

**Strategic Research and Innovation Agenda
(SRIA)
of the
European Open Science Cloud (EOSC)**

Version 0.8 18 October 2020



**EUROPEAN OPEN
SCIENCE CLOUD**

Acknowledgements

This document has been written by the EOSC Executive Board [[EOSC EB](#)] and published by the EOSCsecretariat.eu project [[EOSC_Sec](#)], which supports the Executive Board in its activities. EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019, Grant Agreement No. 831644. The document develops, and incorporates feedback on, the Consultation document, which was for the most part based on the EOSC Partnership Proposal [[EOSC PP](#)] and drafts for the Strategic Research and Innovation Agenda for EOSC. Thanks are due to the members of the EOSC Executive Board Working Groups [[EOSC WGs](#)] for their contributions to this document.

Disclaimer

The European Commission is not liable for any consequences stemming from the reuse of this publication. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Table of Contents

| | |
|--|-----------|
| LIST OF ABBREVIATIONS | 7 |
| GLOSSARY | 11 |
| FOREWORD | 13 |
| SUMMARY | 14 |
| HOW TO READ THIS DOCUMENT | 26 |
| 1 NEW WAYS OF SCIENCE | 28 |
| 1.1. <i>The opportunity</i> | 28 |
| 1.1.1. Research in the digital age | 28 |
| 1.1.1. European leadership | 29 |
| 1.2. <i>The request</i> | 29 |
| 1.2.1. From Gutenberg to Berners-Lee | 29 |
| 1.2.2. Lindau Declaration | 30 |
| 1.3. <i>Open science</i> | 30 |
| 1.3.1. Brief historical context | 31 |
| 1.3.2. Open science facets: documents, data and software | 33 |
| 1.3.3. Open science adoption: progress and resistance | 36 |
| 1.3.4. Limits to open science | 38 |
| 1.4. <i>Next Generation Infrastructure</i> | 39 |
| 1.4.1. Learning lessons from the recent past | 39 |
| 1.4.2. Networking: the next-generation internet (NGI) | 42 |
| 1.4.3. Hardware: the computing continuum | 43 |
| 1.4.4. Software: Visualise, Analyse, Predict | 43 |
| 1.4.5. Data: findable, accessible, interoperable and reusable | 44 |
| 1.4.6. Machines for scientists: EOSC foundations | 45 |
| 1.5. <i>EOSC Objectives Tree</i> | 45 |
| 1.5.1. Open science practices and skills are rewarded and taught, becoming the ‘new normal’ | 47 |
| 1.5.2. Standards, tools and services allow researchers to find, access reuse and combine results | 47 |
| 1.5.3. Sustainable and federated infrastructures enable open sharing of scientific results | 49 |
| 2 SCIENCE AND DATA IN EUROPE | 51 |
| 2.1. <i>European Research Area</i> | 51 |
| 2.2. <i>Priorities of the new Commission</i> | 53 |
| 2.3. <i>The European strategy for data</i> | 53 |
| 2.3.1. Europe-wide common data spaces | 54 |
| 2.4. <i>Horizon 2020</i> | 56 |
| 2.5. <i>Horizon Europe</i> | 57 |
| 2.6. <i>International dimension</i> | 59 |
| 3 EOSC IN THE MAKING | 61 |
| 3.1. <i>Brief EOSC history</i> | 61 |

| | | |
|----------|--|-----------|
| 3.2. | <i>Transition period 2019-2020</i> | 62 |
| 3.3. | <i>National infrastructures</i> | 64 |
| 3.4. | <i>Strengthening the community</i> | 66 |
| 3.4.1. | <i>EOSC Association</i> | 66 |
| 3.4.2. | <i>Governance</i> | 67 |
| 3.4.3. | <i>Process</i> | 67 |
| 4 | GUIDING PRINCIPLES | 69 |
| 4.1. | <i>Introduction</i> | 69 |
| 4.2. | <i>Multi-stakeholderism</i> | 70 |
| 4.3. | <i>Openness: ‘as open as possible, as closed as necessary’</i> | 72 |
| 4.3.1. | <i>Open access</i> | 72 |
| 4.3.2. | <i>Trust in science through science reproducibility</i> | 73 |
| 4.3.3. | <i>Facing global challenges through multi-disciplinary programmes</i> | 74 |
| 4.4. | <i>FAIR guiding principles: making science transparent and reproducible</i> | 74 |
| 4.4.1. | <i>Web of FAIR Data and Related Services for science</i> | 74 |
| 4.4.2. | <i>Diversity of FAIR practices</i> | 75 |
| 4.4.3. | <i>Community standards</i> | 75 |
| 4.4.4. | <i>Research artefacts sustainability</i> | 76 |
| 4.4.5. | <i>FAIR metrics and certification</i> | 76 |
| 4.5. | <i>Federation of Infrastructures</i> | 76 |
| 4.5.1. | <i>First iteration – Minimum Viable EOSC</i> | 77 |
| 4.5.2. | <i>EOSC-Core</i> | 77 |
| 4.5.3. | <i>EOSC-Exchange</i> | 79 |
| 4.5.4. | <i>Federated data and services</i> | 79 |
| 4.5.5. | <i>Future Outlook</i> | 82 |
| 4.6. | <i>Open Science services: machines in support of people</i> | 84 |
| 4.6.1. | <i>Digital systems for Science</i> | 84 |
| 4.6.2. | <i>Hardware</i> | 85 |
| 4.6.3. | <i>Software</i> | 85 |
| 4.7. | <i>Recommendations</i> | 89 |
| 4.7.1. | <i>Recommendation 1: Fund awareness-raising, training, education and community-specific support</i> | 90 |
| 4.7.2. | <i>Recommendation 2: Fund development, adoption and maintenance of community standards, tools and infrastructure</i> | 90 |
| 4.7.3. | <i>Recommendation 3: Incentivise development of community governance</i> | 91 |
| 4.7.4. | <i>Recommendation 4: Translate FAIR guidelines for other digital objects</i> | 92 |
| 4.7.5. | <i>Recommendation 5: Reward and recognise improvements of FAIR practice</i> | 93 |
| 4.7.6. | <i>Recommendation 6: Develop and monitor adequate policies for FAIR data and research objects</i> | 94 |
| 5 | IMPLEMENTATION CHALLENGES | 96 |
| 5.1. | <i>Identifiers</i> | 96 |
| 5.1.1. | <i>Status</i> | 96 |
| 5.1.2. | <i>Gaps</i> | 97 |
| 5.1.3. | <i>Priorities</i> | 98 |
| 5.2. | <i>Metadata and ontologies</i> | 99 |
| 5.2.1. | <i>Status</i> | 99 |

| | | |
|----------|--|------------|
| 5.2.2. | Gaps..... | 100 |
| 5.2.3. | Priorities | 100 |
| 5.3. | <i>FAIR metrics and certification</i> | 100 |
| 5.3.1. | Status..... | 100 |
| 5.3.2. | Gaps..... | 102 |
| 5.3.3. | Priorities | 102 |
| 5.4. | <i>Authentication and authorisation infrastructure</i> | 103 |
| 5.4.1. | Status..... | 104 |
| 5.4.2. | Gaps..... | 104 |
| 5.4.3. | Priorities | 106 |
| 5.5. | <i>User environments</i> | 106 |
| 5.5.1. | Status..... | 106 |
| 5.5.2. | Gaps..... | 109 |
| 5.5.3. | Priorities | 111 |
| 5.6. | <i>Resource provider environments</i> | 113 |
| 5.6.1. | Status..... | 115 |
| 5.6.2. | Gaps..... | 117 |
| 5.6.3. | Priorities | 120 |
| 5.7. | <i>EOSC Interoperability Framework</i> | 124 |
| 5.7.1. | Status..... | 124 |
| 5.7.2. | Gaps..... | 125 |
| 5.7.3. | Priorities | 127 |
| 6 | BOUNDARY CONDITIONS | 130 |
| 6.1. | <i>Rules of Participation</i> | 130 |
| 6.1.1. | Status..... | 130 |
| 6.1.2. | Gaps..... | 130 |
| 6.1.3. | Priorities | 131 |
| 6.1.4. | Considerations..... | 132 |
| 6.1.5. | Impacts | 134 |
| 6.2. | <i>Landscape monitoring</i> | 135 |
| 6.2.1. | Status..... | 135 |
| 6.2.2. | Gaps..... | 135 |
| 6.2.3. | Priorities | 136 |
| 6.2.4. | Monitoring areas | 137 |
| 6.2.5. | Monitoring indicators..... | 138 |
| 6.3. | <i>Funding models</i> | 140 |
| 6.3.1. | Status..... | 140 |
| 6.3.2. | Gaps..... | 141 |
| 6.3.3. | Priorities | 141 |
| 6.4. | <i>Skills and training</i> | 142 |
| 6.4.1. | Status..... | 142 |
| 6.4.2. | Gaps..... | 143 |
| 6.4.3. | Priorities | 144 |
| 6.5. | <i>Rewards and recognition</i> | 149 |
| 6.5.1. | Priorities: | 149 |
| 6.6. | <i>Communication</i> | 149 |

| | |
|--|------------|
| 6.6.1. Priorities | 151 |
| 6.7. <i>Widening to public and private sectors and going global</i> | 151 |
| 6.7.1. Widening to public and private sectors | 151 |
| 6.7.2. Going global..... | 155 |
| 7 EXPECTED IMPACTS | 164 |
| 7.1. <i>Introduction</i> | 164 |
| 7.2. <i>Improved trust, quality and productivity in science</i> | 164 |
| 7.3. <i>Development of innovative services and products</i> | 167 |
| 7.4. <i>Improved impact of research in addressing societal challenges</i> | 170 |
| 8 RISK MANAGEMENT | 172 |
| 8.1. <i>Introduction</i> | 172 |
| 8.2. <i>Methods</i> | 172 |
| 8.3. <i>Results and recommendations for sustainable EOSC risk governance</i> | 175 |
| 9 CONCLUSIONS..... | 178 |
| REFERENCES | 180 |

List of Abbreviations

[In progress]

| | |
|--------------------|---|
| 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 | Appication 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 |
| CLARIN | Common Language Resources and Technology Infrastructure |
| COAR | Confederation of Open Access Repositories |
| CODATA | Committee on Data of the International Science Council |
| CORDIS | Community Research and Development Information Service |
| 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 |
| EB | Executive Board |
| EBU | European Broadcasting Union |
| EC | European Commission |
| ECTS | European Credit Transfer and Accumulation System |
| EEG | Electroencephalography |
| EERA JPWind | European Energy Research Alliance Joint Programme on Wind Energy |
| EIC | European Innovation Council |
| EMBL | European Molecular Biology Laboratory |
| ENVRI | Environmental Research Infrastructure |
| EOSC | European Open Science Cloud |

| | |
|-------------------|--|
| EPA | Economic Partnership Agreement |
| ePIC | Persistent Identifier Consortium for eResearch |
| EPOS | European Plate Observing System |
| ERA | European Research Area |
| ERC | European Research Council |
| ERIC | European Research and Innovation Committee |
| 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 |
| FaaS | Function as a Service |
| FAIR | Findable, Accessible, Interoperable and Reusable |
| FAO | Food and Agriculture Organisation of the United Nations |
| FDMM | [to be added] |
| FDO | FAIR Digital Object |
| 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 |
| GDPR | General Data Protection Regulation |
| GODCC | Global Open Data Commons Charter |
| G7 | Group of Seven |
| GIDA | Global Indigenous Data Alliance |
| H2020 | Horizon 2020 |
| HAL | Hyper Articles on Line |
| HBP | Human Brain Project |
| HEI | Higher Education Institute |
| HNSciCloud | Helix Nebula the Science Cloud |
| HPC | High-Performance Computing |
| HTC | High-Throughput Compute |
| I-ADOPT | Interoperable Descriptions of Observable Property Terminology |
| IAM | Identity and Access Management |
| IAU | International Astronomical Union |
| IAWN | International Asteroid Warning Network |
| ICE | Infrastructure and Cloud research & test Environment |
| ICOS | Integrated Carbon Observation System |
| ICT | Information and Communication Technologies |
| IdP | Identity Provider |
| IETF | Internet Engineering Task Force |
| IoT | Internet of Things |
| IP | Intellectual Property |
| IPA III | Instrument for Pre-accession Assistance III |

| | |
|---------------------|---|
| IPR | Intellectual Property Rights |
| IRUS-UK | Institutional Repository Usage Statistics UK |
| ISO | International Organisation for Standardisation |
| IVOA | International Virtual Observatory Alliance |
| JSON | JavaScript Object Notation |
| JUST | Judicious, Unbiased, Safe and Transparent |
| KNIME | Konstanz Information Miner |
| KPI | Key Performance Indicator |
| KRI | Key Risk Indicator |
| LHC | Large Hadron Collider |
| MoU | Memorandum of Understanding |
| MS | Member States |
| MVE | Minimum Viable EOSC |
| NDICI | Neighbourhood, Development and International Cooperation Instrument |
| 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 |
| OCRE | Open Clouds for Research Environments |
| OECD | Organisation for Economic Cooperation and Development |
| ORCID | Open Researcher and Contributor ID |
| ORFG | Open Research Funders Group |
| OS | Open Science |
| OSPP | Open Science Policy Platform |
| PaaS | Platform as a Service |
| 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 |
| RACER | Relevant, Accepted, Credible, Easy to monitor and Robust |
| RDA | Research Data Alliance |
| RDI | Research Data Infrastructure |
| RFO | Research-Funding Organisation |
| RI | Research Infrastructure |
| RoP | Rules of Participation |
| RPO | Research-Performing Organisation |
| RPN | Risk Priority Number |
| SDG | Sustainable Development Goal |
| SIP | Strategic Implementation Plan |
| SLA | Service-Level Agreement |

| | |
|-----------------|---|
| SLAC | Stanford Linear Accelerator Centre |
| SME | Small and Medium-sized Enterprise |
| SOLID | Social Linked Data |
| SRIA | Strategic Research and Innovation Agenda |
| SSHOC | Social Sciences & Humanities Open Cloud |
| SRA | Strategic Research Agenda |
| SWOT | Strengths, Weaknesses, Opportunities and Threats |
| TAIX | [to be added] |
| TU Delft | Delft University of Technology |
| USP | Unique Selling Point |
| VO | Virtual Observatory |
| W3C | World Wide Web Consortium |
| WAI | Web Accessibility Initiative |
| WDS | World Data System |
| WG | Working Group |
| WISE | WISE Information Security for Collaborating e-Infrastructures |
| WLCG | Worldwide LHC Computing Grid |
| XML | Extensible Markup Language |
| XSEDE | Extreme Science and Engineering Discovery Environment |

Glossary

[In progress]

| | |
|---|--|
| Data | An encompassing term used in the EOSC context for all digital outputs of research including datasets, metadata, publications and software code. |
| Data Infrastructure | An (inter)national or institutional infrastructure that stores, handles and provides a level of access to (possibly FAIR and open) research data. |
| e-Infrastructure | An (inter)national or institutional infrastructure that enables research through technical hardware and digital services (such as storing, computing or connecting) for sharing and exploiting research data. |
| EOSC Association | International Non-Profit Association (AISBL) founded in Brussels on 29 July 2020 to represent those (eligible) stakeholders wishing to formalise their role in EOSC. The Association intends to sign a Memorandum of Understanding (MoU) with the European Commission and thus form a European Partnership. |
| EOSC-Core | The basic architecture, standards and services that form the technical backbone of EOSC and are necessary to operate a Web of FAIR Data and Services. |
| EOSC-Exchange | The value-added services that will build upon the EOSC-Core and offer its users additional functionality to perform Open Science and share and exploit FAIR (and open) data. |
| EOSC Ecosystem | The encompassing set of federated (e-)infrastructures, research infrastructures, stakeholder organisations and projects that contribute to and/or use EOSC. |
| EOSC Governing Bodies | The current interim EOSC governance structure, comprising the Governance Board and Executive Board, whose mandate will end 31 December 2020, and thereafter the governance structure of the EOSC Association. |
| EOSC Partnership | The Co-programmed European Partnership between the EOSC Association and the European Commission that will consolidate the outputs of EOSC projects from Horizon 2020 and further develop EOSC through structured funding in Horizon Europe and in-kind contributions from the member countries and stakeholders. |
| EOSC Partnership Proposal | The published proposal for a Co-programmed EOSC Partnership [EOSC PP]. |
| 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. |

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

Foreword

[not included in V0.8]

Summary

[Work in progress]

The overall purpose of this document is to define the general framework for future strategic research, development and innovation activities in relation to the European Open Science Cloud (EOSC), to be further defined in the context of the candidate EOSC European Partnership proposed under the Horizon Europe Programme.

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

The Strategic Research and Innovation Agenda (SRIA) begins by describing the background to EOSC. It outlines the history and landscape of the digitisation of research in Europe (Section 1), the EC/EU policy context for open science and open data (Section 2), and the development and structure of the EOSC initiative (Section 3). The SRIA goes on to outline the guiding principles that are driving EOSC (Section 4), and the challenges and prerequisites to implementing the EOSC ecosystem (Sections 5 and 6). It discusses the anticipated benefits of EOSC (Section 7), before considering the importance of risk management to ensuring its sustainability (Section 8). The SRIA ends by drawing together the main points and conclusions (Section 9). A list of related documents is provided in Appendix A.

New ways of science

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, resulting in better science, increased trust in science, and the 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 deliver Europe's contribution to enabling 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.

While many scientists, innovators, research funders and policy makers around the world recognise the potential of digital technologies to transform the way research is conducted, not all scientists are convinced that the opportunities afforded by Open Science in all its facets – documents, data and software – are greater than its drawbacks. In particular, many influencers still argue that research artefacts should be kept closed as they are assets that research teams ought to keep for themselves if they want to stay competitive.

In addition, while Open Science practices have become a widespread reality, underpinned by policies, guidelines, incentives and roadmaps, these enabling technologies and policies, etc. are not yet fully deployed across all disciplines and countries, and are definitely not yet fully integrated and interoperable; scientific practices often still follow traditional patterns.

Moreover, there are acknowledged limits to Open Science with respect to privacy, security, property and sovereignty that will have to be respected in order for it to deliver value to society while mitigating undesirable consequences.

There are lessons to be learned from the evolution of digital services made available to researchers 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. These include trends such as the rapid expansion of deployment from user research communities to global societal uptake, the bottom-up development of standards, advancement through a combination of private and public funding and through collaborations between different communities (academic, public, private, commercial, non-profit), the importance of formalising semantics and of a design that allows the continuous exploitation of new technologies.

Other developments from which EOSC will benefit and to which it will contribute include the next-generation internet; a new paradigm called the Digital Continuum; artificial intelligence; the two principles of read-write capability and managed data accessibility, which have subsequently become enshrined in the FAIR guiding principles; and the foundational essentials of persistent identifiers, metadata and ontologies, and authentication and authorisation.

Building the European deployment of Open Science requires three main challenges to be addressed:

- 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 governing bodies have analysed these challenges, stating the problems, identifying the barriers, defining the objectives and highlighting the benefits, as summarised in the EOSC Objectives Tree (Figure 1.7).

Science and data in Europe

EOSC is an integral part of, and supports, the European Commission's strategy for realising the European Research Area (ERA), launched 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. In particular, EOSC helps deliver the policy priorities of Open Innovation, Open Science and Open to the World [[EC Open Vision](#)] and the goal of findable, accessible, interoperable and reusable (FAIR) data. The EOSC ecosystem will also be a central element supporting the revitalised ERA set out in a 2020 EC Communication [[EC COM New ERA](#)], with its four strategic objectives to prioritise investments and reforms in research and innovation towards the green and digital transition; improve access to excellent facilities and infrastructures for researchers across the EU; 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; and strengthen mobility of researchers and free flow of knowledge and technology.

It contributes to the six priorities driving the EC's work programme for 2019 to 2024, as outlined by Commission President Ursula von der Leyen [[UvdL Agenda](#)], especially 'A Europe fit for the digital age', and, through its focus on openness and interoperability, to the interlinking of the nine data spaces identified in *A European strategy for data* [[EC Data Strategy](#)].

EOSC can also play a key role in contributing to the specific objectives of the Horizon Europe framework programme and in supporting the implementation of its proposals for open and FAIR data [[Horizon Europe](#)].

There is also a global dimension to EOSC: comparable regional and national developments in Open Research Data Commons and/or Open Science Clouds will enable scientific cooperation throughout the world, while complying with a number of principles (e.g. regarding data portability, interoperability and security) that will ensure the competitiveness, transparency and quality of the international Open Science ecosystem.

EOSC in the making

The EOSC initiative is the tangible outcome of a number of key European and global policy and position milestones regarding Open Science, including the EC's *Open Innovation, Open Science and Open to the World – a vision for Europe* [[EC Open Vision](#)] communication, the establishment of the European Open Science Policy Platform (OSPP) [[EC OSPP](#)], and positions taken by the Group of Seven (G7) countries, the Lindau Nobel Laureates and UNESCO.

In the initial phase of development, from 2016 to 2020, funded by the EC through project calls in the Horizon 2020 framework programme, more than 35 projects have laid the foundations of EOSC, with a roadmap published in 2018 to direct its future strategic implementation [[EOSC Roadmap](#)].

To bring the community together and ensure a smooth transition to the second implementation phase (2021–2027) under Horizon Europe, a three-tiered transition governance structure was established to run from 2019 to 2020, comprising an Executive Board, Governance Board and Stakeholder Forum [[EOSC Gov](#)], supported by the EOSCsecretariat project [[EOSC Sec](#)]. The activities of the Executive Board, and of the six working groups it created, have been directed by a Strategic Implementation Plan [[EOSC SIP](#)] and a work plan [[EOSC Work Plan](#)]. The overarching objective of the Executive Board is to provide recommendations on the governance of EOSC in the second phase of implementation and to hand over all outputs to the new governance structure. This will take the form of a new legal entity, the EOSC Association, involving research and innovation stakeholders across the EU and beyond. The Association provides a means recognised by the EC to serve the EOSC community, promote alignment of EOSC contributions at all levels and support the Open Science development in Europe. It will be the focal point of the Co-programmed Partnership with the European Commission and this SRIA sets out the elements needed to implement the EOSC Partnership.

With regard to governance, the EOSC Association is open to any interested stakeholder organisations adhering to a set of principles. It is managed by three bodies: the General Assembly, the Board, and the Secretary General. In addition to those bodies, a Strategy Committee representing Member States and Countries associated to the Horizon Europe

Framework Programme will sit outside the Association. Its role will be to provide advice at policy and strategy levels. The coherence of the programme and the synergies (internal and external) will be ensured by the Secretariat of the Association.

