on certifying services required to enable FAIR within EOSC', December 2020 https://op.europa.eu/en/publication-detail/-/publication/70aa74b5-53bf-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-184009543

[WG_Skills_DigitalSkills]   https://op.europa.eu/en/publication-detail/-/publication/af7f7807-6ce1-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-190694287

[WG_Sustain_FAIRLady]   'Solutions for a Sustainable EOSC: A FAIR Lady report from the EOSC Sustainability Working Group', 2020, ISBN 978-92-76-25594-9, doi 10.2777/870770, KI-02-20-999-EN-N https://op.europa.eu/en/publication-detail/-/publication/581d82a4-2ed6-11eb-b27b-01aa75ed71a1/language-en/format-PDF/source-175468053

[WG_Sustain_Strawman]   https://www.eoscsecretariat.eu/file/swg-solutionsforasustainableeosc00pdf. 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.

[WG_Sustain_Tinman]   https://www.eoscsecretariat.eu/file/solutionsforasustainableeosc-tinmandraft02dec19pdf. 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.

[What_If]	https://pair-code.github.io/what-if-tool/
[Wikipedia_Gray1]	https://en.wikipedia.org/wiki/Jim_Gray_(computer_scientist)
[Wikipedia_Gray2]	https://en.wikipedia.org/wiki/Data_science
[Wikipedia_OS]	https://en.wikipedia.org/wiki/Open_science
[Wildgaard_2020]	L. Wildgaard, E. Vlachos, L. Nondal, A. V. Larsen and M. Svendsen, (2020, January 31). 'National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management (DM Forum)' (Version 1). Zenodo. http://doi.org/10.5281/zenodo.3609516
[Zenodo]	https://about.zenodo.org/
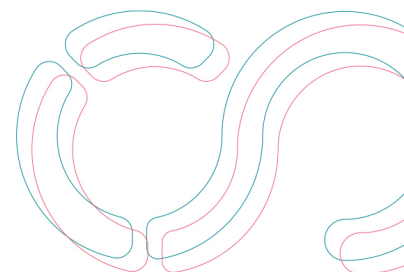
List of Abbreviations

|	
AAI	Authentication and Authorisation Infrastructure
AARC	Authentication and Authorisation for Research and Collaboration
AC	Associated Countries
AEGIS	AARC Engagement Group for Infrastructures
AGU	American Geophysical Union
AI	Artificial Intelligence
AISBL	Association Internationale Sans But Lucratif / International Non-Profit Association
API	Application Programming Interface
ARCHIVER	Archiving and Preservation for Research Environments
BDVA	Big Data Value Association
BPA	Blueprint Architecture (AARC project)
CARE	Collective benefit, Authority to control, Responsibility, Ethics
CCSDS	Consultative Committee for Space Data Systems
CEF	Connecting Europe Facility
CERIF	Common European Research Information Format
CERN	European Organisation for Nuclear Research
COAR	Confederation of Open Access Repositories
CODATA	Committee on Data of the International Science Council
CSF	Critical Success Factor
CSV	Comma-Separated Values
DCAT	Data Catalog Vocabulary
DDI	Data Documentation Initiative
DG CONNECT	EC Directorate-General for Communications Networks, Content and Technology
DG RTD	EC Directorate-General for Research and Innovation
DIAS	Data and Information Access Services
DICOM	Digital Imaging and Communications in Medicine
DIN	Deutsches Institut für Normung / German Institute for Standardisation
DMP	Data Management Plan
DOI	Digital Object Identifier
DONA	Digital Object Architecture Foundation
DORA	Declaration on Research Assessment
DPC	Digital Preservation Coalition
DSCC	Data Stewardship Competence Centre
EB	Executive Board
EBU	European Broadcasting Union
EC	European Commission
ECRIN	European Clinical Research Infrastructure Network
ECTS	European Credit Transfer and Accumulation System
EEG	Electroencephalography
EERA JPWind	European Energy Research Alliance Joint Programme on Wind Energy
EIC	European Innovation Council
eLTER	Long-Term Ecosystem Research in Europe
EMBL	European Molecular Biology Laboratory
EMBRC	European Marine Biological Resource Centre
ENVRI	Environmental Research Infrastructure
EOSC	European Open Science Cloud
EPA	Economic Partnership Agreement
ePIC	Persistent Identifier Consortium for eResearch

