EOSC Future

D2.7a
Recommendations for the SRIA

The EOSC Future project is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020, Grant Agreement number 101017536

Version 1.0
November 2021

D2.7a / Recommendations for the SRIA

Lead by **GÉANT**
Authored by Sarah Jones (GÉANT)
Reviewed by Ron Dekker (TGB), Gareth O'Neill (TCB), Franciska de Jong (CLARIN) and Athanasia Spiliotopoulou(JNP)

Dissemination Level of the Document

Public

Abstract

This report provides strategic recommendations on priorities for EOSC development to inform the forthcoming Horizon Europe work programme for 2023/24. The deliverable was brought forward to ensure recommendations could be included within the timeframe specified by the European Commission to release a draft work programme in June 2022. Inputs were solicited from each Work Package and the key advisory bodies of EOSC Future, namely the Strategy and Oversight Board (SOB) and Technical Coordination Board (TCB).

Version History

Version	Date	Authors/Contributors	Description
V0.1	19/10/2021	Sarah Jones (GÉANT)	Initial draft for consultation with WP partners
V0.2	31/10/2021	Rudolf Dimper (ESRF), Mark van den Sanden (SURF), Tiziana Ferrari (EGI), Najla Rettenberg (RDA), Iryna Kuchma (OpenAIRE), Antti Pursula (CSC), Paolo Manghi (OpenAIRE)	Additions to the recommendations from EOSC Future
V0.3	06/11/2021	Gareth O'Neill (TGB), Franciska de Jong (CLARIN)	Review of document
V0.4	07/11/2021	Sarah Jones (GÉANT)	Processing comments and amendments
V0.5	17/11/2021	Sarah Jones (GÉANT)	Circulation for internal QA
V0.6	21/11/2021	Ron Dekker (TGB), Athanasia Spiliotopoulou (JNP), Sarah Jones (GÉANT)	Final reworking of comments
V0.7	25/11/2021	Sarah Jones (GÉANT)	Final version for circulation to consortium
V1.0	01/12/2021	Sarah Jones (GÉANT), Ron Dekker (TGB), Mike Chatzopoulos (ATHENA)	Final Version submitted to EC

Copyright Notice

Table of Contents

Table of Tables

Table of Figures

List of Abbreviations

Acronym	Definition
AI	Artificial Intelligence and Machine Learning
AAI	Authentication and Authorization Infrastructure
BPA	BluePrint Architecture – referring to AARC for authentication and authorisation
CSA	Coordination and Support Action
EOSC-P	EOSC Partnership
FOSB	Flemish Open Science Board
HPC	High Performance Computing
HTC	High Throughput Computing
MVE	Minimal Viable EOSC
QA	Quality Assurance
R&R	Rewards and Recognition
RI	Research Infrastructure
SOB	Strategy and Oversight Board – a governance body in the EOSC Future Project
SKGs	Scientific Knowledge Graphs
SRIA	Strategic Research and Innovation Agenda – a document to define the general framework for future research, development and innovation activities in relation to the EOSC
TCB	Technical Coordination Board – a governance body in the EOSC Future Project

1 Executive Summary

This deliverable collates feedback from the project consortium to the current Strategic Research and Innovation Agenda (SRIA) for EOSC, identifying which aspects are already addressed by the EOSC Future project and where additional activities are needed. It has been scheduled to coincide with the European Commission's consultation processes to seek inputs for the Multi-Annual Roadmap for 2023-24 to help inform the shape of the forthcoming work programme.

Background context on the initial development of the SRIA and the timlelines for contributions is provided in the introduction. Section 3 describes the current state of play by examining the existing priorities and how these are addressed in the EOSC Future project. The preliminary topics proposed by the EOSC Association are also outlined in this section, since these were taken as inputs for consortium feedback. The main chapter of relevance is section 4 which outlines the recommendations coming from the EOSC Future project.