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-related work programmes in Horizon Europe). The Association will also coordinate the identification of needs for the development of EOSC and will provide input to all relevant stakeholders, including the Commission. Regarding EOSC-relevant Horizon Europe work programmes, they will be adopted by the EC following relevant Horizon Europe comitology procedures.

Guiding principles

Based on the shared principles that have emerged during the evolution of EOSC, 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 and ensure EOSC helps researchers;
- **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’;
- **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 scientific 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).

In addition, the EOSC FAIR Working Group has identified the following recommendations [[WG FAIR Report](#)], which echo previous priorities identified in the ‘Turning FAIR into reality’ Expert Group report:

- Fund awareness raising, training, education and community-specific support;
- Fund development, adoption and maintenance of community standards, tools and infrastructure;
- Incentivise development of community governance;
- Translate FAIR guidelines for other digital objects;

- Reward and recognise improvements of FAIR practice;
- Develop and monitor adequate policies for FAIR data and research objects.

These recommendations have provided 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, in order to overcome the implementation challenges and realise the boundary conditions for deploying EOSC.

Implementation challenges

The EOSC governing bodies have identified fourteen action areas to help deploy the EOSC ecosystem. For each action area, the status has been assessed, gaps identified and priorities proposed. The seven areas relating to the primarily technical challenges and prerequisites to implementing the EOSC ecosystem 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. The priorities are:
 - Develop standardised identifiers for resource types that have not as yet become standard practice.
 - Develop a ‘meta resolver’ that can deal with any type of relevant identifier.
 - Define specifications (schemata) for persistent identifier (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.
- **Metadata and ontologies.** These have evolved organically over time, addressing the needs of individual communities and sub-communities. An overarching, coordinated approach is required, to ensure interoperability. The priorities are:
 - 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 federating/harvesting, crosswalks and tools for metadata management.
 - 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.
- **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. The priorities for FAIR metrics include:
 - Support the assessment and improvement of the RDA FAIR Data Maturity Model.

- Assess and test the proposed EOSC FAIR data metrics in a neutral forum.
- Support the definition of evaluation tools; their thorough assessment and evaluation including inclusiveness; comparison of tools; identification of their biases and applicability in many different contexts.
- Support the definition of FAIR for software and of the assessment framework for key elements of the FAIR ecosystem, in particular PID services and semantics.

The priorities for FAIR certification include:

- Support the current efforts to align certification schemas with FAIR.
 - Test the proposed schema in a variety of communities to gather feedback and update the proposed framework accordingly.
 - Support data and service providers to progress towards certification.
 - Support the establishment and maintenance of registries of certified components of the ecosystem.
- **Authentication and authorisation infrastructure (AAI).** The purpose of 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. Its 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. The priorities are:
 - 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 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.
 - **User environments.** User environments are the digital platforms users go to in order to interact with EOSC and EOSC resources. These include portals, dashboards, landing websites and, in general, services through which the EOSC resources are accessed and made useful to researchers. Throughout the distributed, federated and clustered architecture of the EOSC ecosystem, the user environments must meet the users requirements and expectations. The priorities include:
 - Discovery of EOSC and user environments: advanced discoverability of portals.
 - Discovery of resources: meta catalogues to aggregate information from the resource catalogues of the service providers, and open interfaces of catalogues.
 - Ordering, access and use: licences, usage terms and conditions, and user authentication and authorisation methods set by service providers; payment principles that are transparent and as easy as possible to use throughout the lifecycle of the research project.
 - Composing resources in a user environment: legal and organisational framework and its implementation in the distributed architecture.
 - Technical support: collaboration with service provider and local level support functionalities and resources; EOSC helpdesk functionalities.

- Community of practice of EOSC researchers: portals, other richer digital platforms and required supporting components providing necessary capabilities and capacity; interoperability with portals, thematic and regional community services and resources; science gateways for composability of the resources from different sources to generate new scientific outputs.
- **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. The priorities include:
 - Onboarding of resources: EOSC supply portal for service registration available; definition of standard validation criteria for EOSC compliance; automation via APIs.
 - Access to resources: resources support the EOSC AAI framework and the Interoperability Framework; EOSC Portal Service Catalogue assessment.
 - Composability of resources: semantic interoperability; implementation of standards and protocols; implementation of APIs.
 - Composability across resource providers: evolving framework for existing collaborations for EOSC; evolution of landscape and sustainability guidelines; sustainable financial model for EOSC resource provisioning.
 - Community of practice: resource provider forum established.
- **EOSC Interoperability Framework.** Achieving a good level of interoperability within EOSC is essential to federate services and provide added value for users, across disciplines, countries and sectors. The draft EOSC Interoperability Framework [[EOSC IF](#)] identifies general principles and organises them into four layers: technical, semantic, organisational and legal. The Framework also contains a proposal for the management of FAIR digital objects in the context of EOSC. The priorities include:
 - Technical level: use open specifications, where available; define a common security and privacy framework and establish processes for EOSC services; 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.
 - Semantic level: provide support for the maintenance of repositories of semantic artefacts, and governance frameworks for such repositories; define clear protocols and building blocks for the federation/harvesting of these repositories; support research communities so as to generate clear and precise definitions for the terms they use, as well as for their metadata and data schemas and their documentation; dedicate urgent, additional resources to communities with less developed or no community standards.
 - Organisational level: complete the current set of Rules of Participation recommendations with aspects related to interoperability for data providers and service providers.
 - Legal level: provide a list of EOSC-recommended licences and their compatibility with Member States' recommended licences to data producers, right-holders and users; 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; consider developing a centralised source of knowledge and support on copyright and licences to users and data generators and to address common Q&A.

Boundary conditions

The seven action areas relating to the social, financial, legal, educational, cultural challenges and prerequisites to implementing the EOSC ecosystem 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. The priorities include:
 - Provide standards for policy, processes and procedures that provide assurance of quality and trust in the services offered through EOSC.
 - 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, applicable to all digital resources made accessible via EOSC.
 - Establish a framework where the RoP can be owned, defined, maintained and enforced.
 - Make provision to evolve the RoP, e.g. to incorporate elements arising from the FAIR, Architecture and Sustainability Working Groups (WGs).
 - Elaborate and review the RoP's conceptual framework, e.g. relating to Terms and Conditions and Acceptable Use Policies, with respect to legal regulations.
 - RoP are about governance, oversight and authority. Without RoP, EOSC becomes no more than a search engine over an unmanaged collection of resources.
- **Landscape monitoring.** The Landscape Working Group has surveyed and documented the infrastructures, initiatives, investments and policies related to the development of EOSC in the European Member States and Associated Countries, as well as some border countries, as at the beginning of 2020. Sustainable long-term monitoring of EOSC landscape developments at national and institutional levels is required to keep the information gathered in the analysis up to date. Together with the development of respective national policies, supported by a set of relevant key performance indicators, this is required in order to allow informed decisions on EOSC. The priorities are:
 - Elaborate a thorough, sustainable monitoring methodology to define not only the criteria and indicators, but also purpose, process, tools, actors, responsibilities and actions. The monitoring should assess both the societal and the technical aspects of EOSC implementation readiness, covering the infrastructure, organisational and strategic landscape and the strategic outlook. The Landscape report provides a framework.
 - Ensure continuous monitoring of the existing readiness of countries to contribute to EOSC.
 - Suggest priorities for action based on the monitoring.

- **Funding models.** Viable funding models are an essential element of ensuring an operational, scalable and sustainable EOSC ecosystem. The Sustainability Working Group has taken an iterative approach to identifying funding models for EOSC, culminating in an 'Iron Lady' report published in October 2020. The report reflects the findings of a series of targeted studies commissioned by the WG, whose subjects included EOSC-Core operational costs and funding models for the full Minimum Viable EOSC. The priorities are:
 - Perform cost assessments for EOSC-Core services and MVE.
 - Ensure sustainable financing for EOSC by developing financing schemes.
 - Develop monitoring schemes for the in-kind contribution of members.
 - Develop synergies between national and EC funding streams as well as a higher level of coherence in the funding from different chapters of the Framework Programme, and across the three pillars of Horizon Europe.
- **Skills and training.** To realise the potential of EOSC for open and data-intensive research, Europe must ensure the availability of highly and appropriately skilled people with an excellent knowledge of standards and best practices for delivering, using, sharing and analysing open and FAIR data, and applications and tools. In alignment with the new ERA priorities on circulation of knowledge and the importance of skills and training [[EC-COM New ERA](#)], EOSC stimulates the development of a large talent pool equipped with the requisite skills, embracing a wide range of data-related profiles. It will contribute to shifting the culture of research towards openness and transparency, to building bridges between different disciplines and organisational models, and to approaching data literacy in various modes and settings, while working on existing initiatives and preconditions. The priorities are:
 - Develop the next generation of open science and data professionals by enhancing professional data career paths, developing data skills profiles, recognising data skills, providing a quality assurance framework, facilitating lifelong learning mechanisms and aligning Data Competence Centres.
 - Bridge the education gap: coordinate and align curricula for students and researchers.
 - Build a trusted and long-lasting knowledge hub of learning materials and related tools by developing a quality assurance and certification framework, devising a common framework for learning pathways, supporting the development of an EOSC Knowledge/Education Hub, facilitating the adoption of open learning environments, and promoting and supporting innovative ways of learning.
 - Develop an EOSC leadership programme to foster the right policy environment for data skills and training.
- **Rewards and recognition.** A culture change needs to be realised in order to increase the quality of 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. Future evaluation of scientists should have a better balance in valuing achievements in education; research; influence (on science and/or society and/or economy and/or teaching); organisation and leadership. In particular, evaluation and promotion criteria should recognise openness and FAIR practices. The priorities are:
 - Produce a country-level inclusive approach to research evaluation, taking Next Generation metrics into account.

- Discuss this approach within and between (all) the institutions in the country.
- Create interaction between the countries on this topic and learn from each other.
- As one of the organisations stimulating Open Science, EOSC should help in providing guidelines for adapting Rewards and Recognition systems aligned with the priorities outlined above.
- **Communication.** EOSC addresses not only researchers but also policy advisors, research funders and resource providers. Nine different stakeholder groups have been distinguished, which can be aggregated into three main categories: Research Service Providers, Research Performers and Research Funders [[EOSC Landscape](#)]. This diversity of stakeholders requires a communication policy that meets the different needs of these groups. EOSC should provide clarity on the why, how and what of EOSC, and should send out its messages in a consistent way. This will be focused on stakeholder engagement, content production, branding and the positioning of EOSC towards the different stakeholders. The priorities are:
 - Perform an in-depth stakeholder analysis.
 - Set up a Strategic Communication Plan.
 - Develop and deploy communication channels.
 - Develop stakeholder messaging that is impactful (addressing the why) and functional (addressing the how and the what).
 - Set a value statement and carry out an impact analysis.
- **Widening to the public and private sectors and going global.** EOSC will be widened beyond the European research community to public and private sectors, while also developing its global reach.

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 broader public and private sectors. 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. Key findings of the study include:

- 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.
- 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.

Recommendations include:

- 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.
- The INFRAEOSC-03 funded project should be used to initiate, implement or prototype, as appropriate, a series of recommended actions.

As noted above, there is also 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, bringing the potential to revisit long-standing scientific and societal challenges as well as to address new ones. For each of the strategic objectives defined in the EOSC Objectives Tree, the international dimension has been considered, priorities identified and deliverables defined. The priorities include:

- 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.
- Engage with the rapidly evolving global policy landscape of Open Science and support the creation of a policy observatory.
- Support the work being undertaken on methods, to complete FAIR – in particular interoperability – at a global level.
- Support the setting up of ‘Wise Persons global fora for Architecture’, to identify and remove the technical barriers that hinder the full potential of Open Science.
- 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.
- Direct Member State participation in the global research ecosystem, while maximising the added value of their bilateral international connections.
- 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.

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.

Expected impacts

The climate crisis, the extinction of species, global poverty and social inequality are only a few of the challenges that humankind has to face in the 21st century. Research plays a crucial role in facing 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 fill infrastructure gaps in unstructured areas, 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. The anticipated benefits of EOSC in the areas of Science, Industry and Society are as follows:

- Improved trust, quality and productivity in science:

- Encouraging collaboration and openness.
- Trusted frameworks for data availability and security.
- Infrastructure planning.
- Broadening discoverability.
- Making new connections.
- Addressing global challenges.
- Enhancing reproducibility.
- Development of innovative services and products:
 - Opportunities to improve support for researchers.
 - Opportunities to improve support for the private and public sector.
 - Opportunities to increase European leadership in open science and strengthen international cooperation.
- Improved impact of research in addressing societal challenges:
 - Research in society.
 - Supporting international collaboration.
 - Lifting science beyond the human scale.

Risk management

As part of ensuring the sustainability of EOSC, a targeted study has been conducted by AON Hewitt [[AON](#)] on behalf of the EOSC Sustainability Working Group in order to introduce clear and structured guidance on how to incorporate risk management into the governance of the EOSC Association. The study included a benchmark analysis, interviews, a SWOT analysis, and the development of a risk matrix.

While a number of risk governance gaps were identified, the study also found a human capital very rich in multi-disciplinary technical skills, sensitivity to governance issues, passion for the activities to be carried out and for the belief in EOSC itself. Moreover, the presence of all, or almost all, the essential pillars for the construction of effective risk management was found, together with a very high and mutual interest in and awareness of the importance of the subject among the EOSC major players, and that between them the stakeholders have the skills and experience required for effective risk governance.

The study sets out 32 recommendations to address the gaps and ensure the effectiveness of an EOSC risk governance. The main recommendations are: launch a comprehensive action plan; establish a governance structure for risk management; define risk policies; design the risk assessment and reporting process; map the skills and competences required; establish a risk awareness programme; set up an infrastructure and data security team; and improve technical resilience.

The identified gaps and recommendations are to be considered by the EOSC Association, the EOSC contributing projects and the EOSC partnership overall to develop a comprehensive risk governance, which in turn will significantly increase the value of EOSC and benefit its stakeholders by supporting its objectives and allowing a more effective use and allocation of capital and resources within the organisation.

Conclusions

[not included in V0.8]

How to read this document

[Work in progress]

While the overall purpose of this document is to define the general framework for future strategic research, development and innovation activities in relation to the European Open Science Cloud (EOSC), to be further defined in the context of the candidate EOSC European Partnership proposed under the Horizon Europe Programme, it contains a 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. Its core sections – those that relate most closely to planning of future research work and EC-funded projects – are highlighted in **blue**.

| If you are interested in . . . | Then please see . . . |
|---|--|
| An overview of the whole document | Summary |
| The history and landscape of the digitisation of research in Europe – the science-supporting technological context from which EOSC has evolved | Section 1 New ways of science, subsections 1.1 to 1.4 |
| The recent EC/EU/ERA policy context for open science and open data | Section 2 Science and data in Europe |
| The development of the EOSC initiative and its governing and legal bodies | Section 3 EOSC in the making |
| The objectives, guiding principles and recommendations that are driving and shaping EOSC | Subsection 1.5 EOSC Objectives Tree and 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 | Section 7 Expected impacts |
| The rationale and recommendations for risk governance in the complex multi-factor environment of EOSC | Section 8 Risk management |
| Other, 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

Terminology note

Readers are reminded that throughout the document, 'data' is used as an encompassing term referring to all digital research outputs, 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 subjects rather than in contradistinction to them, while 'scientist' refers to all researchers, academics and practitioners in all domains.

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 and landscape of the digitisation of research in Europe, establishing the technological context from which the European Open Science Cloud (EOSC) has evolved. It includes lessons to be learned, developments from which EOSC will benefit and to which it will contribute, and challenges that remain, together with the role EOSC will take in alleviating them.

1.1. *The opportunity*

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.

Better science

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.

Increased trust in science

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.

Meeting global challenges

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.

1.1.1. European leadership

When the Horizon Europe programme begins in 2021 [[Horizon Europe](#)], Europe will be well placed to lead the world in exploiting digital capabilities at the service of science. After three years of preparation, EOSC was launched in November 2018. More than 30 research and innovation projects have developed foundational technologies and initial services on top of which Europe can now build. These efforts have also allowed the establishment of a Europe-wide community that is now 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 fully operational as part of the Horizon Europe programme will enhance Europe's leadership position. Through the coordination and concentration of effort the 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

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 not all researchers are convinced that the opportunities it affords to them individually are greater than the drawbacks. To change this, establishing a new paradigm for rewards and recognition is essential: it can no longer be based on publications alone. Also, many leaders

within the research community still 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.

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. It will be officially signed by Nobel laureates and published during the 70th interdisciplinary Lindau Meeting, which has been postponed to 2021 due to the coronavirus pandemic.

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 prudence 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](#)]

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]

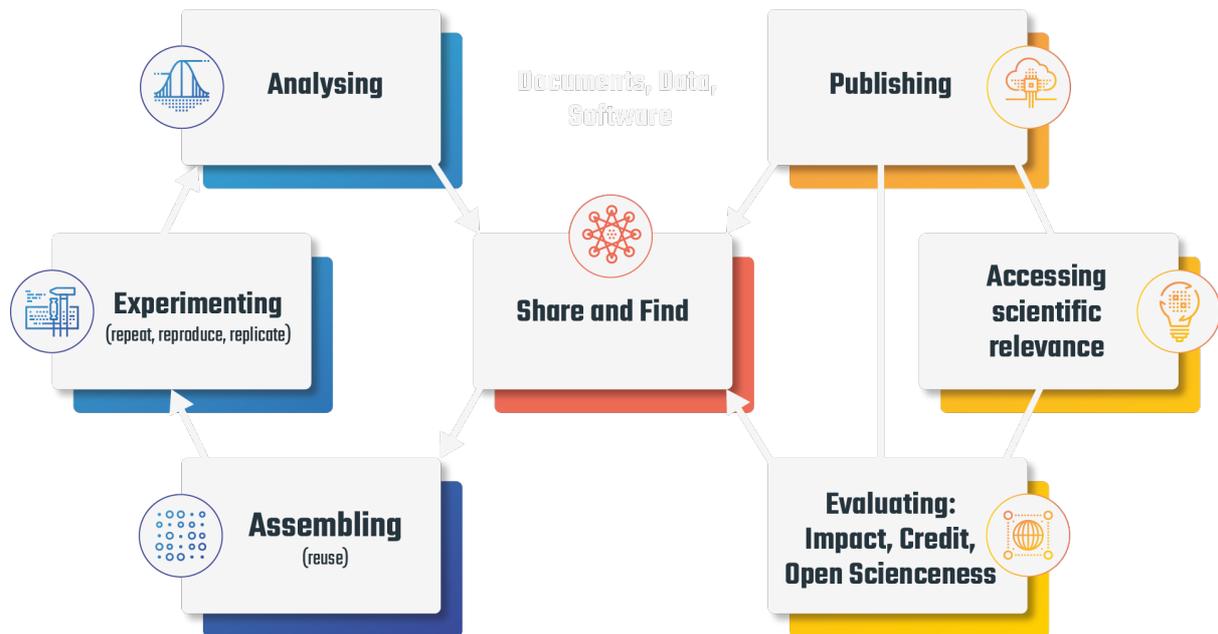


Figure 1.1: Research activity flows

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.’ [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 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 sociopolitical 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).

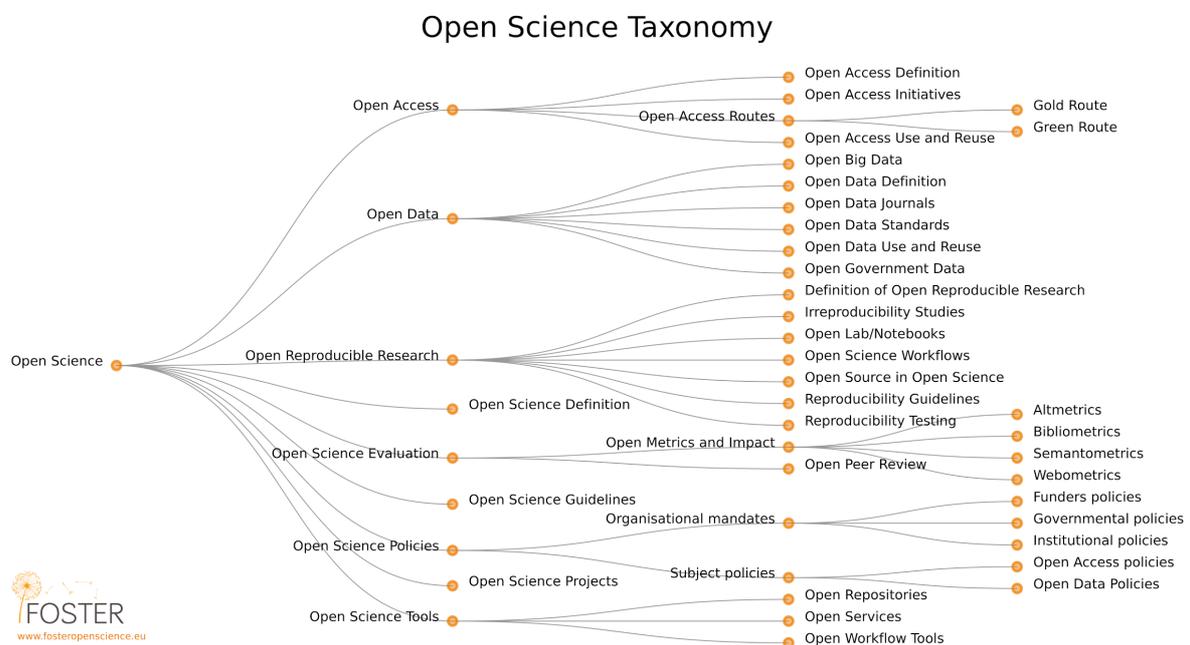


Figure 1.2: Open Science Taxonomy (from the FOSTER project)

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:



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.

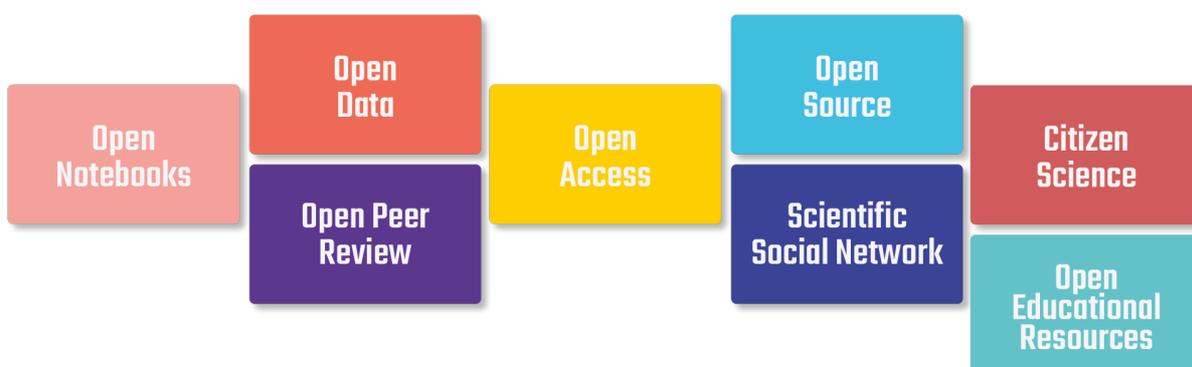


Figure 1.4: Open Science facets (from the FOSTER project)

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, however, which up to this point were mostly owned by publishing corporations. On February 14, 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 [[Software Heritage](#)] is an initiative launched by Inria, the French Institute for Research in Computer Science and Applied Mathematics 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.