EOSC Association AISBL
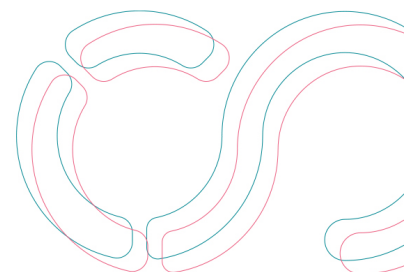
|	
ERA	European Research Area
ERC	European Research Council
ERM	Enterprise Risk Management
ESCAPE	European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures
ESFRI	European Strategy Forum on Research Infrastructures
ESO	European Southern Observatory
ESRC	Economic and Social Research Council
EU	European Union
EXIF	Exchangeable Image File
ExPaNDS	European Open Science Cloud (EOSC) Photon and Neutron Data Service
FAIR	Findable, Accessible, Interoperable and Reusable
FAO	Food and Agriculture Organisation of the United Nations
FITS	Flexible Image Transport System
FP7	The European Union's seventh research and innovation funding programme which ran 2007–2013
FIM4R	Federated Identity Management for Research
G7	Group of Seven
GDPR	General Data Protection Regulation
GIDA	Global Indigenous Data Alliance
GO	General Objective
H2020	Horizon 2020
HAL	Hyper Articles on Line
HE	Horizon Europe
HEI	Higher Education Institution
HNSciCloud	Helix Nebula the Science Cloud
HPC	High-Performance Computing
HR	Human Resources
HTC	High-Throughput Compute
I-ADOPT	InteroperAble Descriptions of Observable Property Terminology
IARU	International Alliance of Research Universities
IAU	International Astronomical Union
IAWN	International Asteroid Warning Network
ICE	Infrastructure and Cloud research & test Environment
ICT	Information and Communication Technologies
IdP	Identity Provider
IETF	Internet Engineering Task Force
Ifremer	Institut français de recherche pour l'exploitation de la mer / French National Institute for Ocean Science
INORMS	International Network of Research Management Societies
IoT	Internet of Things
IP	Intellectual Property
IPR	Intellectual Property Rights
ISO	International Organisation for Standardisation
IVOA	International Virtual Observatory Alliance
JSON	JavaScript Object Notation
JUST	Judicious, Unbiased, Safe and Transparent
KPI	Key Performance Indicator
LERU	League of European Research Universities
LHC	Large Hadron Collider
LIBER	Ligue des Bibliothèques Européennes de Recherche / Association of European Research Libraries
MAR	Multi Annual Roadmap
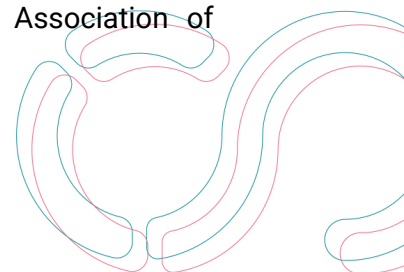
MIABIS	Minimum Information About BIobank data Sharing
MoU	Memorandum of Understanding
MS	Member States
MVE	Minimum Viable EOSC
NEO	Near-Earth Object
NetCDF	Network Common Data Form
NGI	Next-Generation Internet
NI4OS-Europe	National Initiatives for Open Science in Europe
NREN	National Research and Education Network
NSF	National Science Foundation
OA	Open Access
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OBM	OpenBioMaps
OCRE	Open Clouds for Research Environments
OECD	Organisation for Economic Cooperation and Development
OO	Operational Objective
ORCID	Open Researcher and Contributor ID
ORFG	Open Research Funders Group
OS	Open Science
OSPP	Open Science Policy Platform
PaNOSC	Photon and Neutron Open Science Cloud
PCP/PPI	Pre-Commercial Procurement / Public Procurement of Innovation Solutions
PID	Persistent Identifier
PPP	Public-Private Partnerships
PRACE	Partnership for Advanced Computing in Europe
PSI	Public Sector Information
R&I	Research and Innovation
R&R	Rewards and Recognition
RDA	Research Data Alliance
RDI	Research Data Infrastructure
RFO	Research-Funding Organisation
RI	Research Infrastructure
RoP	Rules of Participation
RPO	Research-Performing Organisation
SDG	Sustainable Development Goal
SIP	Strategic Implementation Plan
SLA	Service-Level Agreement
SLAC	Stanford Linear Accelerator Centre
SME	Small and Medium-sized Enterprise
SO	Specific Objective
SOLID	SOcial LInked Data
SPARC Europe	Scholarly Publishing and Academic Resources Coalition Europe
SRA	Strategic Research Agenda
SRIA	Strategic Research and Innovation Agenda
SSHOC	Social Sciences & Humanities Open Cloud
TCP-IP	Transmission Control Protocol / Internet Protocol
TU Delft	Delft University of Technology
USP	Unique Selling Point
VO	Virtual Observatory
VSNU	Association of Universities in the Netherlands
W3C	World Wide Web Consortium
WAI	Web Accessibility Initiative
WCRI	World Conferences on Research Integrity
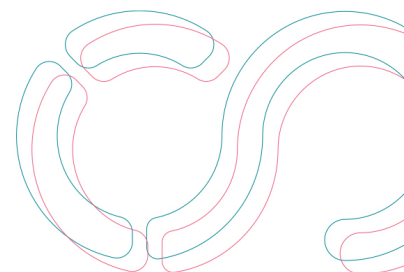
WDS	World Data System
WG	Working Group
WISE	WISE Information Security for Collaborating e-Infrastructures
XML	Extensible Markup Language
YERUN	Young European Research Universities Network
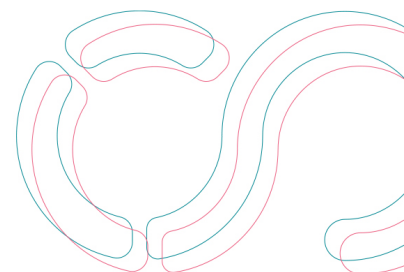
Glossary