The eleven priorities noted by the project focus on:

- widening EOSC through strategic alignments with the Research Infrastructures, data spaces and research performing organisations,
- providing better connections between EOSC services and resourcing models to enable more seamless access across them,
- ensuring trusted digital repositories that support quality data and long-term preservation,
- addressing the human support and skills needed to facilitate uptake of EOSC,
- enabling tools that help monitor research activity to act as an incentive.

The recommendations were shared with the DG-RTD and DG-CNECT of the Commission in early November to act as inputs to their considerations. A first draft of the Work Programme will be available in early 2022.

2 Introduction

In 2020, the EOSC Executive Board in collaboration with its Working Groups drafted a Strategic Research and Innovation Agenda for the European Open Science Cloud. The core purpose of the SRIA is to provide input to develop the Work Programmes for EOSC in Horizon Europe. An extensive consultation was held with the wider community in which hundreds of responses were received and processed. The SRIA was finalised in Spring 2021.[1] It provides some background context on Open Science and how EOSC has been developed, as well as overarching strategic objectives, guiding principles and 14 core challenge areas around which further work is recommended. The SRIA concludes with a multi-annual roadmap which prescribes which actions should take place in which phase of implementation and notes the level at which they should be addressed i.e. by the European Commission, by Member States or by institutions.

This deliverable is intended as input to the revision of the SRIA. As the flagship EOSC project addressing many aspects of the SRIA as noted in chapter 2, it is important that the EOSC Future project consortium reflects on what activities should be prioritised next. The deliverable was originally scheduled for June 2022 (M15) when the next iteration of the work programme is due to be released. In order to influence the shape of that Work Programme, however, inputs are needed much earlier to feed into EC meetings and drafting processes. The EOSC Partnership (EOSC-P) is required to provide inputs before the EC internal meeting on 20th November 2021, so this deliverable has been brought forward to be developed in October and November 2021. The complete internal QA process for the deliverable will occur after sharing the core recommendations with the European Commission in order to maximise the time available to collect inputs from the project members. The deliverable is planned to be finalised and submitted in December 2021 (M9).