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’. 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).

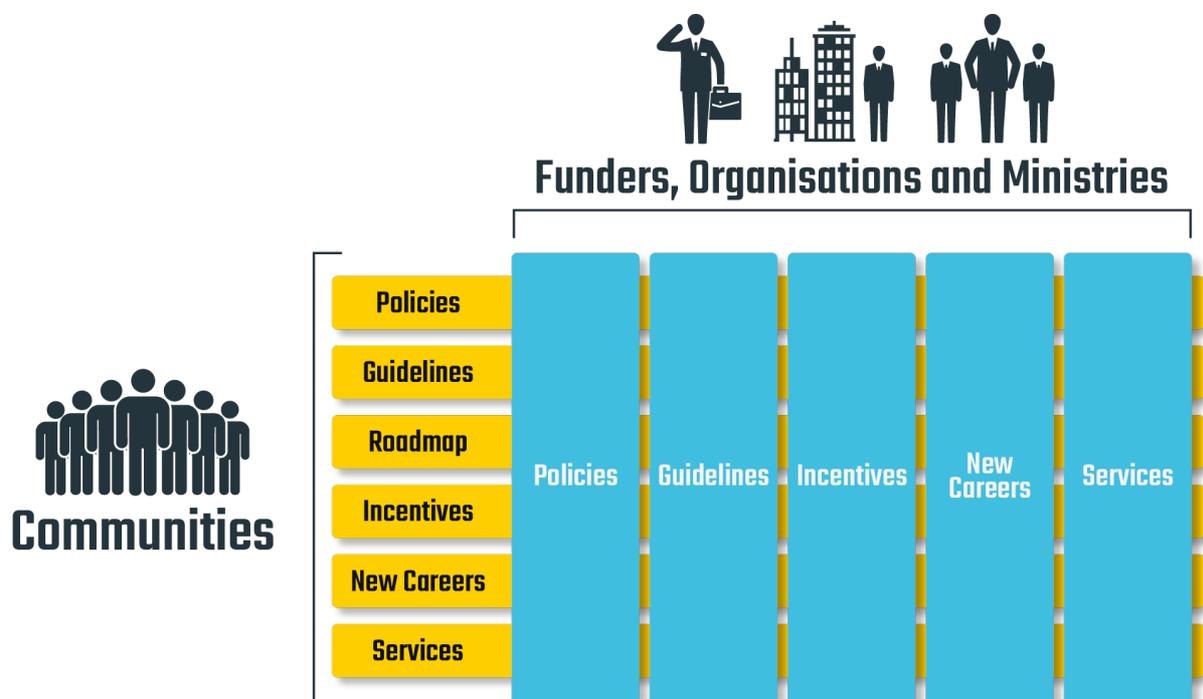


Figure 1.5: Open Science at the crossroads between communities and funders, organisations, and ministries

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 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’.

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.

| | 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, Startups | 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, Catalogs, Electronic Commerce | Encyclopedia, Social networks, Telephone, Digital Photography, Television | Elastic Cloud, Electronic Publishing, Videoconferencing | |
| Standards | IETF | X Window 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, Fiber Optics (network), Wi-Fi, REST, Git, Python | Smartphones, Fiber Optics (last mile), Hadoop, Cloud Computing, Grid Computing, GitHub | Digital Continuum, Internet of Things, Machine Learning |

Table 1.1: EOSC in its technological context

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 (eg mailing lists, file transfers, news) to assemble hundreds of computing and telecommunications scientists and design the internet architecture.

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 Tim 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. The Java and

JavaScript **programming languages** were developed by technology companies (Sun Microsystems and Netscape respectively). 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. Tim also requested that W3C would have multiple hosts and Inria, the French Research Institute for Computer Science and Automation, 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.

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 encyclopedia** 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 re-use of research data’.

1.4.1.6. 2020 Lessons learned

The emergence of so many new digital products and services, over the years, 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’, Berckman, 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 [Berckman HCC]. The **Computing Continuum** is described in Figure 1.6.

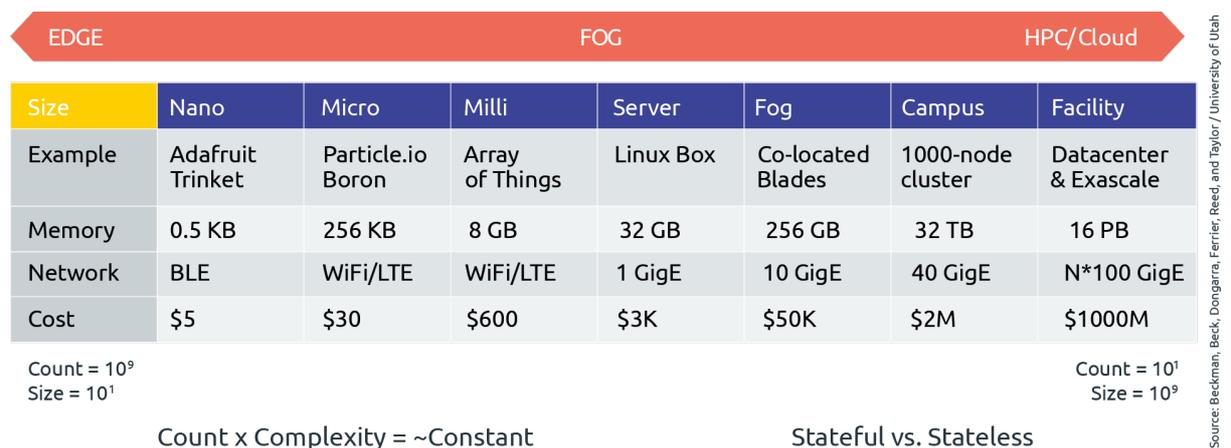


Figure 1.6: The Computing Continuum (from the EuroHPC SRA)

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.’ [EuroHPC SRA]

The EuroHPC 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 contribute to the Digital Continuum by providing the management of scientific information necessary for scientists to use the Computing Continuum and implement their ideas.

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 Horizon2020, 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, Tim 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.

Based on these foundations, once again, it will be possible to see 'thousands of flowers bloom'. Using the computing continuum, equipped with EOSC mechanisms, generic and specific services will be developed using current and novel technologies.

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.

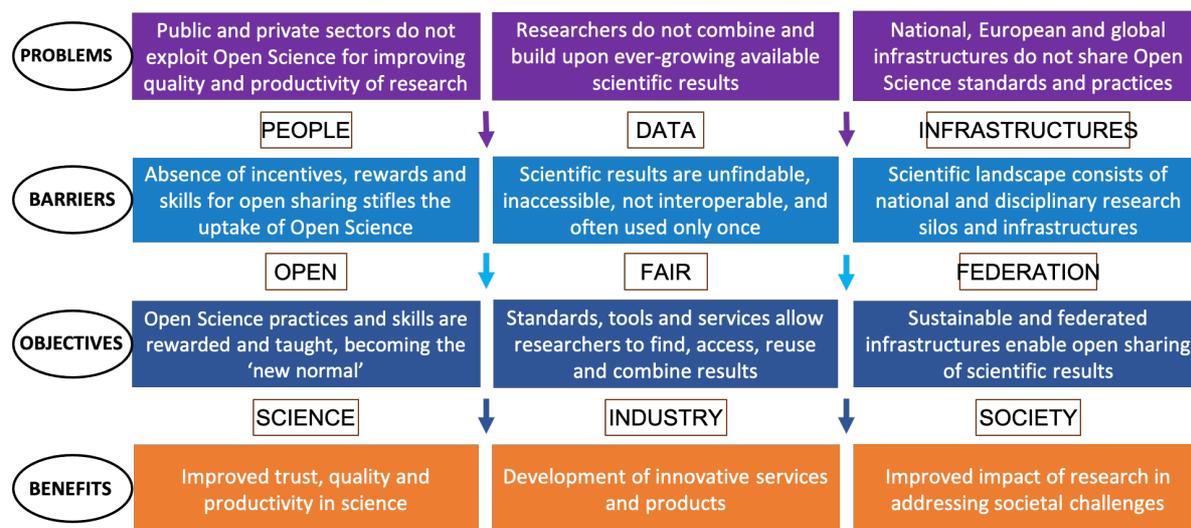
1.5. EOSC Objectives Tree

In summary of this introduction to EOSC, 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 1.7) presents these three challenges by stating the problems, identifying the barriers, defining the objectives and highlighting the benefits.

European Open Science Cloud Objectives Tree



v12

Figure 1.7: 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 objectives defined in its Objectives Tree, each of which is discussed below.

1.5.1. 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 the Europe-wide infrastructure for open research. The more scientists are convinced of the value of the EOSC federated infrastructure, the higher that value will be, following the network effect that led to the deployment and success of 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.

1.5.2. Standards, tools and services 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 (**‘knows 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 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.

1.5.3. Sustainable and federated infrastructures enable 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.

It 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.

Section 1 of this SRIA has presented an overview of the development of scientific practice, as influenced by changing ideas and evolving technologies, and has summarised the objectives that EOSC aims to achieve in order to address the challenges posed by such a contextual legacy. Section 2 of the SRIA provides further context, focusing on the recent, current and future status of and strategy towards science and data in Europe.

2 Science and data in Europe

The European Open Science Cloud 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 open science and open, FAIR data, and how EOSC will contribute to the EC's strategic objectives and priorities.

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 2018, the Council of the European Union made a call to revitalise the European Research Area, in 2020, with a new Commission Communication. In December 2019, Member States advised on the future of the ERA through an opinion of the European Research and Innovation Committee.

In April 2020, the European Commission introduced the ERAvsCorona Action Plan as part of the EU response to the coronavirus pandemic [[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](#)]. Based on excellence, competitive, open and talent-driven, the new European Research Area will improve Europe's research and innovation landscape, accelerate the EU's transition towards 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 and Youth)
on 30 September 2020

The Communication sets out strategic objectives and actions to be implemented in close cooperation with the Member States, in order to prioritise investments and reforms in research and innovation, improve access to excellence for researchers across the EU and enable research results to reach the market and the real economy. The Communication also 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.

The EU will work towards accomplishing the above strategic objectives, in close cooperation with the Member States, through 14 actions that are linked to each other and will be instrumental in realising the European Research Area [[EC ERA Actions](#)].

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); 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 [[UvdL Agenda](#)]. These priorities include:

- 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 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. This came out very clearly in the community consultation that took place in summer 2020 on the future Strategic Research and Innovation Agenda (SRIA). 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;
- 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.¹ 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. *Horizon 2020*

Horizon 2020, the 8th EU Framework Programme for Research and Innovation, has been essential to start implementing EOSC along the six action lines foreseen in the EOSC Roadmap (data, services, architecture, access, rules and governance) [[EOSC Roadmap](#)]. Actions of direct relevance to EOSC were included in the European Research Infrastructures (including e-Infrastructures) Work Programme 2016–2017 (€70m under the calls 'European Open Science Cloud for Research', 'Data and Distributed Computing e-infrastructures for Open Science' and 'Platform-driven e-infrastructure innovation') and Work Programme 2018–2020, (€250m under the dedicated call entitled 'Implementing the European Open Science Cloud'). The latter call aimed to help realise an EOSC that truly supports interdisciplinary research and Open Science, responds to the emerging needs of the scientific community and supports the whole research data lifecycle.

The proposed model supported by the call was to consider the development of an EOSC ecosystem and its pan-European service access mechanism – the EOSC-hub – providing all European researchers with seamless, non-discriminatory and secure access to public and commercial services and appropriate access modalities to a wider user community such as

In machine learning and artificial intelligence (AI), data form the **source** that determine the algorithms that will be formed. These data nowadays are often in silos for various reasons. However, access to interoperable data is becoming more and more essential. To achieve this, legal and commercial hurdles have to be eliminated. By proper data management (making data FAIR) visiting data in order to combine them can become the norm. Without such an approach many AI-applications will fail.

industry, public sector, citizen scientists, etc. The call was designed to support, amongst other things, the coordination between national initiatives aiming to make data FAIR, as well as the connection to EOSC of priority European research infrastructures (in particular the ESFRI ones) able to benefit from sharing a wealth of services and curated resources in a wide range of scientific domains. The call also supported the piloting of innovative financial schemes and/or pan-European joint procurements for storage, computing, software and other resources or services in close cooperation with funders in Europe, with the objective to seize opportunities arising from an aggregated demand by the researchers. Finally, the call also supported the setting up of an appropriate EOSC governance structure, taking into account the outcomes of previous efforts and the active contribution of all scientific stakeholders.

EOSC has driven a wide R&I agenda in Horizon 2020 since 2017. The call for proposals / grant approach used in Horizon 2020 has been successful in prototyping EOSC in its initial phase of implementation. It has 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 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 will not become available before 2022.

2.5. Horizon Europe

Horizon Europe, the next EU Framework Programme for Research and Innovation, is expected to be launched on 1 January 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 brings 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’;
- 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 identified in the EOSC Partnership Proposal, 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.

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 motivation by all Member States to establish an EOSC Co-programmed Partnership in Horizon Europe has been clearly expressed through the EOSC Governance Board since 2019. This reflects 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 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 is 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.

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 EOSC 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 will also enshrine a number of principles regarding international cooperation, with which potential partners 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.

3 EOSC in the making

The European Open Science Cloud initiative is the tangible outcome of a number of key European and global policy and position milestones regarding Open Science. This section outlines those milestones, together with the stages of EOSC's development, including its governance structure and activities during the transition period 2019–2020 and during the next, second phase of its implementation 2021–2027. It also summarises the landscape of national infrastructures and initiatives across Europe related to the development of EOSC.

3.1. *Brief EOSC history*

Open Science has been a policy priority of the European Commission (EC) 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 (ERA). 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.

The policy drive for Open Science is shared by the Science and Technology Ministers of the Group of Seven (G7) countries, who, also in 2016, established a G7 Open Science Working Group to share common international principles for Open Science [[G7 OS](#)]. The Lindau Nobel Laureates fully support the transition to Open Science in their Lindau Declaration of 2020 on Sustainable Cooperative Open Science [[Lindau Dec](#)] and UNESCO is developing an international standard-setting instrument on Open Science in the form of a UNESCO Recommendation on Open Science to be adopted in 2021 [[UNESCO OS](#)]. Several European countries are developing and implementing their own national policies for Open Science, such as France [[FR OS](#)] and the Netherlands [[NL OS](#)], and more are expected to follow. The current focus for many countries is on supporting open access to research publications and enabling researchers through findable, accessible, interoperable and reusable (FAIR) and open data.

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'. 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 2016 to 2020, the EC made a financial investment of approximately €350 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.² 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 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](#)]. Finally, 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](#)].

3.2. 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 serves the first implementation phase of EOSC in 2018–2020 under Horizon 2020, but also prepares for the second implementation phase of EOSC under the new funding programme of Horizon Europe for 2021–2027. The roadmap envisions 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,

² See Annex X of the landscape report for a list of EOSC projects funded under Horizon 2020 [[EOSC Landscape](#)].

consisting of eight members representing organisations and three independent experts, advises and supports 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, oversees and supports the activities of the Executive Board and ensures an effective implementation of EOSC [[EOSC GB](#)]. The Stakeholder Forum, consisting of the full EOSC community of organisations, projects and initiatives, allows 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 is supported by the EOSCsecretariat project, which not only functions as the governance secretariat, but also manages 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 is 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 is defining requirements for developing, assessing and certifying EOSC services in order to foster cross-disciplinary interoperability through FAIR [[EOSC WG FAIR](#)]. WG Landscape is mapping the landscape and readiness of existing research infrastructures in Europe that could be connected to EOSC [[EOSC WG Land](#)]. WG Rules of Participation is designing the rules to define the rights and obligations governing transactions between EOSC users, providers and operators [[EOSC WG RoP](#)]. WG Skills & Training is providing a framework for a sustainable training infrastructure to support the uptake of EOSC [[EOSC WG Skills](#)]. Finally, WG Sustainability is 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 are steered by a Strategic Implementation Plan (SIP), which defines 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 sets 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 is 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 consolidate all project outputs and ensure 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 transition governance bodies founded a new legal entity called the EOSC Association which will have as members all relevant stakeholders in the EOSC ecosystem and will enter into a contractual agreement with the European Commission to direct the Partnership under Horizon Europe. (Further information is provided in Section 3.4.)

3.3. *National infrastructures*

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 builds 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 include the 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 has 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 has 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 have policies in place regarding open access to scholarly publications, but only 34% have a policy in place regarding FAIR data (though, encouragingly, 44% have one either in planning 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; but only three countries so far (Bulgaria, Denmark, Romania) include mention of EOSC in their criteria for funding. More than half of responding countries have nominated contact points for Open Science (53%) and for EOSC (42%).

Taken together, sources show that there has been a significant investment across Europe in e-infrastructures and data-oriented infrastructures. For EOSC to reach its full potential, these investments 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 to now as part of national investment strategies, there is evidence that future policies and strategies will increasingly align around the concept. While it has proved hard to obtain definitive and quantitative data on national levels of investment, it is 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 applies to many of the data-intensive ESFRI landmark infrastructures. The scale of such investments over the past decade is certainly in the billions of euros and hence much larger than the planned central investment in the EOSC core. This shows 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. GÉANT).
- **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 landscaping exercise will continue until the end of 2020. In its entirety, the activity will indicate which infrastructures are 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 will make EOSC implementation gradual and dynamic. Links between the national thematic infrastructures and e-infrastructures, including data infrastructures, and national open access (OA) repositories, will be investigated. The readiness of the states will depend upon acceptance of EOSC.

3.4. *Strengthening the community*

3.4.1. EOSC Association

The **EOSC Association** was established on 29 July 2020 as a not-for-profit international association (AISBL), established in Belgium, involving research and innovation stakeholders across the EU and beyond. The Association provides a means recognised by the EC to serve the EOSC community, promote alignment of EOSC contributions at all levels and enable Open Science in Europe.

The Association will be the focal point of the EOSC Partnership with the European Commission and this Strategic Research and Innovation Agenda (SRIA) sets out the elements needed to implement the EOSC Partnership.

The 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). The Association is open to existing and newly developed organisations with shared values to those of EOSC from the public and private sectors. An open and inclusive European Partnership will help ensure directionality (common vision and objectives) and additionality (complementary commitments and contributions at all levels). It will help provide a framework to reach consensus amongst those committed to achieving results.

The EOSC Partnership aims to expand on the 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 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. EOSC has the ambition to be a virtual environment for Europe's 2 million researchers and 70 million science and technology professionals from every research, innovation and educational area. The Association provides a recognised European voice for this purpose for advocacy and representation for 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.

One of the primary tasks of the Association is to continuously develop the Strategic Research and Innovation Agenda, which shall influence future EOSC activities at institutional, national and EU level (including the EOSC-related work programmes in Horizon Europe). The SRIA Version 1.0, initiated by the EOSC Executive Board under Horizon 2020, accompanies the Memorandum of Understanding (MoU) between the European Commission and the EOSC Association, kicking-off the EOSC Co-programmed Partnership.

The Association will support 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

will identify 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 will enable 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 will coordinate and foster technical environments and promote the skills that enable the federation of existing and new scientific data infrastructures. The Association's Partnership with the EC and engagement with its stakeholders will allow it to maintain alignment between the operations sponsored by the Association and the EC's Open Science strategy.

3.4.2. Governance

The EOSC Association is open to any interested stakeholder organisations adhering to a set of principles:

- Decentralisation principle: the legal entity is as small as possible.
- Transparency principle: the statutes, bylaws, the membership contract and any other organisation document are public.
- Openness principle: any new member is able to join at any time, subject to compliance with simple rules of participation for new members.

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 most recent Framework Programme for research can join as full Members, while other organisations are able to participate as Observers.

From a governance perspective, the Association is managed by three bodies: the General Assembly, the Board, and the Secretary General. In addition to those bodies, a Strategy Committee representing Member States and Countries associated to the Horizon Europe Framework Programme will sit outside the Association. Its role will be to provide advice at policy and strategy levels.

The coherence of the programme and the synergies (internal and external) will be ensured by the Secretariat of the Association. While being as lightweight as possible, the Secretariat will primarily focus on technical (coordination) and communication roles.

3.4.3. Process

The Association will coordinate the identification of needs for the development of EOSC and will provide input to all relevant stakeholders, including the Commission. Regarding EOSC-relevant Horizon Europe work programmes, they will be adopted by the EC following relevant Horizon Europe comitology procedures. Calls for proposals 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. As a

principle, the Association as a legal entity would not bid to Horizon Europe or other calls for proposals where this would represent a conflict of interest or competition with its members.

Projects funded via actions defined in the Horizon Europe work programmes will contribute to EOSC implementation strategy by, for example, delivering services for the EOSC-Core and EOSC-Exchange, which form the Minimum Viable EOSC as referred to in the 'Solutions for a Sustainable EOSC' consultation document [[WG Sustain Report3](#)]. Services funded by other channels, for example by national funders and private sector providers, may also contribute to the MVE.

During the course of the Partnership and in implementing this SRIA, the Association will coordinate EOSC-related activities within its remit. It will focus on technical, communication and administrative roles. The technical role will essentially bring consensus and convergence in defining or contributing to the development and adoption of standards and good practices.

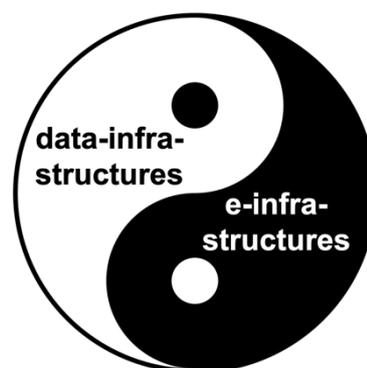
If needed, the Association will exercise or outsource operational responsibilities such as managing specific services. This role will be limited in size and avoid conflicts of interest with its membership. When the resources necessary to operate a service become too significant, responsibility will be either transferred or hosted by another entity in a transparent and cost-effective manner.