As open as possible, as closed as necessary   Slogan of Open Science recognising that while the drive is for all research outputs to be open as far as possible and as soon as possible, there will be valid privacy, security, property and sovereignty reasons for not opening or for delaying the opening of research. For research artefacts, the interpretation of "as closed as necessary" is "as restricted as necessary".

Authentication and Authorisation Infrastructure (AAI)   Digital infrastructure that allows users to be identified and subsequently allowed to access data and/or services.

Co-programmed European Partnership   A contractual arrangement between the European Commission and private or public partners to realise together a common goal through financial and in-kind contributions.

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 international, national or institutional infrastructure that stores, handles and provides a level of access to (possibly FAIR and open) research data.

Data Stewardship   Support offered at all levels to researchers and institutions by expert data stewards with respect to the management of data.

e-Infrastructure   An international, 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 Advisory Groups The EOSC Association Advisory Groups are a structure to allow Association members and others to help steer the implementation of EOSC. The Advisory Groups provide an "umbrella" for a set of EOSC Task Forces that are highly related and have the same Liaison person from the EOSC Association Board of Directors.

EOSC Association   International Non-Profit Association (AISBL) founded in Brussels on 29 July 2020 to represent those (eligible) stakeholders wishing to formalise their role in EOSC. The Association signed a Memorandum of Understanding (MoU) with the European Commission and thus formed the Co-programmed European Partnership on EOSC.

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 Ecosystem   The encompassing set of federated (e-)infrastructures, research infrastructures, stakeholder organisations and projects that contribute to and/or use EOSC.

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 Executive Board An expert group reporting to the European Commission forming part of the interim EOSC Governance, comprising eight individuals

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

– 219 –

representing stakeholders and three independent experts tasked with the transition to the next implementation and governance stage of EOSC after their mandate ends on 31 December 2020.

EOSC Governance   The interim EOSC governance structure, comprising the EOSC Executive Board of experts and the EOSC Governance Board of national representatives, whose mandate ends on 31 December 2020, and thereafter the governance structure of the EOSC Association.