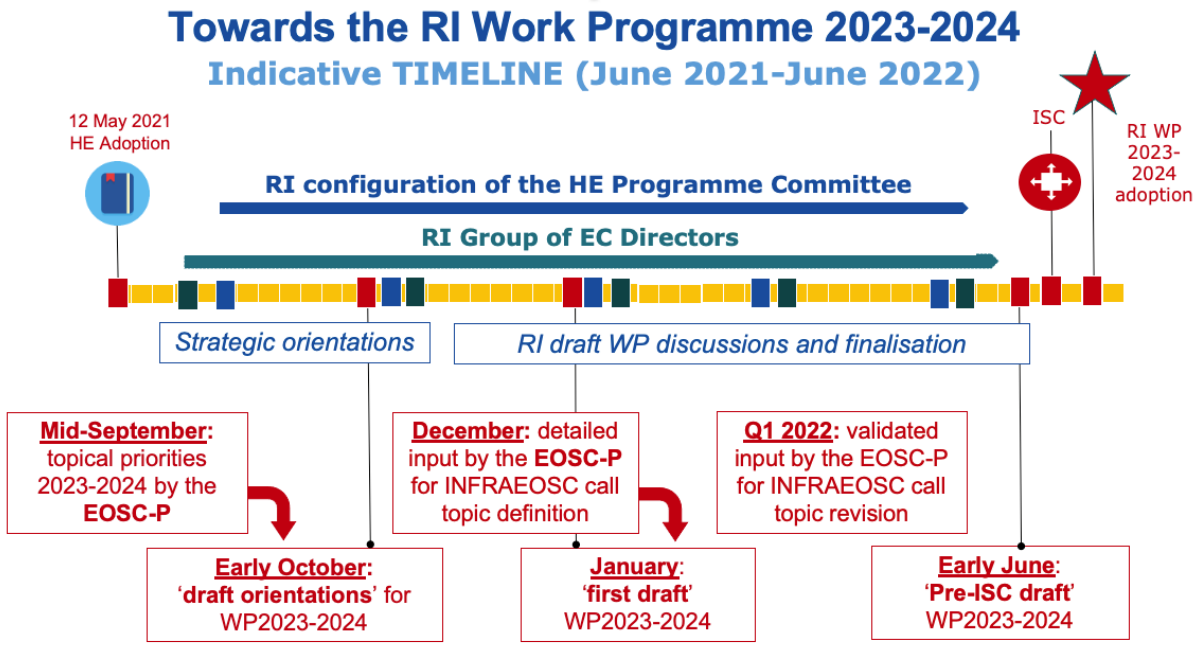

Figure 2.1: European Commission timeline for developing the EOSC Work Programme 2023-24

*Note that input by the EOSC-P is required by November not December 2021 as shown in the diagram.

[1] Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC), v1.0, 15th February 2021, https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf

3 The current state of play

This section outlines the current state of play, explaining what is recommended in the SRIA and where those activities are being addressed in EOSC Future. It also notes priorities put forward by the EOSC Association which were used as inputs for the recommendations coming from EOSC Future as outlined in section 4.

3.1 SRIA February 2021

The EOSC objectives tree notes three core objectives and contextualises these by identifying the problems and barriers they seek to address, as well as the resulting benefits to be gained by meeting them. At a general level, these objectives address the human and policy dimension, the data and content to be exploited and the infrastructure and services needed to enable this. These objectives should be realised by the EOSC Partnership which is a collaboration between the European Commission, on behalf of the European Union, and the EOSC Association together with its members.