The promotional role of the Association will be based on multi-directional communication, actively supporting the users' engagement and feedback in all shapes and forms, as well as promoting EOSC results and success stories to convince new users of the added value of EOSC. The Association will also, on behalf of its stakeholders, communicate with the EC and society at large by speaking with one voice.

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 useless for 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. These form what can be called a 'yin and yang' relationship. One is not possible without the other. It is more or less a semantic discussion as to whether the whole of data-infrastructures and e-infrastructures should be called EOSC or whether EOSC is only a part of that.



The overarching principle is that research must be placed at the centre of the EOSC initiative. And thus, engagement with research communities is fundamental to understand their requirements and ensure that the way in which EOSC operates and the services offered help the researchers. 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 this 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';
- **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.

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 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 favour 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 European 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 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;
- International research centres;
- Etc.

A full and precise description of the EOSC stakeholders is provided by the EOSC Landscape report [[EOSC Landscape](#)].

The implementation of EOSC during its initial phase has largely focused on EU-level activities³ carried out through Horizon 2020 (H2020) projects. The involved consortia 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 accomplish it 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 take part in the development of the present strategy and have their voice heard.

³ Total EU investment of about €350 million in the period 2017–2020.

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);
- 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'*

The rise of the digital age allows the ways research is conducted to change in multiple dimensions, 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 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 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 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 science 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 of 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.

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, other products and services, 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 which 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 and 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'.

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 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. 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 objective of the first iteration is to bootstrap EOSC by establishing a Minimum Viable EOSC (MVE) such that 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.

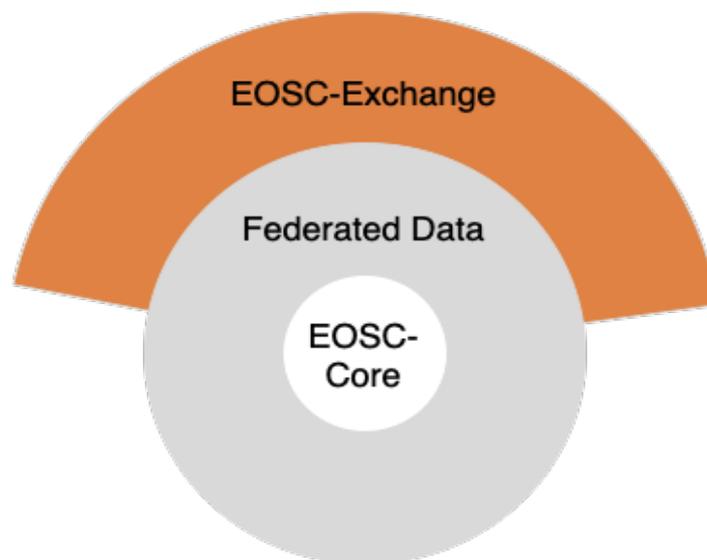


Figure 4.1: Schematic representation of key elements of the Minimum Viable EOSC

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 *EOSC-Core* will be constituted of widely used production-quality services, which are already deployed by the EOSC-related projects and communities, to provide the following functionality:

- **A shared open science policy framework**, which effectively embeds a data compliance framework for open and FAIR data. It defines and applies the rules of how the data elements are published, shared and reused.
- **Authentication and authorisation infrastructure (AAI) framework**, a trust and identity service for researchers to seamlessly access all EOSC resources⁴.
- **Data access framework**, whose primary role is to offer data as a service. It enables open interfaces where data consumers can discover and use data.
- **Service management and access framework**, whose role it 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.
- **A minimum legal metadata framework**, for ensuring openness and interoperability, privacy and security and all related legal and ethical issues (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), issues related to the General Data Protection Regulation (GDPR), etc.).
- **An open metrics framework**, which sets the rules (usage, performance, value for money) 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 awards 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](#)].
- **Persistent identifiers (PIDs)**, services to generate, resolve and validate persistent identifiers.
- **Helpdesk**, framework for linking national/thematic/institutional service desks that can provide training/consultancy on EOSC-Core services.
- **Portal**, an EOSC service implementing a web portal, which provides one form of access and use of EOSC resources.

The EOSC-Core will also include procurement services as well as 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. Additionally, during the study on the EOSC-Core operational costs [[EOSC-Core Costs](#)], two further candidate (component) services were identified:

- A collaboration and communication service (organisational);
- A messaging service (technical) that facilitates the interoperability of EOSC-Core services.

⁴ EOSC Resource extract from definition in the EOSC Glossary: *EOSC Resources include services, datasets, software, support, training, consultancy or any other asset* [[EOSC Glossary](#)].

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.

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 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 projects⁵ 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

⁵ [[EOSC Landscape](#)] – section 3.5.1.

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 there are efforts underway to establish national-level research data. 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.

The current offer of services and resources is managed by the EOSC Portal [[EOSC Portal](#)] via the EOSC Catalogue and Marketplace [[EOSC Marketplace](#)]. In that sense, the EOSC Portal serves as an 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 develop a rich platform offering a wide range of services and resources through the EOSC Portal, EOSC requires the participation of service providers. Services and resources are provided and maintained by different providers under the observance of the current legal frameworks and under a variety of licences and access requirements. (These include: accessible by users outside its original community; described through a common template focused on value propositions and functional capabilities; at least one service instance is running in a production environment available to the user community; production and publication of research data is FAIR; release notes and sufficient documentation are available; helpdesk channels are available for support, bug reporting and requirements gathering) [[EOSC Providers](#)].

In spring 2020, the resources listed in this catalogue are offered by 73 service/resource providers and aggregators in 13 categories: Aggregator, Analytics, Application, Compute, Consulting, Data, Networking, Operations, Other, Security, Software, Storage and Training. Taken together, they 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 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 sector⁶, 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 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.

The European Investment Bank published a report [\[EIB FFS\]](#) which included a section on EOSC and found that:

'The unique selling point (USP) of the EOSC is the magnitude of data in the context of the convergence of HPC, Big Data and machine learning.'

Enabling the private sector to make use of EOSC resources in such a manner 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.

⁶ 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 [\[EC Procurement\]](#).

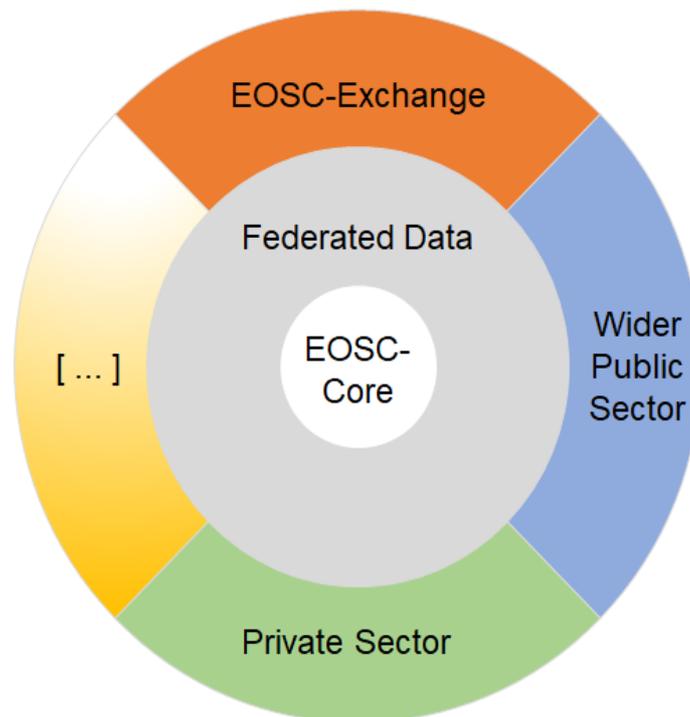


Figure 4.2: Pictorial representation of the relationships between the elements of EOSC

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*,⁷ 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

⁷ *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.

by step. This may involve revisiting and adjusting the datasets and e-infrastructures involved in the ongoing EOSC-related projects.

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 business 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 beyond the end of 2020 and extend it over the full length of Horizon Europe, as depicted in the graph in Figure 4.3 below.

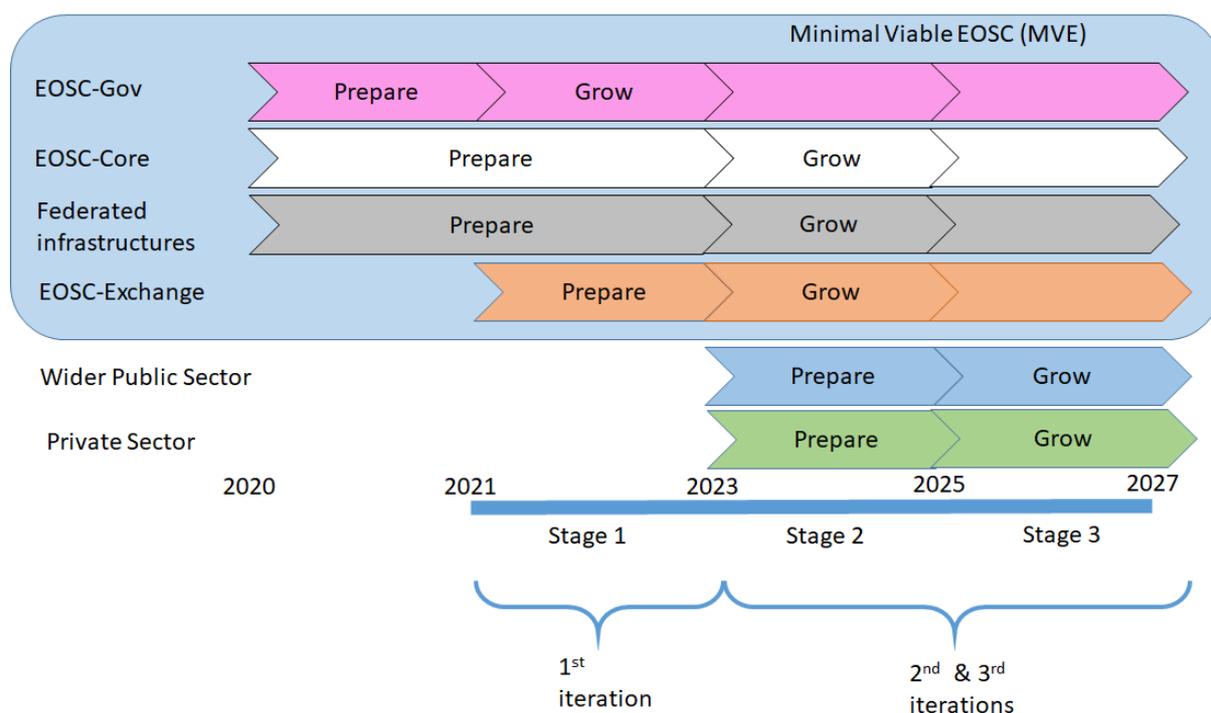


Figure 4.3: Schematic representation of timelines of EOSC iterations

Consequently, it is recommended that a transition period of 3 years (2021–2023) be anticipated to establish MVE, building on projects to be funded via INFRAEOSC-03-2020 and

INFRAEOSC-07-2020 funding calls, and by further means, such as open calls under the EOSC European Strategic Partnership, 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 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 (eg operating systems, programming languages environments), ... |
| Research Team | Computing & Storage Servers, | Databases, |

| Systems/Users | Hardware | Software |
|-------------------------|---|---|
| | Specific equipment, ... | 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 Equipments, ... | General Purpose Applications & Platforms, Very Large Repositories (Publications, Data, Software) |
| EOSC | Minimal Shared Resources | Federating EOSC-Core |

Table 4.1: Commonly used systems – of which EOSC offers an integrated view by federating existing infrastructures

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.

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 delivered a report covering the state of the art, best practices and open issues, workflows and use cases before offering recommendations for next steps.

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 4.4. Research software developments have created their own schemes.

Software schemes

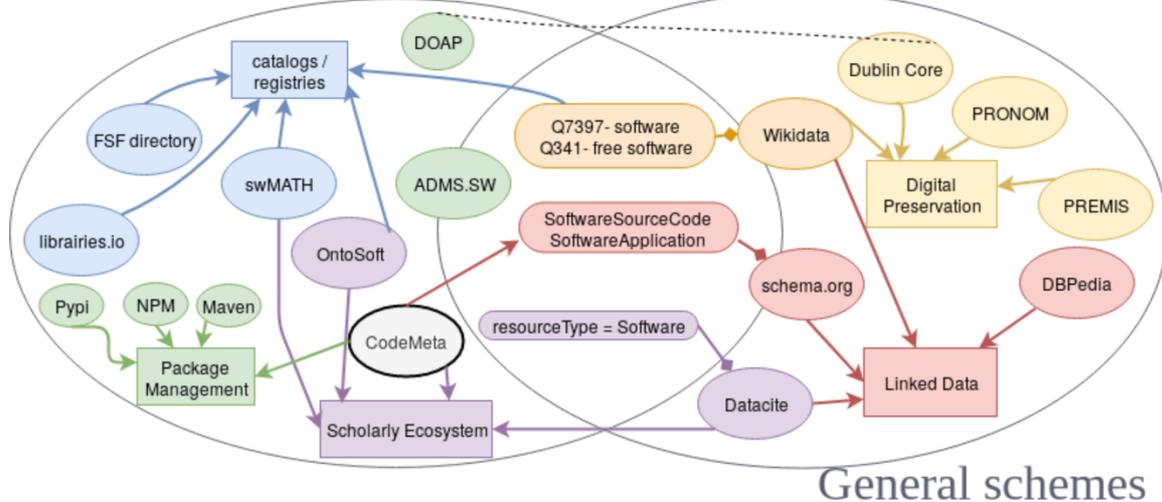


Figure 4.4: Metadata for software

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 4.5.

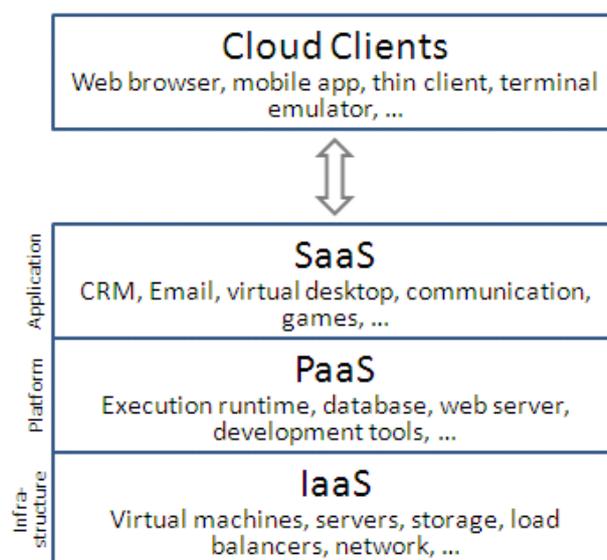
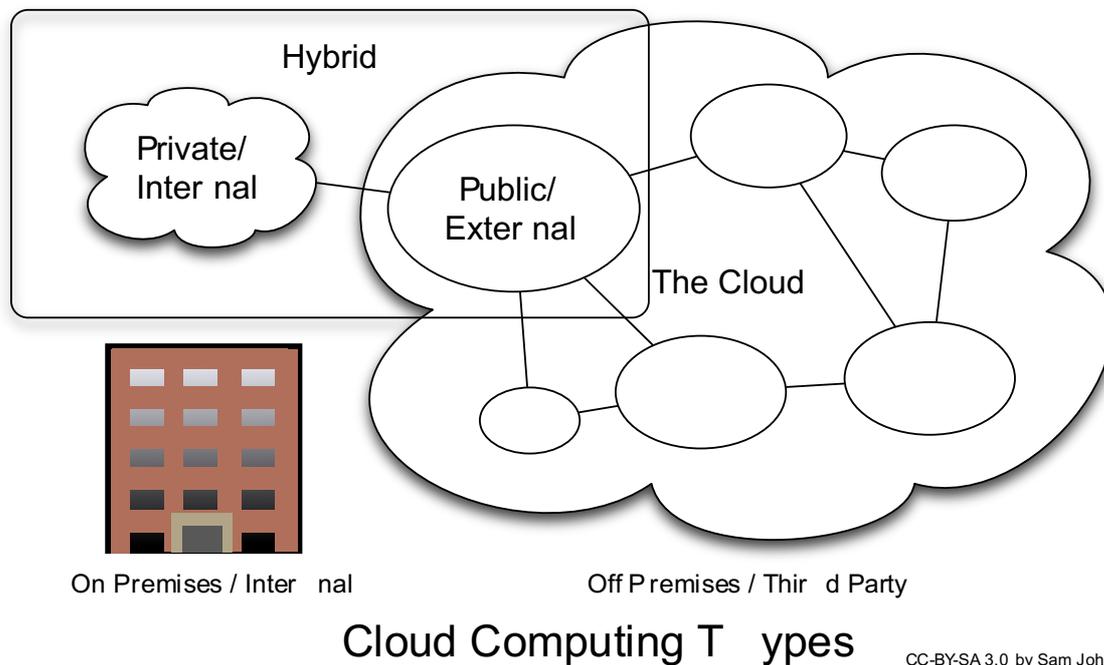


Figure 4.5: Cloud Computing Layers (from Wikipedia)

When designing a new Cloud offering, the way resources are shared can belong to Cloud Computing types as shown in Figure 4.6.



Cloud Computing Types

CC-BY-SA 3.0 by Sam Johnston

Figure 4.6: 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 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, EOSC-Core itself will be based on software services. In order to deploy EOSC services in a controllable manner, special attention has to be put into managing the software involved in implementing EOSC-Core functionalities.

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 scientific 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 open science 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 Report](#)]. 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:

- Recommendation 1: Fund awareness raising, training, education and community-specific support;
- Recommendation 2: Fund development, adoption and maintenance of community standards, tools and infrastructure;
- Recommendation 3: Incentivise development of community governance;
- Recommendation 4: Translate FAIR guidelines for other digital objects;
- Recommendation 5: Reward and recognise improvements of FAIR practice;
- Recommendation 6: Develop and monitor adequate policies for FAIR data and research objects.

| | 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 they apply to

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

4.7.1. Recommendation 1: Fund awareness-raising, training, education and community-specific support

| |
|---|
| <p>Stakeholders: EOSC, Research funders, Institutions</p> |
| <p>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.</p> |
| <p>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.</p> |

4.7.2. Recommendation 2: Fund development, adoption and maintenance of community standards, tools and infrastructure

| |
|---|
| <p>Stakeholders: EOSC, Research funders, Coordination fora, Standards bodies, Data service providers</p> |
|---|

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.⁸ 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.

Example: 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,⁹ 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. Recommendation 3: Incentivise development of community governance

Stakeholders: EOSC, Research funders, Coordination fora

Rationale: Standards need to be developed by/with the community for them to be accepted and successfully implemented. For this to happen, clear community

⁸ Those who successfully developed standards often cite years to ensure sufficient community consultation and co-development.

⁹ For examples of projects funded by the Wellcome Trust Open Research Fund, see [[Wellcome ORF](#)].

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).

Example: 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. Recommendation 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](#)].

4.7.5. Recommendation 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. Recommendation 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.

These 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 of them, status, gaps and priorities are highlighted; priorities are split into short, medium and long-term priorities.

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 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 Marketplace.
- **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.

- **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’, 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.

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.

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, projects or services, are 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 (but 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.

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.
- 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.

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.

5.3.1.1. 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.

5.3.1.2. 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.

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 underway 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.

5.3.3.1. Priorities for FAIR metrics

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

Priority 1.1: Support discipline 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 seek 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 Group, to seek global agreement with the international EOSC counterparts.

Priority 3: Support the definition 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 particular 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.

5.3.3.2. 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 schemas with FAIR.

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

Priority 3: Support data and service providers to progress towards certification.

Priority 4: Support the establishment of criteria and a methodology to certify other key elements of the FAIR ecosystem.

Priority 5: 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 a registry of registries.

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 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 AAI and infrastructure proxies can leverage eduGAIN [eduGAIN], the federation of national 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 R&E.

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 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 R&E

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

EOSC 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 startups/SMEs.

5.5.1. Status

Users, user groups and service providers have various expectations and requirements, such as:

- Seamless and easy access to resources: scientific services, research infrastructures, data and others;
- Spaces to share and reuse scientific data, including services for sensitive data;
- A networking and innovation environment to open new opportunities for collaboration;
- Interoperability with international communities beyond Europe;
- A high-bandwidth service for transferring data between distributed facilities;
- A solution for the long-term preservation of large quantities of open data;
- High-performance storage and compute resources for data analysis, accessible through cloud technologies such as containers, Function as a Service (FaaS) or Platform as a Service (PaaS) technologies;
- A federated search capability for searching and finding scientific products;
- A set of services for data simulation and analysis, ranging from generic services such as Jupyter Notebook to domain-specific applications per scientific application, including cloud and high-performance computing (HPC) resources;
- Service registration, helpdesk, monitoring and accounting;
- Clear exposure of conditions for accessing the resources.

These services will be provided from the distributed service architecture based on offerings contributed by the EOSC service providers. The list of expectations and requirements is a mix of those that are realistic and achievable in the short term and those that are more aspirational and long term.

Discovery of EOSC and user environments

In order to benefit from EOSC, users must be able to discover user environments, both through the central EOSC Portal and also through regional and thematic portals. 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.

Discovery of resources

Once users have discovered user environments, 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 within user environments.

Present discovery of resources at the EOSC level occurs primarily through the EOSC Portal [[EOSC Portal](#)] and, within it, through two lists of services (from the EOSC-hub and eInfraCentral projects [[EOSC-hub](#); [eInfraCentral](#)]) merged, from the user perspective, into a single list. However, they remain two lists in the background, with plans to merge them more fully in the immediate future. 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.

Ordering, access and use

Having selected resources of interest, users must be able to order them (directly or indirectly), access them with appropriate authentication and authorisation mechanisms, and then use them (either as part of the user environment or elsewhere).

Through the EOSC Portal, resources are listed, and three types of ordering and access are possible. Wide access / open access services that require no authorisation are linked, such that users click through to them. Other services that require ordering but which are not integrated with any central ordering system are linked and, on discovery, users must click through and order them directly from the provider. A third group, which is a small subset (generally resources associated with the EOSC-hub project), can be ordered via the Portal, generating a service request to the provider via either email or application programming interface (API), which the provider then fulfils.

Usage of all services occurs in the environment offered by the provider, rather than being embedded in the user environment.

Composing resources in a user environment

Beyond access and ordering of 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.

Technical support

Users will require support to assist with their use of resources, and it may not be clear to them where an issue lies, especially in the case of composed services. Hence, there must be a technical support function which assists users to deal with issues either with the user

environment or with the resources themselves. This means that the first line of support – for example, in campuses – must understand the problem scope and be able to communicate it to others. Each of the services should be well documented and be accompanied by an up-to-date tutorial.