EOSC Governance Board   A working group reporting to the European Commission forming part of the interim EOSC Governance, comprising individuals representing Member States and Associated Countries of the European Union tasked with overseeing the transition to the next implementation and governance stage of EOSC after their mandate ends on 31 December 2020.

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 European Partnership on EOSC.

EOSC Task Forces   The Task Forces address key areas of implementation. They liaise with EOSC projects to offer feedback on developments, as well as identify strategic gaps and areas for investment to input to SRIA. An open call was held to define the membership of the Task Forces. This resulted in several hundred members of the community offering their expertise as volunteers to shape the future direction of EOSC.

EOSC Working Groups   Six Working Groups of experts from the EOSC ecosystem reporting to the EOSC Executive Board focusing on architecture, FAIR, landscape, sustainability, rules of participation and skills, which have delivered reports and recommendations for EOSC.
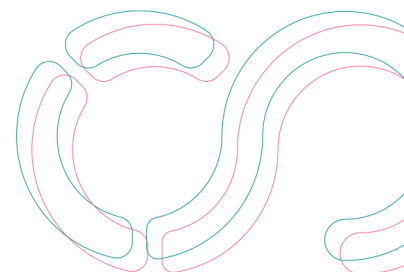
European Data Spaces   Nine data commons that will incentivise the sharing of data and facilitate the use of data across the areas of manufacturing, mobility, health, finance, energy, agriculture, public administration, skills and the Green Deal.

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.

European Research Area (ERA)   A system of scientific programmes integrating scientific resources of the European Union to enable researchers and scientific knowledge and technology to circulate more freely and to make Europe more competitive in its scientific goods and services.

FAIR Data   Data that comply as far as possible with the FAIR principles.

FAIR Metrics   A quantitative approach for evaluating the extent to which data have been made FAIR.

EOSC Association AISBL

Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu   - 220 –
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

eosc

FAIR (Principles)	The set of guidelines for making research (meta)data findable, accessible, interoperable and reusable that ultimately ensures standardised machine actionability.
Horizon 2020	The European Commission's eighth funding framework programme for research and innovation, which ran from 2014 to 2020.
Horizon Europe	The European Commission's ninth funding framework programme for research and innovation, which will run from 2021 to 2027.
Interoperability	The ability for data and metadata to be analysed and combined together with other data and metadata through shared standards.
Machine Actionability	The ability for data, through standardised metadata, to be read and understood by machines regarding the nature and use of the data.
Metadata	Information that is provided with data to explain the nature of the data and how the data could and should be exploited.
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.
Ontologies	Agreed vocabularies that provide formal names for objects and the relationships between those objects within and across scientific domains so that the nature and use of data are understood by machines.
Open Data	Data in an open format that can be freely used, reused and shared by anyone for any purpose.
Personal Identifier (PID)	A digital description that uniquely and persistently identifies an individual so that the correct digital object of research output can be linked to the correct researcher responsible for its creation.
Research Artefacts	Any product of the research lifecycle such as methodologies, data, software, publications, reviews and learning materials.
Research Infrastructure	An international, national or institutional infrastructure that enables research communities to perform research.
Research Outputs	Any product of the research lifecycle such as methodologies, data, software, publications, reviews and learning materials.
Rules of Participation	The terms, policies, processes and procedures required to provide assurance of sustainability, transparency, quality and trust in the practices and services offered through voluntary participation in EOSC.
Strategic Research and Innovation Agenda (SRIA)	An agenda giving an overall picture of the developments needed for the domain described. The EOSC SRIA is a set of recommendations from many EOSC stakeholders, edited by the EOSC Executive Board and handed over to 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 related services that researchers
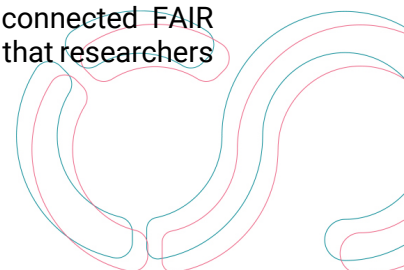
need to exploit these datasets for their research that are brought together and offered through EOSC.

Work Programme   A grouping of funding calls for, typically, a two- or three-year period within a funding framework programme.

EOSC Association AISBL