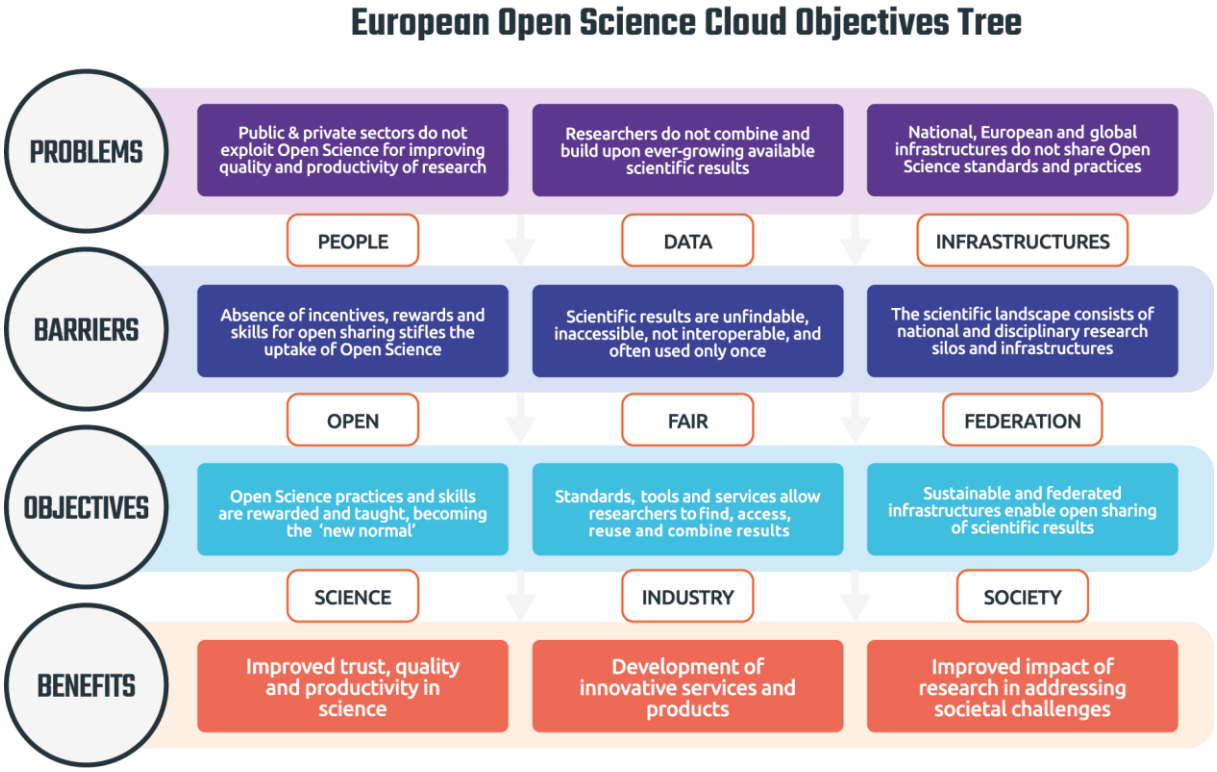

Figure 3.1: The European Open Science Cloud Objectives Tree

In order to address these overarching objectives, there are 14 action areas noted in the SRIA. Several of these are relevant to and can be mapped against EOSC Future project activities as noted in the table 3.1 below.

Table 3-1: SRIA priorities mapped against EOSC Future activities

Ref	Topic	Recommendations	EOSC Future activities
5.1	Identifiers	– Establish mature and recognised PID infrastructures for emerging resource types. – Develop a 'meta resolver' that can deal with any type of relevant identifier. – Define specifications for PID records. – Produce type definitions for the most common data formats. – Support the creation and use of a PID graph – Develop tools to certify PID infrastructure.	Draft interoperability framework for a PID meta-resolver and guidelines for PID service providers for minimum Kernel Type Information (D3.2).

5.2	Metadata and ontologies	– Develop governance structures to coordinate the work on metadata and ontologies. – Support registries of metadata schemas and ontologies with clear protocols for harvesting, crosswalks and metadata management. – Maximise the re-uptake of information in systems that communicate in a semantically interoperable manner. – Develop EOSC guidelines for a minimum metadata description. – Develop services that build on metadata registries.	Guidelines for minimum metadata to support the discovery, metadata exchange, and cross-walks of research products across communities (D3.2).
5.3	FAIR metrics and certification	– Assess and test the proposed EOSC FAIR data metrics in a neutral forum. – Support the definition and implementation of evaluation tools. – Support the definition of FAIR for software. – Support data and service providers to progress in the FAIRness of their holdings. – Align repository certification standards and assessment schemas with FAIR. – Establish core criteria to certify other key elements of the FAIR ecosystem. – Establish registries of certified components of the ecosystem.	
5.4	Authentication and authorisation infrastructure (AAI)	– Establish and implement a common framework for managing user identity & access. – Ensure long-term attribute availability, assurance, freshness and provenance. – Scale the current proxy (BPA) architecture and supporting infrastructure. – Address user experience challenges. – Provide solutions for identity beyond the research and education community. – Enable identity for the individual scientists. – Develop future trust fabrics and authorisation models in support of dynamic collaborations.	EOSC federated authorisation and authentication, enabling seamless integration of community AAI with EOSC AAI federation (D7.3).
5.5	User environments	– Integration of existing catalogues and portals. – Information about resources should be aggregated to enhance discovery. – The distributed architecture model must address legal and organisational frameworks to enable researchers to compose resources. – Strong engagement and consultation is required to ensure interoperability and integration of portals, thematic and regional community services and resources.	Science cases for development of EOSC Architecture (D3.1). Front office requirements analysis, leading to design spec and software release (D5.1, D5.2, D5.3).
5.6	Resource provider environments	– Ensure more efficient onboarding of resources and integration with existing community catalogues and repositories. – Enable the composability of resources and across resource providers.	EOSC providers software and resource catalogue Workflows developed from the moderation process to onboard