The EOSC Portal offers a basic support system through the candidate EOSC-level helpdesk offered by EOSC-hub, which covers the Portal and core services of EOSC (AAI, the Marketplace, etc.). However, it is not connected to provider helpdesks, so if an issue is not with the Portal or core, it cannot be programmatically escalated to resource providers. However, when onboarding services, providers must list helpdesks (email or ticket based), which are exposed to users through the user environment on the Portal. As such, users can access support for both the Portal and core services and also for EOSC resources from the communities, but they are not interconnected.

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 user environments

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.

Discovery of resources

EOSC should 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 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.

Ordering, access and use

To all users, ease of ordering, access and use of services is essential. Current ordering provides some pilot indication of what integrated EOSC ordering may offer, but is not a mature solution. Ordering systems must be strengthened and the opportunities for providers must be clarified, as well as the work implied for providers, such that they take up these opportunities and users can benefit from integrated ordering. Where there are wide/open access services that do not require ordering, there must be a way to gauge uptake of services via EOSC to then show impact.

The licences, usage terms and conditions, and user authentication and authorisation methods set by service providers are key components of access to the services. The AAI principles of EOSC offer basic starting points for the technical implementations of EOSC services and set up the distributed service architecture.

Service payment methods and principles are essential parts of the usability of and access to services. In the distributed architecture, service providers define their own payment principles for various services and user groups, such as policy- or usage-based payments, freemium, etc. For users it is important that the payment principles are transparent and as easy as possible to use throughout the lifecycle of the research project.

The rules that apply to charging or payment from users must be clarified for end users and for the many providers who need to charge in some way, to ensure their services are available to users.

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 (EOSC-Exchange) supporting the full cycle of scientific workflows. EOSC itself has a role as a usability evaluator.

Technical support

Training and support of the open science principles, methods and technologies are essential parts of the success of EOSC. Technical support delivered by service providers of EOSC is a part of this.

As mentioned in Section 5.5.1 Status, there are technical support options in terms of helpdesks for the EOSC Portal and for resources, but they must be integrated where possible through some sort of ‘bus’, such that a user need not first have to diagnose where an issue lies. This will be challenging but is needed if composability is actively pursued.

More complex technical support exists locally but not yet in a distributed way or at an EOSC level. Technical support for integrating researcher-facing services with core services is needed, such that a user does not perceive the join between EOSC and its many service providers more than necessary.

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

To address the gaps and to prioritise proposed actions it is necessary to outline the vision of EOSC at a certain milestone. This will be done according to the following three phases:

- Phase 1: 2021–2023;
- Phase 2: 2024–2025;
- Phase 3: 2026–2027.

| Area | Challenge | Solution | Phase 1 | Phase 2 | Phase 3 |
|---|---|--|---------|-------------|-------------|
| Discovery of EOSC and user environments | Users should find those EOSC services and resources they need | Advanced discoverability of portals | First | Operational | Operational |
| Discovery of resources | Users discover resources from the service providers | Meta catalogues to aggregate information from the resource catalogues of the service providers | First | Operational | Operational |

| Area | Challenge | Solution | Phase 1 | Phase 2 | Phase 3 |
|---|--|--|---------|-------------|-------------|
| | | Open interfaces of catalogues | Second | First | Operational |
| Ordering, access and use | Ease of ordering, access and use of services | Licences, usage terms and conditions, and user authentication and authorisation methods set by service providers | First | Operational | Operational |
| | | Payment principles are transparent and as easy as possible to use throughout the lifecycle of the research project | Second | First | Operational |
| Composing resources in a user environment | Ensuring the position of the users and their work | Legal and organisational framework and its implementation in the distributed architecture | First | Operational | Operational |
| Technical support | Training and support of the open science principles, methods and technologies are essential parts of the success of EOSC | Collaboration with service provider and local level support functionalities and resources | First | Operational | Operational |
| | | EOSC helpdesk functionalities | Second | First | Operational |
| Community of practice of EOSC researchers | Sharing best practices across the community | Portals, other richer digital platforms and required supporting components such as distributed data, computing and storage providing necessary capabilities and capacity | First | Operational | Operational |
| | | Interoperability with portals, | First | Operational | Operational |

| Area | Challenge | Solution | Phase 1 | Phase 2 | Phase 3 |
|------|-----------|---|---------|---------|-------------|
| | | thematic and regional community services and resources | | | |
| | | Science gateways for composability of the resources from different sources to generate new scientific outputs | Second | First | Operational |

Table 5.1: User environments priorities by phase

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.

If EOSC is recognised as a system of systems,¹⁰ it means that it should be inclusive rather than selective, i.e. all metadata standards from communities are acceptable, all service framework standards (service pipe-lining and workflows, e.g. Galaxy, KNIME, Taverna, etc.) adopted by the communities are acceptable, etc. In particular, ‘basic’ EOSC participation should be at zero cost, while services can participate in EOSC with different degrees of engagement and cost of participation based on opportunities rather than obligations. (See also Section 6.1 Rules of Participation.)

An EOSC with only a few resource providers offering limited resources provides little added value to the European scientific community and to the two million researchers in Europe. The added value of EOSC exists only when many of the resource providers serving the scientific community can enter and offer resources. Therefore, for EOSC to be successful, it requires a low barrier to entry for resource providers to comply with the rules set for EOSC and to promote their resources through EOSC resource catalogues.

The vision for EOSC is to serve a wide variety of users and stakeholders (e.g. researchers, research infrastructures, service providers, service developers, funders, organisations, project managers, SMEs, citizens, etc.). It is to create a virtual environment that provides easy access to already existing resources and to allow EOSC users to build complex solutions out of a variety of resources. To embrace open science at its core, EOSC should stimulate FAIRness throughout the full research data lifecycle and should provide incentives for resource providers to support the paradigm ‘as open as possible, as closed as necessary’.

¹⁰ EOSCpilot Deliverable D5.1: Initial EOSC Service Architecture, Section 3.1. See [\[EOSCpilot_D5.1\]](#).

To allow sustainable resource provisioning within EOSC, next to a technical interoperability framework, non-technical frameworks must be set up for resource providers, on the levels of legislation (e.g. national and Europe-wide) and organisation, to allow the usage of resources ‘free at the point of use’.

In the context of this document, resource provider and resource provider environments are defined as follows.

Resource provider

A group or an organisation providing some type of resource (e.g. computing, storage, data sources and scientific products such as publications, research data, software, experiment reports, etc.). Typically, providers are from the wider community rather than from the core operations of EOSC.

Resource provider environments

These comprise the interface between the resource provider, a community and EOSC. This interface will be a framework of processes, tools, approved standards, APIs and other elements that enable resource providers to bring their resources into EOSC. Resource provider environments are therefore key to EOSC in that they are the way EOSC brings in the supply side from the broader community, from generic e-infrastructure services through to thematic services coming from the ESFRI clusters, research infrastructures (RIs) and potentially from other public domain and private sectors. Through this interface with EOSC, resources are then available to researchers, and are able to add value to European research.

Resource provider actions

Specific actions to be incorporated within resource provider environments to enable the inclusion of resource providers within EOSC are multiple:

- **Onboarding of resources.** To make resources (e.g. computing, storage, data sources and scientific products such as publications, research data, software, experiment reports, etc.) available through EOSC, resource providers must enrol their resources in an EOSC resource catalogue, either by an automated process or manually; comply with the EOSC Rules of Participation; and integrate where needed with the EOSC-Core components for monitoring, accounting, to collect usage and open science metrics, and to provide support.
- **Access to resources.** Access means the right of an end user to use a resource or a service offered by a resource provider as well as the way to reach it. This has technical implications, e.g. to authenticate and authorise users to access a resource, as well as non-technical administrative aspects, e.g. to clearly define access models, licences and/or usage terms and conditions under which a resource is made available.
- **Composability of resources** is the ability to assemble resources (e.g. computing, storage, services, data sources, datasets, publications or other research products) to build solutions by overcoming their heterogeneity and interoperability barriers. This may require the uptake and adoption of the interoperability frameworks and guidelines being produced by EOSC. This does not mean that EOSC should act as a new standards body, but instead should stimulate the use of interoperability frameworks and act as an enabler to support the adoption. Build solutions are in themselves resources that can be offered through EOSC.

- **Composability across resource providers** is the ability to allow sustainable (e.g. legal and organisational) resource provisioning across resource providers originating from different RIs, e-infrastructures, organisations and/or countries. This requires a legal and organisational framework within EOSC to be adopted by resource providers and Member States.
- **Community of practice.** To accelerate the uptake and adoption of the interoperability frameworks and guidelines being produced by EOSC, it is essential to build a community of practice [[Wikipedia CoP](#)] among the resource providers. The community of practice concept can also play a role in the development of the Rules of Participation, Interoperability and Open Science Framework for resource provider environments. This will support and stimulate the evolution of EOSC.

5.6.1. Status

Onboarding of resources

The EOSC Portal provides a mechanism for discovering the resource provider environments and requesting onboarding. Other paths are based on individual or personal contacts with members of projects such as EOSC-hub, EOSC Enhance, eInfraCentral, OpenAIRE-Advance and the thematic and regional cluster projects. These are all channels into an increasingly unitary onboarding process to achieve the basic listing of resources in the EOSC Portal and EOSC in a wider sense. Beyond this, the options for discovery are rather ad hoc, as these added-value options for resource providers are still being created and clearly established.

The EOSC resource provider landscape is highly distributed and diverse. Resource providers are distributed across all European Member States and are highly diverse, 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, providing core functionalities i) on the demand side, for researchers to discover, browse and order EOSC services and resources, and ii) on the supply side, for providers to onboard and register their resources (i.e., services) into a single EOSC Portal catalogue following standardised metadata, classifications and APIs.

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.

Access to resources

Although limited access has been facilitated through the EOSC Portal as a result of small commitments from the resource providers, access to resources for science is severely lacking. This is largely because the scientific landscape consists of fragmented and disconnected disciplinary research silos. An open and inclusive EOSC-wide solution for authentication and authorisation could address some of the challenges. At the moment, the majority of RIs follow a model similar to EOSC, which is based on the AARC BPA [[AARC BPA](#)]. However, allowing seamless access across resource providers from different RI and e-infrastructure domains remains one of the main challenges. (For further detail, see Section 5.4.)

In general, users from the target user community of a resource provider can authenticate themselves to a resource and therefore have access. Authorisation allowing access for users outside the targeted user community still has some significant barriers. Through federated authentication (e.g. eduGAIN), users can identify and authenticate themselves via their institution credentials. However, while users can authenticate, it does not automatically mean that a user is authorised to access a resource. Federated authorisation on the basis of user identity and/or on the role a user has within a community or organisation, or on the basis of attributes, is still challenging on both technical and non-technical levels. For example, harmonising the acceptable use policies¹¹ across resource providers and/or providing excellence-based access to HPC resources would be particularly useful for users outside the community to which the resource provider belongs but are very difficult to automate.

Within RIs and e-infrastructures, several AAI frameworks (IAM, EGI Check-in, EUDAT B2ACCESS, GÉANT eduTEAMS, CORBEL life science AAI, etc.) have been deployed. Within the EOSC-hub project, the initiative has been taken to integrate the AAI infrastructures from EGI, EUDAT, INDIGO and GÉANT, to exchange user identity and attribute information across these e-infrastructures. Within the EOSC Architecture WG a task force has been working on defining an EOSC AAI architecture (see Section 5.4).

The second aspect of access to resources is how to facilitate the interoperability and discoverability of services and resources. This is covered by the EOSC Portal Service Catalogue and is currently being developed by the EOSC Enhance project.

Composability of resources

In the last decades, research infrastructures 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-infrastructures services have been developed to provide bespoke solutions, some level of composability between RI and e-infrastructure services exists, for example between community workflows, HPC 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. From experience, adapting existing infrastructures developed for one community to be used by another community is non-trivial due to the underlying assumptions that are typically made to facilitate composability of resources. In general, the devil is in the detail. 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.

Composability across resource providers

In the last two decades, research communities and generic resource providers have organised themselves in many RI and e-infrastructure organisations, developing domain-specific

¹¹ See for example 'The WISE Baseline Acceptable Use Policy and Conditions of Use' [[WISE AUP](#)].

solutions to serve their mandated and targeted usage scenarios. The RIs and e-infrastructures are built out of resources providers with a common aim, to serve a specific research community (e.g. EPOS, CLARIN, ICOS, WLCG, etc.) and/or to harmonise and optimise service provisioning within a certain scope of services (e.g. PRACE, EGI, EUDAT, OpenAIRE, FENIX). The infrastructures have organised themselves autonomously via partnership agreements in developing and advancing the service and resource offering. Providing resources within the community domain or national boundaries is therefore, in general, allowed because of these agreements and/or because of legislation, since there is a clear mandate. However, when a user is from another community domain and/or another country, the situation is not always clear because of community and/or national, legal and/or organisational constraints. This introduced a non-technical interoperability gap to allow sustainable resource provisioning across RIs and e-infrastructures.

For a long time this complexity was partially covered by project participation agreements, which proxied for a clearer mandate or framework to support cross-domain or cross-border use of services. One of the main challenges of EOSC is to mitigate these limitations, moving beyond project-based 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.

Community of practice

As mentioned in the previous section, RIs and e-infrastructures have been organising themselves, building up communities of practice specific to their community, service domain and/or specific technologies or topics. The establishment of these communities of practice must be considered as one of the highest-valued outcomes of these partnerships and/or collaborations. Because EOSC consists of many projects, building communities of practice to specific technical and quality standards would facilitate several aspects of EOSC and give the users a real added value.

An early example of a community of practice is the WISE Information Security for Collaborating e-Infrastructures (WISE) community [[WISE](#)]. The aim of the WISE community is to define security policies and working practices that work across e-infrastructures and to inform each other about security incidents. The WISE community has established two frameworks, one on security policies [[WISE SCI](#)] and the other for risk assessments [[WISE RA](#)]. The frameworks are acknowledged by the participating e-infrastructure organisations (e.g. EGI, PRACE, EUDAT, WLCG, XSEDE, HBP and others). The WISE community is also collaborating on security awareness and on training.

5.6.2. Gaps

Onboarding of resources

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.

Concerning the discoverability of the resources and services, the EOSC Portal is still a work in progress. The usability of the EOSC Portal is crucial, as the Portal will represent and carry the reputation of EOSC in its entirety. An approach for measuring the usability of the EOSC Portal, whether for the users or the providers, is still missing. To provide incentives for resource providers to join EOSC, it is essential to prove that providers actually gain value from participating. Therefore, it is necessary for EOSC to offer tools to assess specific key performance indicators (KPIs) and increased benefit, e.g., user visits, increased ordering, recommendations to targeted communities, etc.

Concerning usage statistics, no specific recommendations have been made so far by EOSC, but standards exist and in some cases are widely used by the communities. For usage statistics the RDA Make Data Count initiative, the IRUS-UK network, and the OpenAIRE UsageCounts network are already operative and supported by publication and data repositories world-wide [[RDA MDC](#); [IRUS-UK](#); [OpenAire UC](#)]. While many of the referenced initiatives have a focus on data, to support Open Science, the focus should be expanded on gathering and monitoring usage statistics across all kinds of resources made available through EOSC.

Access to resources

A holistic identity and access management (IAM) system is lacking. Potential EOSC resource provider environments make different decisions on the basis of the local organisational, national, community and/or discipline-specific regulations or guidelines. Resource providers have existing infrastructures and services; these services are provided for a specific scope and purpose, using particular technologies. Therefore it is not easy to change or to modify them to achieve a common standard.

Providing a federated access model and infrastructure allowing seamless access across resource providers from different RI and e-infrastructure domains on the basis of either user identity, a role a user has and/or attributes is still one of the main challenges to solve on a technical and non-technical level.

Composability of resources

Due to the lack of an EOSC interoperability framework, or through having a choice of too many, the evolution of resources and technologies has been determined by the choices made by RI, e-infrastructure and resource providers addressing specific requirements. The

development of bespoke solutions drove the proliferation of the standards and APIs in use by resource providers, introducing technical challenges, and limiting the interoperability and reusability of resources within EOSC.

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.

In the EOSCpilot and EOSC-hub projects, and via the EOSC Architecture Working Group (e.g. AAI and PID Policy task forces), initiatives have been taken to harmonise and produce interoperability guidelines for EOSC-Core services (e.g. accounting, monitoring, helpdesk, etc.) and in the areas of an EOSC AAI and EOSC PID Policy.

These initiatives must be considered as a start that must be extended and evolve over time into an EOSC Interoperability Framework consisting of a rich set of guidelines to be adopted by resource providers across EOSC.

At this moment, EOSC is missing a forum that stimulates the definition and evolution of adopted interoperability frameworks beyond the community and infrastructure domains.

Composability across resource providers

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 3 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.

Community of practice

EOSC has been evolving and currently consists of about 30 projects. Building communities of practice for specific EOSC-related topics that span across EOSC is lacking; most of them are specific to a project or to a collaboration between projects. The EOSC working groups and the

task forces originating from these working groups must be considered as the exception. Building up these communities of practice should be one of the main focus points of EOSC. The communities of practice must be open for any resource provider participating or interested in EOSC. This is essential for the exposure and adoption of EOSC practices.

5.6.3. Priorities

5.6.3.1. Implementation timeframe and milestones

To address the gaps and to prioritise proposed actions, three phases of implementation are foreseen (Table 5.1), which align with the anticipated funding schedule.

| Phase | Period | Description |
|-------|----------------------|---|
| 1 | 2021–2023 | Creating the European Open Science Cloud operations (EOSC-Core) to provide authentication and authorisation infrastructure (AAI) and other necessary core functions of the Minimum Viable EOSC. |
| 2 | 2024–2025 | Expanding the Minimum Viable EOSC with access to added-value services, applications and tools (EOSC-Exchange) supporting the full cycle of scientific workflows. First pilots/demonstrators on linking EOSC beyond the research communities to the wider public sector and the private sector from 2024 onwards, for addressing societal challenges. |
| 3 | 2026–2027 and beyond | Deployment of the Web of FAIR Data and Services , including the EOSC-Core, EOSC-Exchange and other framework conditions for interoperability and machine actionability of data. Connection of the European contribution to a Web of FAIR Data and Services to other Open Science commons across the world. Continuous support to enhance the ecosystem of the Web of FAIR Data and Services for the research community. |

Table 5.1: EOSC implementation timeframe

The phases serve the following two major milestones [[EC Data Strategy](#)]:

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

The mapping of stages and milestones is shown in Table 5.2.

| | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 |
|--------------------|---------|------|------|---------|------|---------|------|
| Milestone 1 | Phase 1 | | | Phase 2 | | | |
| Milestone 2 | | | | Phase 2 | | Phase 3 | |

Table 5.2: Mapping of EOSC implementation phases and milestones

5.6.3.2. Alignment of priorities with the strategic objectives

Table 5.4 shows core objectives in the context of the actions identified for the resource provider environments, namely: onboarding of resources; access to resources; composability of resources; composability across resource providers; and community of practice. These objectives will be aligned with key aims and goals of EOSC as a whole within these phases to monitor and track progress.

| | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 |
|---------|----------------------------|------|------|-------------------------|------|--|------|------|
| Phase 1 | Automation & consolidation | | | | | | | |
| Phase 2 | | | | Auto-scale & resiliency | | | | |
| Phase 3 | | | | | | Best-in-class services and customer experience | | |

Table 5.3: Resource provider environments objectives by phase and year

Definitions

- **Automation.** To be able to scale out EOSC significantly, with a large number of resource providers, the EOSC-Core processes and composability between resources 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.
- **Consolidation.** EOSC should be a stronger organisation, therefore efforts should be consolidated to build this stronger organisation in which the resource providers are included.
- **Scale-out.** To attract users to EOSC by offering a rich portfolio, the resource portfolio should be scaled out, with more resources and capacity made available. From a resource providers' perspective this requires more on-demand resource provisioning and auto-scaling of resources, for which automation and consolidation are prerequisites.
- **Resilience.** To enable the scale-out of resource providers and resources, EOSC and resource providers need to adopt resilient and robust operational processes to reduce overheads in resource provisioning and incident and change management across resource providers.
- **Best-in-class services and customer experience.** This objective refers to accomplishing the full vision of EOSC as the main channel for resource providers to offer best-in-class resources for scientific research to all stakeholders in Europe, which goes beyond basic offering of resources to offering them in a robust, resilient, transparent and, as far as possible, in a self-service manner.

| Area | Challenge | Solution | KPI | Phase 1 | Phase 2 | Phase 3 |
|-------------------------|---|--|---|---------|-------------|-------------|
| Onboarding of resources | Distributed nature and diversity across RIs and e-infrastructures | EOSC Supply portal for service registration available | Ratio of new resources onboarded in self-service mode | 50% | 95% | 100% |
| | | Definition of standard validation criteria for EOSC compliance | Completion of the auto-validation standard | 20% | 80% | 100% |
| | | Automation via APIs | Coverage (functional and performance) of regression testing (CI/CD) | 10% | 30% | 100% |
| | | Resources automatically onboarded via thematic and regional compliant service catalogues | % of the new resources that are onboarded via thematic and regional service catalogue | 20% | 60% | 60% |
| | | Resources automatically publish usage statistics | % of resources automatically publishing usage statistics | 10% | 40% | 60% |
| Access to resources | Identity management compliance | Resources support the EOSC AAI framework | Amount of onboarded resources available through single sign-on | 40% | 60% | 80% |
| | Access management compliance | Resources support the interoperability framework for accounting and reporting | Automated accounting and billing | 50% | 100% | 100% |
| | Terms and conditions adaptations for interoperability | Interoperability framework for terms and conditions | | First | Operational | Operational |

| Area | Challenge | Solution | KPI | Phase 1 | Phase 2 | Phase 3 |
|----------------------------|--|--|---|---------|---------|-------------|
| | | Harmonise terms and conditions with EOSC Rules of Participation. | % of ToC from RP compliant with the EOSC RoP | 20% | 50% | 80% |
| | Automation | AAI and interoperability | | Second | First | Operational |
| | Usability of EOSC Portal for discovering resources | EOSC Portal Service Catalogue assessment | % increase of resources per year requested via EOSC Service Catalogue | 100% | 100% | 100% |
| Composability of resources | Semantic differences | Semantic interoperability | % of resources onboarded in the Research Product Catalogue compliant with Semantic Interoperability Framework | 10% | 40% | 70% |
| | Defined standards | Aspect of interoperability framework | Number of guidelines included in the EOSC Interoperability Framework | 10 | 20 | 30 |
| | | | Number of communities' guidelines included in the EOSC Interoperability Framework | 5 | 12 | 20 |
| | Use of APIs | Implementation of standards and protocols | % of resources that comply with 1 or more guidelines from the EOSC Interoperability Framework | 10% | 40% | 60% |

| Area | Challenge | Solution | KPI | Phase 1 | Phase 2 | Phase 3 |
|---|---|--|---|---------|-------------|-------------|
| | Use of different tools and services | Implementation of APIs | | Third | Second | Operational |
| Composability across resource providers | Scalability constraints of project-based collaborative agreements | Evolving framework for existing collaborations for EOSC | | First | Operational | Operational |
| | Limitations due to national and research mandates | Evolution of landscape and sustainability guidelines | Number of Member States and/or ERICs that have accepted the EOSC Rules of Participation | Second | First | Operational |
| | Financially sustainable solutions | Outcome of sustainability and interoperability framework guidelines | | Second | First | Operational |
| | | Project-based VA model converted to sustainable financial model for EOSC resource provisioning | % of resource providers in EOSC Service Catalogue compliant with new financial model | N/A | 10% | 50% |
| Community of practice | Building the community | Resource provider forum established | Number of resource provider communities formed | 4 | 8 | 12 |

Table 5.4: Resource provider environments priorities by phase

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 handling 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 minimal 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 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 are 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.

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**.
- **Consistent, machine-readable consent agreements** to ensure permissions to access or use data are clear and regulations such as GDPR 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, 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.

6 Boundary conditions

The remaining seven action areas identified by the EOSC governing bodies to help deploy the EOSC ecosystem relate to the social, financial, legal, educational and cultural challenges and prerequisites to its implementation. They 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, again providing an assessment of status, identifying gaps and proposing priorities.

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) 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 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 provide 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 of 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 provide 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

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 it.

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 available, quality may be the value-defining characteristic. Here, transparency and community governance is 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).

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, such as those provided by the RDA WG 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.

6.1.5. *Impacts*

The priority is for the RoP to help build trust and confidence in EOSC in the phase 2021–23. This applies to services provided through EOSC and to research outputs produced from use of EOSC resources. The activities can be grouped into three main categories:

- Consistent and wide (or increasing) application of the FAIR principles.
 - Requiring data management planning and documented standards and protocols for data sharing and reuse, including accessibility, through EOSC.
 - Encouraging adherence to an open charter for datasets incorporating rich, community-defined and FAIR metadata, including granularity levels, versioning, provenance, sustainability commitments, quality goals and licensing terms.
 - Encouraging research publications produced to be open access.
 - Encouraging research data outputs to be as FAIR as possible and ideally open.
 - Requiring use of EOSC-recognised PID services.
- Harnessing the potential of RDIs more fully and effectively.
 - Requiring compliance with the EOSC service description template and onboarding processes.
 - Encouraging RDIs to work towards certification, with community standards, helping to produce FAIR outputs.
 - Encouraging compliance with the EOSC Interoperability Framework to ensure services achieve a minimum level of interoperability (semantic, legal, technical and organisational).
 - Encouraging FAIR certification of repositories (e.g. with CoreTrustSeal).
 - Standardisation of access policies.
 - Encouraging adoption of standards.
 - Requiring compliance with the EOSC AAI framework for managing user identity and access.
- Encouraging open science skills, recognising and encouraging practitioners, and rewarding open science.
 - Encouraging the growth of new professions of research supporters and EOSC enablers by recognising certificates.

- Requiring professional certificates as part of services' EOSC-approved accreditation.
- Encouraging provision of professional data stewards to support research activities.
- Encouraging recognition of open science activities in research career assessments.

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 related to the development of the European Open Science Cloud, as at the beginning of 2020. The resulting report 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 (61%) of these policies 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.

6.2.3. Priorities

It is important to elaborate a thorough 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 (RPOs), research infrastructures (Ris), e-infrastructures, etc.), up to the monitoring of the overall environment of the national landscape (national policies and strategies, research-funding organisation (RFO) 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.

[Priorities from Open Consultation document:

- 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.
- 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.]

6.2.4. Monitoring areas

The Landscape report structure provides a framework for future monitoring exercises and for the structure of the areas to be monitored. Monitoring should cover:

A. The infrastructure landscape

- a. The 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. The description of the supporting infrastructure(s) contributing to Open Science targets at the national or regional level (e.g. universities, public and private RPOs, thematic infrastructures, etc.).
- c. The overall description of the remaining research environment relevant to EOSC, including the private sector.

B. The organisational landscape

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

C. The strategic landscape

- a. The description of the institutional structure(s) at national level accountable for defining and implementing EOSC-related policies and strategies, including their hierarchical structure.
- b. Description of the EOSC-related policies and strategies with direct and indirect impact on EOSC.
- c. 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. Monitoring 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. 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 in Section 6.2.2, 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.

6.2.5. Monitoring indicators

The monitoring should focus on the full data lifecycle. It might even be worth considering making it an obligatory component of the national progress reporting. In order to do so, a set of KPIs shall be defined by an expert team and tested, through a pilot project, on a selected and representative set of countries, enabling transparent and unbiased assessment.

KPIs provide valuable information both for the operators of RIs and for their stakeholders to optimise progress towards objectives through changes in inputs and activities. The KPIs need to reflect the varying needs of different stakeholders and scientific communities. The indicators should be relevant, accepted, credible, easy to monitor and robust (RACER), and accompanied by a reference sheet that provides a definition, data source(s), method of calculation, and other information concerning calculation or applicability. It should be mentioned that KPIs are not suitable for a comparison of the performance of RIs. Their purpose is to indicate the dynamic development, to serve as a warning system on which actions and decisions can be based, and facilitate lessons learned for possible changes in future. The entire process aims to ensure quality. The purpose of KPIs is to measure performance (efficiency of used resources), successful delivery of results, and progress towards set goals.

KPIs should be used within a RI to maintain a standard, rather than across RIs to increase competition. For this reason, KPIs should look within the infrastructure rather than across infrastructures, to monitor progress. Data collection, storage and preservation play an essential role here. The development of KPIs that may be applied effectively to the many, diverse RIs across Europe is not straightforward. Some of the KPIs are relevant for multiple RIs; others should be tailored to the specific objectives and missions of each RI. It is also important to consider the level of maturity and the scientific domain of the RI. KPIs should be defined as a result of a dialogue with the RI. The RIs and their stakeholders should consider applying the proposed objectives and KPIs for their own monitoring purposes. KPIs could be linked to impacts, including indirect impacts, e.g. missions, sustainable development goals (SDGs). Sharing best practice and lessons learned in refining KPIs and to define new KPIs helps to establish the monitoring system at RI level.

Operational costs should be taken into account as an important element for success in the future. Early operational costs are often absorbed by the construction phase. It is crucial to set up a cost book with the risks.

Once established, these KPIs have to be regularly updated, either in self-evaluation, or by a trusted entity or committee.

Examples of KPIs based on objectives

- Existence of standardised national Open Science and FAIR data strategies, including the description of these policies (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Existence of central/national contact point for Open Science (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Existence of national policy on Open Access publishing and Open Access to publications, and if YES, does it include financial incentives and support schemes? (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Existence of national policy on Data and Services, and if YES, does it include Open Access to data including financial incentives and support schemes? (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Existence of national policy on open learning including financial incentives and support schemes (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Existence of national, regional or sectorial research evaluation schemes of universities and RPOs, which account for existing institutional implementation of Open Science principles and Open Access schemes (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Existence of reference to EOSC in current national, regional or sectorial policies (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Progression of the institutional structure(s) at national level accountable for defining and implementing EOSC-related policies and strategies, including their hierarchical structure
- Existence of EOSC-dedicated funding streams or criteria in national funding mechanisms or programmes (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Existence of 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 (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Evolution (investment) of the backbone stakeholder(s) at national level, which is contributing to EOSC services (**KPI – number of relevant stakeholders in the individual categories, i.e. data infrastructures, e-infrastructures, high-performance computing (HPC) infrastructure, data repositories, and other services**)
- Sustainable funding of operational cost for the backbone stakeholder(s) at national level that is contributing to EOSC services (**Binary/ descriptive KPI: YES/NO/In planning + text**)
- Evolution of the Open Science mindset at the national or regional level (**KPI – number of universities, public and private RPOs, thematic infrastructures that have adopted Open Access policies, number of Open Access repositories, etc.**)
- Impact indicators (**KPI – e.g. share of Open Access users and publications per country, including users associated with industry, number of publicly available datasets, usage of data contributed by global providers**)

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 Report1](#)] in September 2019 on which community feedback was gathered, leading to a 'tinman' report [[WG Sustain Report2](#)], which was completed in December 2019. Analysis of the feedback received on the tinman report prompted the commissioning of a series of targeted studies, starting with the EOSC-Core operational costs.

This study involves the identification of the opportunities presented by and nature of the EOSC ecosystem, use cases and revenue models. Scenarios are being developed in collaboration with stakeholders, related projects and experts to understand cost structures. The first deliverable included a preliminary ecosystem model for EOSC, while the intermediate deliverable expanded the model, building on the initial interviews with service providers and users. This work has highlighted some difficulties in identifying the costs associated with EOSC services because the accounting 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 Iron Lady report, published in October 2020 [[WG Sustain Report3](#)].

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.

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 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 can 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.

6.3.2. Gaps

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, there has to be a national/EU funding model 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 are needed.

6.3.3. Priorities

- 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.

- 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 with an excellent knowledge of standards and best practices for delivering, using, sharing and analysing open and FAIR data, and applications and tools (services). EOSC will contribute to shifting the culture of research towards openness and transparency, to building bridges between different disciplines and organisational models, and to approaching data literacy in various modes and settings, while working on existing initiatives and preconditions.

This vision of a strong research data ecosystem that exploits digital technologies and has data and software at its core necessitates a comprehensive skills and education strategy for EOSC. Skills and training around EOSC are indeed essential for mainstreaming open science practices in research and thus essential in 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.

6.4.1.1. What types of skills?

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. Workforce capacity development is relevant to individual and institutional actors in the whole EOSC ecosystem. Institutional capacity in a broad sense, e.g. of organisations, units or teams, is crucial for a systematic and sustainable development of EOSC. Individual-level skills and competences form the basis, but ‘the group as a whole is more than the sum of its parts’.¹² 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. Figure 6.1 illustrates a detailed grouping of EOSC actors covering (i) the data-intensive science from the software/infrastructure perspective, (ii) the data sharing and reuse processes, (iii) the discipline domain exploration and analytical view, and (iv) the public’s contributions into the EOSC value chain.

Changing environment

Skills and roles will change and adapt depending on the needs but an analysis of the relevant skills and subsequent clustering, under an EOSC ecosystem, might lead to defining new job profiles (or strengthening recent ones) that could become essential for EOSC.

¹² Angus Whyte, Jerry de Vries, Rahul Thorat et al., D7.3: Skills and Capability Framework, 2018, p. 13 [[EOSCpilot D7.3](#)].

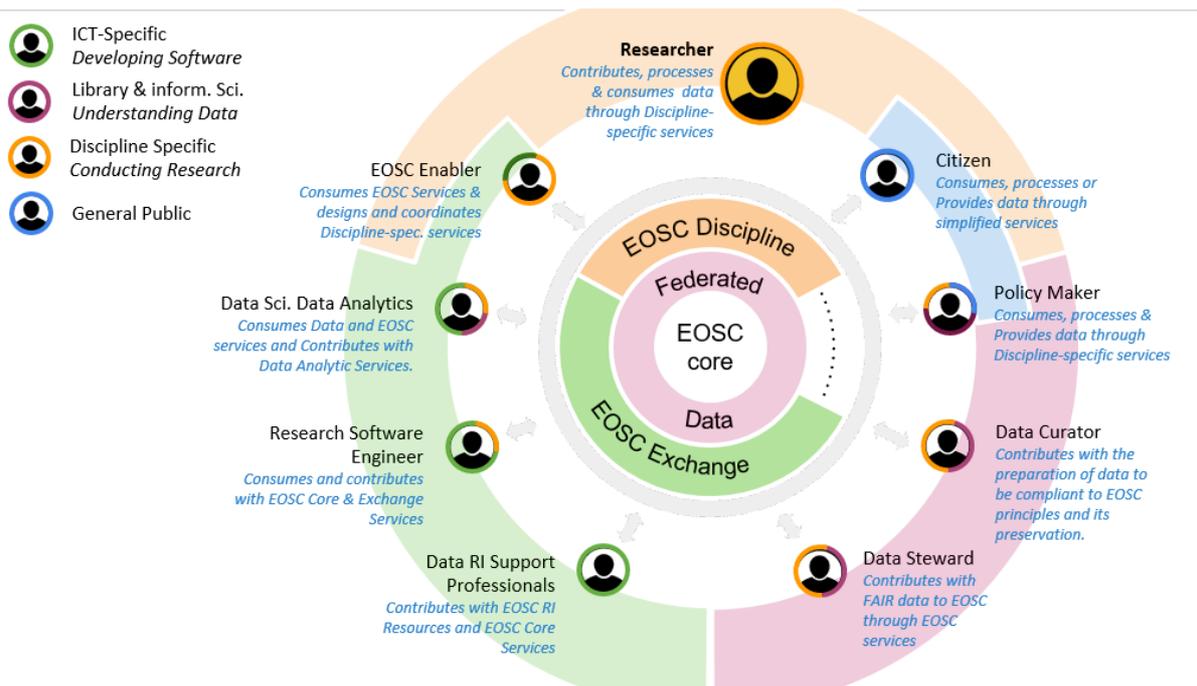


Figure 6.1: Actors in the EOSC ecosystem and related skills needs.

6.4.2. Gaps

Lack of data core expertise

Using or developing tools for handling data is becoming an increasingly important part of research.¹³ 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. University curricula across domains (scientific, engineering, medical, socio-economical, humanities, ethical and legal) do not yet commonly include open science skills or data-related skills (e.g. skills related to stewardship, data analysis, etc.)¹⁴. This is particularly true at Bachelor and Master level. And although several universities have curricula and PhD courses dedicated to the data scientist, there is a lack of widespread training on multi-disciplinary usage of digital research objects and on the interoperability of textual and numerical datasets, software and workflows, that are at the basis of EOSC. A concerted effort in skills and capacity development is needed to build and exploit the full potential of EOSC.

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. Over the past few years several European initiatives have begun to fill the gap of data professional profiles and to identify the competences that will be required, amongst them the H2020 projects EDISON [EDISON] and EOSCpilot.¹⁵ But there is still far from a common agreement

¹³ OECD report: The Digitalisation of Science, Technology and Innovation [OECD_DSTI].

¹⁴ Stoy, Lennart, Saenen, Bregt, Davidson, Joy, Engelhardt, Claudia, & Gaillard, Vinciane. (2020). D7.1 FAIR in European Higher Education (Version v1.0_draft) [FAIRsFAIR D7.1].

¹⁵ Whyte, A., Leenarts, E., de Vries, J. et al., D7.5 Strategy for Sustainable Development of Skills and Capabilities, EOSCpilot, 2019 [EOSCpilot D7.5].

on how these correspond to different career paths and related certification accreditation mechanisms, and are therefore embedded in research settings. It is necessary to provide recognition for these roles and define career paths which make them a viable choice.

Disparities

Although the reliance on the emerging new scholarly data and software support profiles are cornerstone elements in the implementation of FAIR data mandates (see e.g. OECD report: Building digital workforce capacity and skills for data-intensive science [[OECD BDWC](#)], RFI report on Digital Competencies [[RFII DG](#)]), a very diverse and uneven picture is seen across Europe. It is expected that geographical, disciplinary, gender, age (career stage) inequalities in building capacities for support staff and researchers will remain even in the middle or long term without a coordinated and coherent approach.

Lack of open science expertise

There is 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. This means employing skilled researchers and software engineers to develop the hardware (e.g. smart sensors) and the software to analyse and make generated data automatically FAIR.

Lack of legal/IPR and data ethics expertise

This is a challenge even among the FAIR data/OS experts. At present, this is typically addressed using legal consultants but as this aspect is very important in some domains (e.g. medical sciences, social sciences, arts and humanities) this approach may not be enough. In many cases, effective IPR and ethics support requires the combination of disciplinary and legal knowledge.

Lack of interdisciplinarity

Lack of interdisciplinarity, coordinated and coherent approaches to skills and competences building and of education and training provision is another area of concern. Differences across disciplines cannot be denied, but there is a need for a baseline approach for data stewardship, balancing the risk of infinite atomisation as different subdomains are traversed and the risk that too general an approach could result in the scientists' distrust of using their expertise.

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 with an overarching objective as a federated entity to coordinate the training offered at institutional, regional and national level, to help standardise educational curricula and to professionalise roles. Four key priority areas have been identified, with a phasing plan in line with the developments 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. Developing an EOSC leadership programme to foster the right policy environment for data skills and training.

All EOSC activities need to be shaped in collaboration with other key European initiatives (e.g., Digital Europe, EuroHPC, Artificial Intelligence) and take into consideration the new ERA priorities on circulation of knowledge and the importance of skills and training in ‘deepening the ERA’ [[EC-COM New ERA](#)].

Priority 1: Developing the next generation of open science and data professionals

Developing professional training programmes is a key priority to educate and re-skill/up-skill the current science workforce (researchers and research support personnel) with the specialised skill sets needed at various levels of an organisation. These programmes will stimulate life-long learning in the domain of data, software and infrastructure and will bring the anticipated culture shift in the organisation. This set of professional experts is expected to improve quality of data-driven research over time. Moreover, recognition of these experts’ careers will contribute to job security for the individual and close the loop of mobility across sectors, a strategic aspect of skills for sustaining jobs, growth and competitiveness.

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 recognition (evaluation and reward mechanisms), so as to make careers in academia attractive enough to compete with data science/engineering jobs in the private and public sector.
- **Develop data skills profiles** through community understanding and consensus mechanisms. This will help design and implement tailored profiles for building and enhancing training programmes, and solicit support from institutions and infrastructures.
- **Recognise data skills.** Promoting the ‘transparency and recognition of skills and qualifications’ is particularly relevant to the European and international nature of EOSC (cross-border, cross-discipline, cross-sector).
- Provide a **quality assurance framework** and certification mechanisms for trainers and trainees leading to career progression (e.g., recognised EOSC badges and seals¹⁶) to ensure continuous improvement, accountability and sustainability. A framework which:
 - Quickly adapts to changes in skill requirements;
 - Applies standards for qualifications and related experience;
 - Requires renewal methods after a set period of time;
 - Measures skills on a highly granular and individual basis (e.g., different sized-institutions);
 - Is recognised across borders through an EOSC credential ecosystem for transparency.
- Facilitate and simplify **lifelong learning mechanisms** for up-skilling in order to be reactive to the demands of the science labour market and to address the data skills

¹⁶ Example: Open Recognition Belgium [[OpenRec BE](#)].

shortages. Explore flexible funding models and offer training in a variety of forms to address diversity and support for flexible learning.

- **Align Data Competence Centres.**¹⁷ Data Competence Centres are proposed as a focal point for EOSC adoption, and their joining forces is crucial to better utilise resources and anticipate demands for new skills or new personnel at the European level. EOSC should facilitate such alignment, including the **establishment of networks of experts**, both at **cross-disciplinary** and **cross-national** levels. Coordination should focus on the co-development of common practices and tools, stimulating and promoting the mobility of students, data professionals and domain experts, while creating mobility opportunities beyond EOSC such as industrial internships (including AI and HPC). Special attention to Social Sciences and Humanities (SSH) is of key importance as there has been less emphasis on data skills in these disciplines. **Cross-national** alignment will support collaboration between different actors in MS/AC by i) fostering and empowering national coalitions, and ii) establishing a European Network of Data Competence Centres.

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 **close the data literacy gap**: 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 data¹⁸ [[Qlik Data Literacy](#)].

Data Science has already entered the higher education domain in MSc and in a cross-disciplinary fashion. Yet, Higher Education Institutions (HEIs) should ensure a cultural change by embedding practical skills acquisition focused on open science, FAIR data and software skills into ongoing educational programmes on a regular and standardised basis from as early as possible, and to ensure engagement with diverse cohorts. Students and researchers are more likely to make an effort to gain such skills if these are relevant for their academic or other non-academic careers.

Activities to be considered within the realm of EOSC include:

- **Align curricula and training with demand.** Generate a consensus on a core European higher education curriculum to deliver data skills at university level, specifically addressing open science, data stewardship and interdisciplinarity skills.
- 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.,

¹⁷ Data Competence Centres are hubs for uniting technical and non-technical activities, for instance by bringing technology and application development together and fostering skills, competence, and best practices. They offer advanced technical expertise, access to the latest knowledge and information on open/FAIR data practices and tools, services integration and use, including the availability of labs, pilot and experimental facilities, and other technological and scientific infrastructure.

¹⁸ The Human Impact of Data Literacy, 2020, Qlik & Accenture, https://thedataliteracyproject.org/files/downloads/Qlik_Accenture_Human_Impact_of_Data_Literacy.pdf

supervisors and mid-career researchers, teachers in HEI) 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 ECTS¹⁹ or other formal certificates.
- Build and operate a **network of researchers-champions in open science**, stimulate and promote their mobility as a means of cross-pollination for EOSC (e.g., Erasmus+, MSCA).
- Promote and support **advanced learning environments** as part of the broader open science agenda and the engagement of researchers in participatory activities (i.e. project design for Citizen Observatories²⁰ linked to EOSC) to develop different skills, not only in data acquisition but also in interdisciplinarity, data analysis, etc.

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

Learning materials (content), learning styles and methodology and learning platforms (tools) play a pivotal role in increasing researcher/professional engagement and in improving skills. Researchers, data scientists, data analysts, data engineers, research software engineers, data curators, managers and data stewards will be able to discover learning materials tailored to their needs via user-specific entry points. Learning materials descriptions will be aligned and available to humans and machines, ensuring their easy discoverability, according to defined standards, open access and reusability in terms of formats and licences.

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. Learning materials, especially those covering data/computational skills, are rapidly becoming obsolete and regular updating is quite resource intensive. The lifecycle of the materials, and the continuous need to update data/ICT professionals, should thus be taken into account, to ensure that training is up to date with technology and policy changes, as part of lifelong learning programmes.
- Devise a **common framework for learning pathways** for different open science and data-related profiles to enable learners to navigate through the content, selecting what is useful to them. Materials will be delivered with data and analytics in mind, providing all learners with quantified measures and analytics that reflect their aptitude, skills and strengths (linked to learning analytics).
- Support the development of an **EOSC Knowledge/Education Hub** as a set of interconnected and decentralised learning platforms and living repositories for knowledge sharing. Support a *federated training catalogue* where people could have visibility of the different training and learning resources scattered about in a fragmented ecosystem. Education Hubs should be developed in close cooperation with universities and higher education centres, and would form an integral part of Data Competence Centres.

¹⁹ European Credit Transfer and Accumulation System, usually used for students (sometimes PhD level, but not everywhere) [ECTS].

²⁰ Mominó, J. M., Piera, J., & Jurado, E. (2017). Citizen Observatories as Advanced Learning Environments. In *Analyzing the Role of Citizen Science in Modern Research* (pp. 192-212). IGI Global.

- Facilitate the **adoption of open learning environments** as key EOSC services and tools, and operate an *EOSC online learning and communication platform* to host co-developed material from ‘orphan’ creators (similar to Zenodo for repositories), a shared and interoperable platform, based on open source software.
- Promote and support **innovative ways of learning** by employing creative methodologies and technology for teaching/training at all levels, from awareness to focused expertise.²¹ Address the whole research lifecycle where all EOSC audiences are made aware of the techniques and stages through the use of easily understandable narratives. Pilot gaming techniques (creativity) and cross-disciplinary learning exercises, transforming the current knowledge-driven approach into an experience driven one. Support the combination of both traditional (classroom) education and online tools.

Priority 4: Developing an EOSC leadership programme to foster the right policy environment for skills and training

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 data skills and training. An EOSC leadership programme would ensure EOSC functions as a catalyst to create this broader enabling environment.

There are currently significant asymmetries in different national approaches to open science and data skills for the broader community. Consequently, national skills and training initiatives will face different challenges in embedding national programmes. Leadership programmes can play a key role in creating a *community of influencers* with the ability to affect change within their national environments, to ensure that policy decisions deliver success in the data revolution.

However, the EOSC leadership programme will be unique in its focus on influencing national policies that support the aims of EOSC. 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 (AI, HPC) and, for universities, to quickly adapt curricula and training that change from responding to industry needs, to driving them.

The EOSC leadership programme will draw on existing leadership programmes that inclusively develop different types of leaders in open science and/or data skills²², and university associations’ programmes on open science leadership: YERUN, LERU, IARU, LIBER, etc. Its aims will include:

- Connecting with actors across different EOSC projects, communities, backgrounds, and identities;
- Empowering each other to become effective EOSC and open science ambassadors in the communities and institutions.

²¹ Example Google game for data science/Machine Learning [[What If](#)].

²² Examples: Open Life Science [[Open Life Sci](#)], eLife Innovation Leaders [[eLife](#)], Mozilla Open Leaders [[Mozilla OL](#)].

6.5. Rewards and recognition

[Work in progress]

Present-day rewards and recognition (R&R) systems are shaped by government-mandated national and institutional policies and regulations, but they are also stimulated by the competitive environment in which academics and institutions compete for funding and other resources. Many R&R systems currently used by research-performing and research-funding organisations tend to incentivise and reward a narrow range of academic activities – e.g. publishing in journals and attracting external research funding – and rely on a limited and often problematic set of evaluation tools (e.g. simplistic publication metrics such as the journal impact factor and the H-index). This leads to unequal appreciation of the various fields of science and hinders knowledge utilisation and the uptake of Open Science practices [Cohen 2019; Saenen 2019]. To address this, high-level principles to guide research evaluation were presented in 2015 in the Leiden Manifesto [Leiden Manifesto], and earlier this year the Declaration on Research Assessment (DORA) initiative summarised five design principles [DORA 1] to help institutions experiment with and develop better research assessment practices. The DORA initiative also maintains a curated list of good practice examples of implementation from institutions showing leadership in this area [DORA 2].

Generalising, a culture change needs to be realised in order to increase the quality of education, research, impact and leadership. More than a technical issue (e.g. ‘better indicators’), a responsible R&R system is also a social issue: a catalyst to foster good research practice and quality in terms of content, openness, scientific integrity and contribution to society. Future evaluation of scientists should have a better balance in valuing achievements in education (if appropriate); research; influence (on science and/or society and/or economy and/or teaching); organisation and leadership. In particular, evaluation and promotion criteria should recognise openness and FAIR practices. A diversity of career paths should be made possible in order to reward good performance in different areas. Evaluation at different levels (institution, department, individual) and for different career stages (from early-stage researchers to full professors) should be based on appropriate criteria. All aspects of Open Science should be stimulated more – for example, the sharing of scientific results with society – and the accessibility of research results should be promoted and rewarded.

6.5.1. Priorities:

- Create a Europe-wide framework for rewards and recognition that includes Open Science.
- Produce a country-level inclusive approach to research evaluation, taking Next Generation metrics into account.
- Discuss this approach within and between (all) the institutions in the country.
- Create interaction between the countries on this topic and learn from each other.

EOSC as one of the organisations stimulating Open Science should help in providing guidelines for adapting Rewards and Recognition systems aligned with the priorities outlined above.

6.6. Communication

[Work in progress]

EOSC addresses researchers, but also policy advisors and research funders as it comprises a system change. Adapting to Open Science will change the way research is conducted, funded and evaluated and how its outputs are valued.

This diversity of stakeholders requires a communication policy that addresses the different needs of these groups. The EOSC Landscape Working Group distinguished nine different groups [[EOSC Landscape](#)]. For communication purposes, these groups can be aggregated into three main stakeholder groups:

- Research Service Providers
 - e-infrastructures, such as PRACE, GEANT, 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.

Each stakeholder group may have different expectations and perceptions of EOSC. Even if these interpretations differ, they can still be consistent. For example, funders may focus on governance and efficiency issues, whereas researchers and providers focus on functionalities.

The Executive Board has 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 consists of six members, coming from the Executive Board, Governance Board and communication experts from the EOSC Secretariat and European Commission. Depending on the topic, more experts can be called in. The Task Force has regular virtual meetings. Its main activities are in the EOSC Work Plan 2020.

The Communication Task Force focuses on stakeholder engagement, content production, branding and positioning of EOSC for the different stakeholders. Currently, there is a lot of preparatory work on EOSC, with many (EC) projects. Therefore, the Task Force focused on designing a new template for (PowerPoint) presentations, using the current logo, to be used by the EOSC Executive Board, Governance Board, Working Groups and EC. It also set a protocol for the use of the EOSC logo by (EC) projects and external entities. This policy was approved by the Executive Board, and published on the website of the EOSC Secretariat. The Task Force also prepared key content for presentations to the different stakeholders.

For the second half of 2020, the focus will be on further content production and elaborating key messages for the different stakeholder groups. One of the actions will be to explain EOSC via Q&As, use cases, best practices, etc. Now that the EOSC legal entity has been established, the Task Force will bring up ownership, licensing and liability issues, both within the legal entity and in its relationships with EC projects, service providers and other contributors to

EOSC. Next, a communication strategy will be prepared, including branding, positioning and promoting tools and services once these become available.

6.6.1. Priorities

Current priorities are on providing the key messages for the different stakeholders and to engage with these communities. Now that the legal entity has been established, the Task Force is also working on licensing and ownership issues, and has started the preparatory work for the communication strategy.

[Priorities from Open Consultation document:

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

6.7. *Widening to public and private sectors and going global*

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.

Study methods

The study identified use cases resulting from cross-domain, data-driven applications created in pan-European collaborations by research communities, citizen scientists, public sector 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.

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,²³ 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

²³ 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.

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 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.

Results

Key findings of the study include the following:

- 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.
- 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 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.

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 FAIR-ification 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.

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 INFRAEOSC-03:

- 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 automation and customisation layers on top of EOSC e-infrastructure (e.g.: AirBnB for compute services).
- Integration of intellectual property tracking.

6.7.1.2. *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.

[Priorities from Open Consultation document:

- 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 established. 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.

Global cooperation through Open Science has the potential to revisit long-standing scientific questions as well as address new ones. This is particularly true for a number of pressing contemporary challenges:

- 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 (e.g. oceanography)).

6.7.2.2. Priorities

This section outlines the priorities for the international dimension of the strategic objectives defined in the EOSC Objectives Tree (Figure 1.7 on page 46), taking into account the need to adapt to and consider diverse capabilities and demands, and the principles, outlined in Section 2.6.

❖ Strategic Objective 1. *Open science practices and skills are rewarded and taught, becoming the ‘new normal’*

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

Priorities

1. 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.
2. Engage with the rapidly evolving global policy landscape of Open Science and support the creation of a policy observatory.
3. Promote European best practices and values abroad, and also identify and integrate best practices and values from third countries wherever relevant.
4. Establish state-of-the-art trainings for third-country users on Open Science, based on the curriculum on Open Science skills in the Skills and Training Working Group (WG) currently being developed, and taking into account local capabilities, whenever possible.
5. Initiate an international data steward network across domains to exchange best practices and success stories.

Deliverables

| Deliverables | 2021–22 period | 2023–24 | 2025–2027 | Responsible |
|--------------|---------------------------------|----------------------------|-----------|-----------------------|
| 1.D.1. | Disseminate Open Science Policy | Seek global integration of | - | EC, EOSC Association, |

| Deliverables | 2021–22 period | 2023–24 | 2025–2027 | Responsible |
|--------------|--|--|-----------|---|
| | Platform (OSPP) 2 Final Report. Discuss with global initiatives where EOSC could contribute to their work. | Open Science-related programmes of work where appropriate | | Members States (MS) and Associated Countries (AC) |
| 1.D.2. | Spread European Open Science (OS) policy experience to the Global Policy Observatory within RDA. | Follow closely the work of the Policy Observatory, to further develop EOSC and the Global Open Data Commons cooperation. | - | |
| 1.D.3. | Exchange best practices between EOSC and third-country solutions. | - | - | EC, EOSC Association, MS and AC |
| 1.D.4. | Disseminate the curriculum of the WG Skills and Training abroad. | Report on the uptake of the cursus. | - | EOSC Association |
| 1.D.5. | Facilitate the cross-domain knowledge exchange between curators/stewards. | - | - | |

❖ Strategic Objective 2. *Standards, tools and services allow researchers to find, access, reuse and combine results*

§ Main milestone. *EOSC provides a trusted platform supporting the development of innovative services and products.*

Priorities

1. Support the work being undertaken on methods, to complete FAIR – in particular interoperability – at a global level (e.g. authentication and authorisation infrastructure (AAI), persistent identifier (PID), FDO, FDMM, etc.).
2. Support the setting up of ‘Wise Persons global fora for Architecture’, to identify and remove the technical barriers that hinder the full potential of Open Science.

3. 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.
4. 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.
5. Provide state-of-the-art trainings developed by the Skills and Training WG to third-country participants, 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.

Deliverables

| Deliverables | 2021–22 period | 2023–24 | 2025–2027 | Responsible |
|--------------|---|---|---|-------------|
| 2.D.1. | Follow up the development of methods at an international scale. | - | - | |
| 2.D.2. | Set up a focused Wise Persons forum made up of technical experts from organisations with global outreach (e.g. Data Together and GÉANT) | Continuous monitoring, identification of technical barriers, proposing solutions, in particular regarding industry take-up. | Implement state-of-the-art solutions for an EOSC 3.0. | |
| 2.D.3. | Create an awareness-raising and dissemination roadmap towards third countries. | Monitor the take-up of EOSC building blocks abroad and the hindering factors. Analyse interoperability with local technology choices. | - | |
| 2.D.4. | - | Report on take-up of EOSC service portfolio | Report on take-up of EOSC service portfolio | |

| Deliverables | 2021–22 period | 2023–24 | 2025–2027 | Responsible |
|--------------|--|---|--|-------------|
| | | in third countries. | in third countries. | |
| 2.D.5. | Develop timely trainings, tailor-made for local capabilities to achieve a level playing field. | Collect stakeholder feedback on the usefulness of trainings, for proposal excellence. | Update training materials based on stakeholder feedback. | |

❖ Strategic Objective 3. *Sustainable and federated infrastructures enable open sharing of scientific results*

§ Main milestone. *The EOSC infrastructure is in operation, providing a Web of FAIR Data and Related Services underpinning research addressing major societal challenges*

Priorities

1. Initiate EOSC Rules of Participation (RoP) for service providers from third countries, noting that compliance with applicable legislation is a prerequisite beyond the RoP.
2. Develop value propositions to third country service providers, to widen the EOSC portfolio.
3. Support third countries with know-how on the setting up of FAIR infrastructures (e.g. repositories, research and e-infrastructures), as well as skills and best practices to achieve operational excellence.
4. Set up contingency measures in case of service disruption with third countries (technical and diplomacy in case of digital blackout).
5. Encourage emerging regional Open Data Commons in countries/regions with commitment to Open Science, eligible for the EU Development Funds.
6. Make the Horizon Europe international project outputs and metadata mandatorily accessible, through close collaboration between EOSC and CORDIS.

Deliverables

| Deliverables | 2021–22 period | 2023–24 | 2025–2027 | Responsible |
|--------------|---|---------|---|--|
| 3.D.1. | Prepare RoP for third countries. | - | Review RoP for third countries. | Working Groups/ Executive Board (EB)/ Director |
| 3.D.2. | Develop value propositions for third-country service providers. | - | Report on the international service exchange of EOSC. | EOSC EB, Director |

| Deliverables | 2021–22 period | 2023–24 | 2025–2027 | Responsible |
|--------------|--|---|--|---|
| 3.D.3. | Develop training, broken down per FAIR infrastructures (technical and human). | Continuous upgrade of the curriculum. | Continuous upgrade of the curriculum. | |
| 3.D.4. | Review the resilience of technical infrastructure and the risk of potential disruption in science diplomacy. | Develop and implement a contingency measure for digital blackout (technical and diplomatic) and a risk management plan, with a clear allocation of responsibilities (technical and diplomatic). | Maintain and upgrade. Build into safety culture of EOSC. | Steering Board, Technical Working Group |
| 3.D.5. | Seek synergies with the IPA III, NDICI, the international chapter of Horizon Europe e.g. TAIX. | - | - | |
| 3.D.6. | Develop cooperation plan with CORDIS, on how to make the (meta)data available. | - | Report on the availability of (meta)data from Horizon Europe projects and the FAIR implementation. | European Commission |

❖ Crosscutting Objective. *Boost the impact of EOSC through collaborations and alliances*

Priorities

- Direct Member State participation in the global research ecosystem, while maximising the added value of their bilateral international connections.
- 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 (GODCC) 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.
- Develop tailor-made value propositions to the global regions and explain how they can engage with EOSC.
- Give support to and cooperate with existing international projects and initiatives, build on their work and contribute to their mission (e.g. Data Together (CODATA, GO FAIR, RDA, WDS)).
- Contribute to evidence-based policy making, by EOSC being an effective and reliable partner to global, regional and local policy makers.
- Enable the formation of international consortia for Horizon Europe calls.
- Enhance the EOSC Stakeholder Forum with third-country participants and promote the opportunities for co-creation and collaboration, enabled through events held by members (e.g. TNC, RDA Plenary, International Data Week, etc.).
- Sustain and strengthen the underlying infrastructure and skilled workforce, enabling seamless data flows through certified local and regional repositories which constitute cultural heritage and digital sovereignty.
- Promote virtual mobility of researchers, thereby reducing carbon footprint.
- 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.

Deliverables

| Deliverables | 2021–22 period | 2023–24 | 2025–2027 | Responsible |
|--------------|---|---|---|---------------------------------------|
| C.D.1. | Build on mapping project such as RISCAPE, prioritise bilateral relations. | EOSC dissemination plan through bilateral channels. | - | MS/AC, EOSC |
| C.D.2. | Develop model MoUs and start negotiations. | Signature of the first MoUs. | Review the impact of the signed MoUs. | EOSC EB and Director |
| C.D.3. | Follow and direct the work of the RDA Interest Group; in parallel, start drafting the Charter, which should be in line with the MoUs. | Promote the Charter, complementing the MoUs. | Review the implementation of the Charter. | Members of EOSC, EOSC EB and Director |

| Deliverables | 2021–22 period | 2023–24 | 2025–2027 | Responsible |
|--------------|---|--|---|-----------------------------|
| C.D.4. | Develop tailor-made value propositions and potential membership structure for different regions. | 1 st set of third countries joining. | 2 nd set of third countries joining. | EOSC EB and Steering Board |
| C.D.5. | Align activities of the international members to the Co-programmed EOSC EPA. | - | - | EOSC GA and Executive Board |
| C.D.6. | - | Review the added value of EOSC to European evidence-based policy making. | Improve EOSC's role in evidence-based policy making based on the results of the review. | |
| C.D.7. | Develop stakeholder feedback survey on whether EOSC was able to direct future consortiums under Horizon Europe. | - | - | Stakeholder Forum/ GA |
| C.D.8. | - | Number of third-country organisations in the Stakeholder Forum. | Number of third-country organisations in the Stakeholder Forum. | Stakeholder Forum |
| C.D.9. | Develop survey to measure virtual mobility satisfaction. | Report on virtual mobility. | Follow-up report on the satisfaction with virtual mobility solutions via EOSC. | EOSC EB/ Working Groups |

| Deliverables | 2021–22 period | 2023–24 | 2025–2027 | Responsible |
|--------------|--|--|--------------------------------|-------------|
| C.D.10. | Develop EOOSC SDG contribution, broken down to targets. | Choose the appropriate metrics to measure success. | Report on SDG contribution. | EOOSC EB |

7 Expected impacts

[Work in progress]

7.1. Introduction

Sustainable collaborative research frameworks

The climate crisis, the extinction of species, global poverty and social inequality are only a few of the challenges that humankind has to face in the 21st century [EC HorizonEurope Missions]. Research plays a crucial role in facing 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 fill infrastructure gaps in unstructured areas, 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 1, the final row of the tree describes the benefits of EOSC in the three dimensions of Science, Industry and Society:

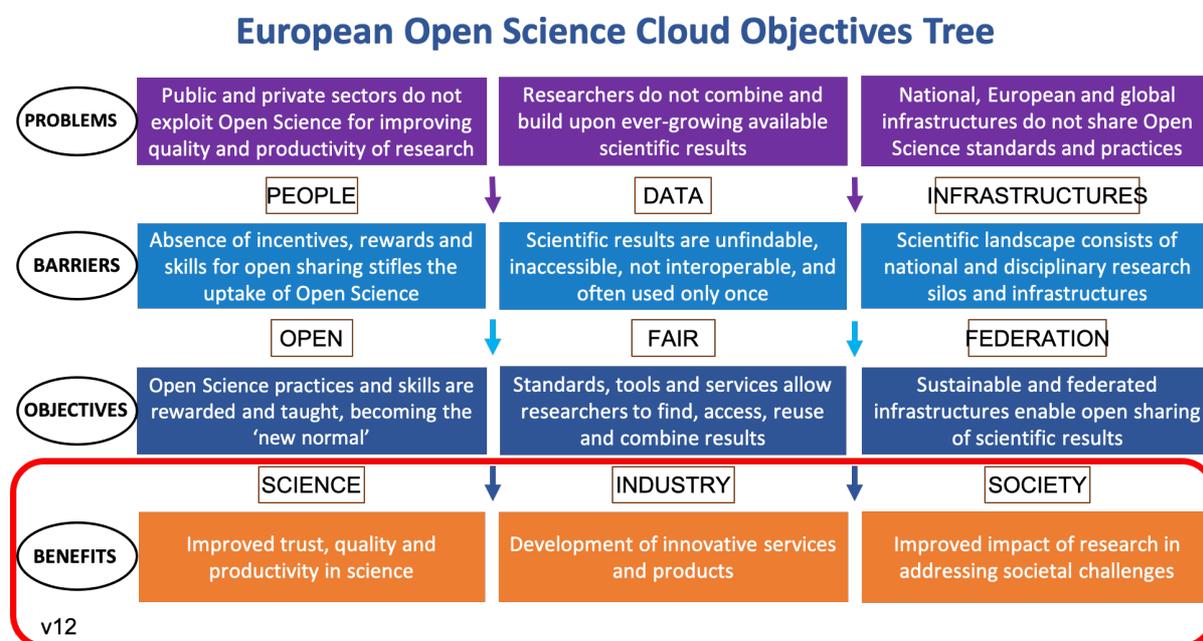


Figure 7.1: EOSC Objectives Tree – benefits

This section considers the impact of EOSC on each of these areas. Section 7.2 addresses the impact of EOSC in improving trust, quality and productivity in science; Section 7.3 looks at the development of innovative services and products; and Section 7.4 discusses the role of research in addressing societal challenges.

7.2. Improved trust, quality and productivity in science

Encouraging collaboration and openness

EOSC will result in cultural changes in the entire research ecosystem. Open Science, which is realised in 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, models, simulations and

methodologies. Making these outputs findable, accessible, interoperable and reusable (FAIR) is therefore a key element of perspective in measuring and rewarding the contribution of research.

Open science and EOSC will have a significant structural effect with the potential to change not only the way research is performed, by creating a pan-European, multi-disciplinary research infrastructure supporting a broad range of a researcher's data and computing needs, but also enabling 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

Already, the basic fact that EOSC represents a secure, safe and transparently trusted virtual environment where scientific data 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 data and, by developing certified services and standards, will enhance the quality of data 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 provide 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 removing 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, so 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 preventing some limitations of human 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 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; the majority of data is stored in local institutional or disciplinary repositories, resulting in disconnected research data silos where data are largely unfindable, 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 apps released to gather this 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.

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.

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 methodologies, instruments, simulations, and analysis and workflow support software. 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 shared, stable and structured 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.

7.3. 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 data, 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 European Strategic Partnership, EOSC 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 impact of EOSC on the capacity of research to address the current and future global challenges will be based on the following key features introduced by EOSC:

- More cross-sectoral and multi-disciplinary research through data sharing and cross-usage of services. For each discipline, the communication and collaboration with scientists in the same and other disciplines will benefit from faster and seamless sharing of publications, data, software, services, tools and other digital research outputs.
- Better and faster sharing of data and results will strengthen collaboration among researchers and disciplines and present opportunities for new levels of integration.

For the cooperation between teams to address multi-disciplinary challenges, the use of data and software across research silos will allow the exploration of new avenues to an extent that has never been possible before.

- Increased added value of the services in the EOSC ecosystem. A European-scale environment for computational, storage, analysis and other data-related services and tools will facilitate multi-disciplinary cooperation, leading to discoveries and solutions in key areas such as environment and health.

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 it 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’.

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 the EOSC will leverage the existing EC investments for the benefit of multiple science and innovation communities. In line with the intervention logic of the Communication, 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.

EOSC 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)²⁴ can be

²⁴ Layers of the interoperability model defined in the European Interoperability Framework [[EC Interoperability](#)].

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.

7.4. 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 ensure that open science becomes 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 greatly 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 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 code 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 has become, by necessity, 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 in 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 EOSC.

8 Risk management

8.1. Introduction

Complex organisations need to consider risk management to prepare for, and consequently to overcome, potential crises. Such Enterprise Risk Management (ERM) includes the activation of an organised risk management structure, as well as a pool of actions and processes, rules and responsibilities through which decisions are taken and implemented in terms of risks. ERM is a system of competences, organisational roles, policies, processes and models of analysis that allows an organisation's management to improve governance and control over its development path.

ERM is aimed at increasing the value of an organisation for the benefit of its stakeholders, supporting its objectives through the preparation of a methodological framework that allows a coherent and controlled performance of each future activity, the improvement of the decision-making process, planning, and creating priorities through a comprehensive and structured understanding of the activity itself. Risk management also contributes to a more effective use and allocation of capital and resources within the organisation, to the protection of the assets, to the corporate image, to the know-how of the organisation and the key people, as well as to the optimisation of operating efficiency.

A targeted study has been conducted by AON Hewitt [[AON](#)] on behalf of the EOSC Sustainability Working Group in order to introduce clear and structured guidance on how to incorporate risk management into the governance of the EOSC Association. To this end, AON studied, analysed and assessed the governance structure of EOSC – both the interim governance in place up until December 2020 and the future EOSC governance (post 2020) – 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 vs. the assessment and management of risks.

8.2. Methods

The basis for this assessment was the analysis of over 20 official documents, reports and presentations produced by EOSC, as well as the output from the EOSCsecretariat.eu (EOSC Liaison Platform), and a benchmark analysis of over 6 organisations and associations, chosen because they are comparable to a complex EOSC structure. In addition to these analyses, AON conducted a set of interviews on a panel of 14 key people from EOSC, which took place from June 15th to July 17th 2020.

Additionally, AON conducted a SWOT analysis²⁵ of the environment that could influence or modify the expected development of EOSC, categorising its characteristics and elements into four parameters:

- Strengths: characteristics of the business or project that give it an advantage over others;
- Weaknesses: characteristics of the business that place the business or project at a disadvantage relative to others;

²⁵ SWOT analysis (or SWOT matrix) is a strategic planning technique used to identify strengths (S), weaknesses (W), opportunities (O), and threats (T) related to business objective and project planning

- Opportunities: elements in the environment that the business or project could exploit to its advantage;
- Threats: elements in the environment that could cause negative implications for the business or project.

The results of this SWOT analysis are outlined below in Figure 8.1.

| | Helpful to achieving the objective | Harmful to achieving the objective |
|---|---|--|
| Internal origin (attribute of the organisation) | Strengths <ul style="list-style-type: none"> • connect different scientific communities • exchange and make research data interoperable • federate data infrastructures • FAIR data management and tools • human capital very rich in multidisciplinary technical skills, sensitivity to governance issues, passion for the activities to be carried out and for the project | Weakness <ul style="list-style-type: none"> • absence of a structured and approved ERM framework: absence of defined risk management roles and responsibilities; absence of risk appetite definition • complexity of the organisational structure and the potential organisational biases • absence of a Scientific Committee within the Board • one size does not fit all due to different infrastructure maturity levels |
| External origin (attribute of the organisation) | Opportunities <ul style="list-style-type: none"> • multidisciplinary • multiculturalism & international environment • political involvement • exchange of best practices • emerging direct competitors | Threats <ul style="list-style-type: none"> • multidisciplinary • multiculturalism & international environment • political involvement • emerging direct competitors • changing regulatory environment • cyberattacks and technological default |

Figure 8.1: EOSC SWOT analysis by AON

A risk model was developed defining five specific risk categories for the EOSC Association:

- **Operational:** the prospect of loss resulting from inadequate or failed processes, procedures, systems, people, policies or technologies;
- **Reputational:** the potential for negative perception that the wide spectrum of stakeholders connected with EOSC may have. Reputation represents a successful indicator for the entire EOSC ecosystem.
- **Governance:** any of various types of risk resulting from an inadequate and ineffective governance
- **Financial:** any of various types of risk associated with financing, funding and economic sustainability in the long term;
- **Strategic:** associated with misalignment of the strategy to the mission and vision needed for EOSC, to the failure of including the strategy in the decision making and failure in deployment and/or execution of the strategy.

These risks have been prioritised by quantitatively evaluating the likelihood of occurrence and the potential impact (Figure 8.2), and graphically reported in a heat map (Figure 8.3). This prioritisation has been achieved by assigning a Risk Priority Number (RPN) to each risk. The definition of the RPN is based on a qualitative evaluation of probability of occurrence and potential impact of the risk. Specific evaluation indexes for impact and likelihood were defined, based on AON’s experience in developing Risk Analysis projects. The evaluation scale of the potential impact is intentionally a non-linear scale to avoid the distortion that a highly impacted risk with low probability has lower priority than a lower impact risk with high probability.

| Severity index | Impact |
|-----------------|--|
| 1 – Low | Insignificant impact Negligible deviation from project objectives |
| 2 – Medium low | Minor impact Significant deviation from project objectives |
| 3 – Medium High | Major impact Project objectives significantly compromised |
| 5 – High | Disruptive impact Complete failure of project objectives |

| Probability index | Likelihood |
|-------------------|-----------------------------------|
| 2 - Low | Remote likelihood of occurrence |
| 3 – Medium Low | Failure unlikely to occur |
| 4 – Medium High | Failure likely to occur |
| 5 - High | Failure extremely likely to occur |

Figure 8.2: Prioritisation according to impact and likelihood

Having determined likelihood of occurrence of each risk and expected impact, RPNs have been directly computed as the product of the aforementioned indexes. In order to obtain a graphical visualisation of the Risk Exposure, a Risk Matrix is presented. In each cell of this matrix the value of the RPN is reported. Colour coding eases the identification of the most relevant risks to be considered.

Looking at the evolution of the risk heat map (see Figure 8.3), it should be noted that the proposed measures significantly reduce the risk profile, but not conclusively. Only the complete and holistic structuring of EOSC governance and adequate EOSC risk governance will allow the synergies between the different actions to be captured and the represented risk profile to be reduced.

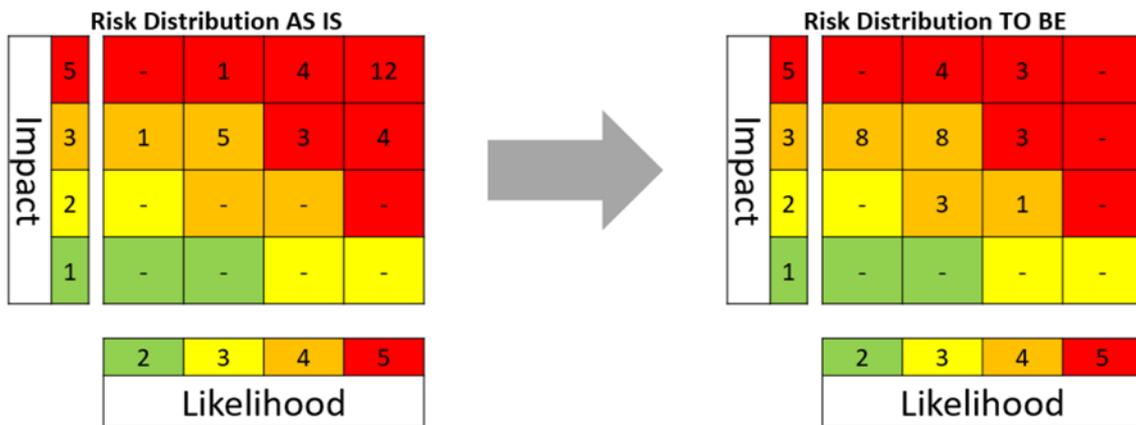


Figure 8.3: Heat map evolution of risk distribution

8.3. Results and recommendations for sustainable EOSC risk governance

The study found a human capital very rich in multi-disciplinary technical skills, sensitivity to governance issues, passion for the activities to be carried out and for the belief in EOSC itself. The presence of all, or almost all, the essential pillars for the construction of effective risk management was also found. The study considered the EOSC major players to have a very high and mutual interest in and awareness of the importance of the subject and that between them the stakeholders have the skills and experience required for effective risk governance.

These actions resulted in the identification of 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. The factors include the organisational model, political influences, multinational and cross-disciplinary usage. Risk management activities for EOSC have been limited to individual project-based analysis and therefore fragmented, i.e. a clear and defined risk governance structure, with assigned roles and responsibilities for risk management, has not yet been established.

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;
- 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 in case, process and procedure to intervene in case of system failure.

However, AON has also found a human capital very rich in multi-disciplinary technical skills, sensitivity to governance issues, passion for the activities to be carried out and for the belief in EOSC itself. Moreover, the presence of all, or almost all, the essential pillars for the construction of effective risk management has been found. The EOSC major players with whom AON conversed during the study showed an interest in and awareness of the importance of the subject that was very high and mutual amongst the relevant players, as well as showing that the relevant skills and experience in the different areas of risk management are spread among stakeholders, even if currently they are not systematically introduced in a complete and effective risk governance.

The study sets out 32 recommendations to address these gaps and ensure the effectiveness of an EOSC risk governance in order to increase the value of the EOSC Association and the benefit to its stakeholders, and to support its business objectives and allow a more effective use and allocation of capital and resources within the organisation.

The main recommendations are as follows:

- **Launch a comprehensive plan** to address the 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, effective, adequate and well-formalised, and appoint roles and responsibilities across the organisational structure (i.e. Risk Management Control Committee, Chief Risk Officer and his/her team with corresponding budget, Risk Owners, etc.);
- **Define the EOSC ERM policies** for regulating roles and responsibilities within the risk governance structure. The policies serve as the strategic guidance reference for risk management and regulate the interactions between the different stakeholders;
- **Design the risk assessment and reporting process** that properly analyses all the main risk areas, including strategic areas and alignment with EOSC's mission and vision, that are affected by internal and external environment (e.g. politics, economy, technological development, regulations, society), and enhances risk intelligence;
- **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 (e.g. strategy, economics and finance, ICT technologies, cyber security, international relations, regulatory, programme and project management, project risk management) and set requirements on the composition of risk management bodies to assure independence in decision making;
- **Establish a risk awareness programme** to support and enhance the culture and skills in risk management across EOSC;
- **Set up an infrastructure and data security team (or committee) 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. cyber attacks, business interruption, damage to data, failure of systems or applications, etc);
- **Improve technical resilience** of the EOSC infrastructures by:
 - Performing specific business impact analysis and identifying the most relevant business interruption risk causes;
 - Defining, introducing and updating the business continuity management plan;
 - Drafting and testing the disaster recovery plan;
 - Defining a set of binding rules about resilience, business continuity management plan and disaster recovery plan for service providers.

From these main recommendations, AON deduces the following actions as a blueprint for the next phase of the EOSC governance:

- Set-up a small working group that has the necessary commitment and represents a correct mix of skills and representation of stakeholders' interests;

- Identify an individual endowed with the necessary experience and analysis skills to guide the ERM deployment project, able to provide a holistic vision of the areas affected by risks, and manage the complexity by proposing different alternatives and transforming them, once identified, into operational processes and matrices of responsibility;
- Integrate the main recommendations of risk management and analysis of the interdependencies between areas in the SRIA evolution;
- Set up and lead a process to continuously analyse and closely monitor the risks that may jeopardise the execution of the identified strategy and the achievement of the strategic objectives and the most relevant measurable targets, also leveraging through the continuous monitoring of a specific and wide set of strategic key performance indicators (KPIs) and key risk indicators (KRIs) that should be designed and implemented;
- Set the priorities of the governance objectives to be targeted and design an incremental plan to achieve them.

The identified gaps and recommendations are to be considered by the EOSC Association, the EOSC contributing projects and the EOSC partnership overall to develop a comprehensive ERM.

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.

9 Conclusions

[not included in V0.8]

Appendix A Related documents

[not included in V0.8]

References

[Work In progress]

- [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>
- [AON] <https://www.aon.com/home/index.html>
- [ARCHIVER] <https://www.archiver-project.eu/>
- [Berckman_HCC] http://dsc.soic.indiana.edu/publications/paper_computing_continuum-2.pdf
or
https://www.researchgate.net/publication/332246123_Harnessing_the_Computing_Continuum_for_Programming_Our_World
- [Budapest_OAI] <https://www.budapestopenaccessinitiative.org/read>
- [CARE] <https://www.gida-global.org/care>
- [CERIF] <https://www.eurocris.org/cerif/main-features-cerif>
- [COAR] <https://www.coar-repositories.org/>
- [CODATA] <https://codata.org/>
- [CodeMeta] <https://codemeta.github.io/>
- [Cohen_2019] 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/>
- [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&items_key=00585595 |
| [DOI] | https://www.doi.org/index.html |
| [DONA] | https://www.dona.net/ |
| [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/good-practices/research-institutes/ |
| [Dzale_2017] | Dzale, Yeumo E., Alaux, M., Arnaud, E. 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_6Priorities] | https://ec.europa.eu/info/strategy/priorities-2019-2024 |
| [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] | 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 2 nd 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_HorizonEurope_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-b292-4f70-96c2-3c18fcee06b/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 |
| [EDISON] | https://edison-project.eu/ |
| [eduGAIN] | https://edugain.org/ |
| [EIB_FFS] | European Investment Bank, ‘Financing the future of supercomputing: How to increase the investments in high performance computing in Europe’, 2018, doi 10.2867/31460 https://www.eib.org/en/publications/financing-the-future-of-supercomputing |
| [eInfraCentral] | https://einfracentral.eu/ |
| [eLife] | https://elifesci.org/InnovationLeaders |
| [Enabling_FAIR] | https://osf.io/jy4d9/ |
| [ELIXIR] | https://elixir-europe.org/ |
| [ENVRI-FAIR] | https://envri.eu/home-envri-fair/ |
| [EOSC_CCF] | https://www.eoscsecretariat.eu/funding-opportunities/co-creation-requests |
| [EOSC-Core_Costs] | [URL to be added when report is finalised] |
| [EOSC_EB] | https://www.eoscsecretariat.eu/eosc-governance/eosc-executive-board |

| | |
|--------------------|---|
| [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] | [to be added] https://www.eoscsecretariat.eu/eosc-liaison-platform/post/eosc-interoperability-framework-out-comment |
| [EOSC_Landscape] | 'Landscape of EOSC-Related Infrastructures and Initiatives: Report from the EOSC Executive Board Working Group (WG) Landscape', 2020, ISBN 978-92-76-21289-8, doi 10.2777/342650, KI-03-20-541-EN-N https://op.europa.eu/en/publication-detail/-/publication/cbb40bf3-f6fb-11ea-991b-01aa75ed71a1 |
| [EOSC_Marketplace] | https://marketplace.eosc-portal.eu/ |
| [EOSCpilot_D5.1] | EOSCpilot Deliverable D5.1: Initial EOSC Service Architecture https://eosc-pilot.eu/sites/default/files/eosc-pilot-d5.1.pdf |
| [EOSCpilot_D7.3] | Angus Whyte, Jerry de Vries, Rahul Thorat et al., D7.3: Skills and Capability Framework, 2018 https://www.eosc-pilot.eu/sites/default/files/eosc-pilot-d7.3.pdf |
| [EOSCpilot_D7.5] | Whyte, A., Leenarts, E., de Vries, J. et al., Strategy for Sustainable Development of Skills and Capabilities, EOSCpilot D7.5, 2019 https://eosc-pilot.eu/content/d75-strategy-sustainable-development-skills-and-capabilities |
| [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_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_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 |
| [EOSC-hub] | https://www.eosc-hub.eu/ |
| [EOSC-Life] | https://www.eosc-life.eu/ |
| [EOSC-Nordic] | https://www.eosc-nordic.eu/ |
| [EOSC-Pillar] | https://www.eosc-pillar.eu/ |
| [EOSCpilot] | https://eoscpilot.eu/ |
| [EOSC-Synergy] | https://www.eosc-synergy.eu/ |
| [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/ |
| [ERAvsCoronaAP] | https://ec.europa.eu/info/files/first-eravscorona-action-plan-short-term-coordinated-research-and-innovation-actions_en |
| [ESCAPE] | https://projectescape.eu/ |
| [EuroHPC_JU] | https://ec.europa.eu/digital-single-market/en/eurohpc-joint-undertaking |
| [EuroHPC_SRA] | https://www.etp4hpc.eu/pujades/files/ETP4HPC_SRA4_2020_web(1).pdf |
| [ESRC_Data_Policy] | 'ESRC Research Data Policy' https://esrc.ukri.org/files/about-us/policies-and-standards/esrc-research-data-policy/ |
| [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] | https://www.nature.com/articles/sdata201618 |
| [FAIRsFAIR] | https://www.fairsfair.eu/ |

| | |
|--------------------|--|
| [FAIRsFAIR_D7.1] | Stoy, Lennart, Saenen, Bregt, Davidson, Joy, Engelhardt, Claudia, & Gaillard, Vinciane, D7.1 FAIR in European Higher Education (Version v1.0_draft), 2020. Zenodo. https://doi.org/10.5281/zenodo.3629683 |
| [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/ |
| [FP1] | http://www.jeupiste.eu/horizon-2020-and-around/historical-timeline-framework-programme |
| [FREYA_PG] | https://www.project-freya.eu/en/blogs/blogs/the-pid-graph |
| [FR_OS] | https://www.ouvrirlascience.fr/open-science/ |
| [G7_OS] | http://www.g8.utoronto.ca/science/2017-G7-Annex4-Open-Science.pdf |
| [GAIA-X] | https://www.data-infrastructure.eu/GAIA-X/Navigation/EN/Home/home.html |
| [Genova_2017] | Genova, F., Arviset, C., Almas, B.M., Bartolo, L., Broeder, D., Law, E. and McMahon, B., 'Building a Disciplinary, World-Wide Data Infrastructure', <i>Data Science Journal</i> , 16, p. 16. DOI: http://doi.org/10.5334/dsj-2017-016 |
| [GIDA] | https://www.gida-global.org/ |
| [GODAN] | https://www.godan.info/ |
| [Handle] | http://www.handle.net/ |
| [Hausman_2019] | Hausman, J., Stall, S., Gallagher, J., & Mingfang, Wu, '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/ |
| [IRUS-UK] | https://irus.jisc.ac.uk/ |
| [ISO_16363] | http://www.iso16363.org/ |
| [Leiden_Manifesto] | http://www.leidenmanifesto.org |

| | |
|----------------------|--|
| [LIBER_Copyright] | https://libereurope.eu/blog/2020/04/14/copyright-coronavirus-statement/ |
| [Lindau_Dec] | http://lindau-declaration.org/ |
| [Mozilla_OL] | https://foundation.mozilla.org/en/initiatives/mozilla-open-leaders/ |
| [nestor_Seal] | https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html |
| [NI4OS-Europe] | https://ni4os.eu/ |
| [NL_OS] | https://www.openscience.nl/ |
| [OCRE] | https://www.ocre-project.eu/ |
| [OECD_BDWC] | OECD. (forthcoming). Building digital workforce capacity and skills for data-intensive science. OECD Publishing, Paris. |
| [OECD_DSTI] | OECD report: 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_RG] | https://www.openaire.eu/blogs/the-openaire-research-graph |
| [OpenAire_UC] | https://usagecounts.openaire.eu |
| [Open_Life_Sci] | https://openlifesci.org/ |
| [OpenRec_BE] | https://be.openrecognition.org/ |
| [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.panosoc.eu/ |
| [Plan_S] | https://www.coalition-s.org/ |
| [Plomp_2019] | Plomp, Esther; Dintzner, Nicolas; Teperek, Marta and Dunning, Alastair, ‘Cultural Obstacles to Research Data Management and Sharing at TU Delft’, <i>Insights</i> 32 (1): 29, 2019. DOI: http://doi.org/10.1629/uksg.484 |
| [Qlik_Data_Literacy] | The Human Impact of Data Literacy, 2020, Qlik & Accenture https://thedataliteracyproject.org/files/downloads/Qlik_Accenture_Human_Impact_of_Data_Literacy.pdf |
| [RDA] | https://www.rd-alliance.org/ |
| [RDA_FAIR_DMMWG] | https://www.rd-alliance.org/groups/fair-data-maturity-model-wg |
| [RDA_MDC] | https://makedatacount.org/ |
| [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>
- [RFII_DG] <http://www.rfii.de/download/digital-competencies-urgently-needed-october-2019/>
- [Saenen_2019] Saenen, B., Morais, R., Gaillard, V., and Borrell-Damián, L., '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>
- [Sempreviva_2017] Sempreviva, Anna Maria; Vesth, Allan; Bak, Christian; Verelst, David Robert; Giebel, Gregor; Danielsen, Hilmar Kjartansson; ... Hermans, Koen W., 'Taxonomy and metadata for wind energy R&D', 2017, December 12. Zenodo. <http://doi.org/10.5281/zenodo.1199489>
- [Software_Heritage] <https://www.softwareheritage.org/>
- [SSHOC] <https://www.sshopencloud.eu/>
- [Stoy_2020] Stoy, Lennart; Saenen, Bregt; Davidson, Joy; Engelhardt, Claudia & Gaillard, Vinciane, 'D7.1 FAIR in European Higher Education (Version v1.0_draft), 2020. Zenodo. <https://doi.org/10.5281/zenodo.3629682>
- [TB-L_Step] <https://inrupt.com/blog/one-small-step-for-the-web>
- [UN_SDG] <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>
- [UNESCO_OS] <https://en.unesco.org/science-sustainable-future/open-science>
- [UvdL_Agenda] https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf
- [WDS] <https://www.worlddatasystem.org/>
- [Wellcome_ORF] https://wellcome.ac.uk/grant-funding/people-and-projects/grants-awarded?scheme_id=3580
- [WG_FAIR_Report] EOSC FAIR Working Group, 'Six recommendations for implementation of FAIR practice', 2020, DOI: 10.5281/zenodo.3904139
<https://zenodo.org/record/3904140#.X2mcmvKhPY>
- [WG_Sustain_Report1] <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_Report2] <https://www.eoscsecretariat.eu/file/solutionsforasustainableeosc-inmandraft02dec19pdf>. 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_Report3] https://www.eoscsecretariat.eu/system/files/solutions_for_a_sustainable_eosc_-_ironlady_draft_19oct20.pdf
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_CoP] | https://en.wikipedia.org/wiki/Community_of_practice |
| [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 |
| [WISE] | https://wise-community.org/ |
| [WISE_AUP] | https://wiki.geant.org/download/attachments/123765566/WISE-SCI-Baseline-AUP-V1.0.1-draft.pdf |
| [WISE_RA] | https://wise-community.org/risk-assessment/ |
| [WISE_SCI] | https://wise-community.org/sci/ |
| [Zenodo] | https://about.zenodo.org/ |