		– Ensure the guidelines for resource providers are appropriate to and respectful of the existing interoperability frameworks available at a community level. – Incentivise resource providers to produce and operate resources that are Open Science by design.	external providers (D6.2). Procurement plan for additional services (D8.3).
5.7	EOSC Interoperability Framework	– Use open specifications to ensure technical interoperability. – Define a common security and privacy framework. – Govern and maintain repositories of semantic artefacts. – Define clear protocols for the federation / harvesting of these repositories. – Define a list of EOSC-recommended licences. – Implement minimum standardised, human- and machine-readable expressions of right statements and use conditions.	EOSC Architecture and Interoperability Framework, including PIDs, AAI, Metadata and Ontologies, Monitoring, Processing, Compute and Publishing (D3.2)
6.1	Rules of Participation	– Apply the RoP to all digital resources made accessible via EOSC. – Define a minimum set of rights, obligations and accountability for all those participating. – Evolve the RoP to include aspects on FAIR, Terms and Conditions and Acceptable Use. – Provide a governance structure for the RoP and coordination of EOSC projects.	Onboarding procedures and workflows (D6.1 & D6.2). Inventory of Core Functions and Inclusion Criteria (D2.5).
6.2	Landscape monitoring	– Ensure continuous monitoring of the existing readiness of countries to contribute to EOSC. – Stimulate the definition and implemention of EOSC-related policies and strategies. – Stimulate EOSC-dedicated funding streams and criteria in national programmes. – Stimulate dedicated funding streams for Open Science at institutional level. – Stimulate investment in national infrastructure contributing to EOSC.	EOSC Observatory and mappings to feed into this (D2.2, D2.3, D2.8). Open Science Monitor (included in D5.3).
6.3	Funding models	– Perform cost assessments for the EOSC Core services and Minimum Viable EOSC. – Develop financing schemes for EOSC. – Develop monitoring schemes for the in-kind contribution of members. – Develop synergies between national and EC funding streams.	
6.4	Skills and training	– Develop the next generation of Open Science and data professionals. – Coordinate and align curricula for students and researchers. – Build a trusted and long-lasting knowledge hub of learning materials and related tools. – Influence national Open Science policy for skills by supporting strategic leaders.	EOSC Knowledge Hub (training catalogue and learning platform) (D9.2). EOSC Training Programmes Results (D9.3).

6.5	Rewards and recognition	– Demonstrate leadership in enacting change towards a culture of trust, openness and risk taking. – Prepare Human Resources to adjust Rewards & Recognition (R&R) structures (including approaches to recruitment and promotion) using next generation and progressive metrics. – Safeguard institutional autonomy. – Identify and remove legal obstacles to empower researchers, organisations and funders to develop and refine better R&R systems.	
6.6	Communication	– Set up a Strategic Communication Plan. – Inform stakeholders about the developments of EOSC. – Develop and deploy communication channels. – Work on licensing and ownership issues.	Liaison with EOSC Association on communications (D10.1). Support of EOSC Symposia (D10.2).
6.7	Widening to public and private sectors and going global	– Widen EOSC stakeholder engagement in a strategic and timely manner, noting that consultation placed this action lowest in relevance for the immediate future.	Alignment with MS/AC & strategic initiatives (D2.4). EuroHPC engagement plan (D2.6). Procurement plan and Business Pilots for additional services (D8.3). Digital Innovation Hub final results (D8.5). Procurement plan for data spaces (D8.6).

3.2 EOSC Association Board recommendations, September 2021

The EOSC Association provided early inputs to the European Commission in September 2021 in the form of a 2-page statement lisiting priority areas for development. These were presented to the EOSC Association Members in a webinar on 15th October 2021 and shared in abbreviated form with the EOSC Future consortium as noted below. There are eight topical priorities recommended (the priority on a second CSA to support the Association was omitted from internal EOSC Future discussions). A summary of the EOSC Association recommendations are provided below as a reference to what was shared with the EOSC Future consortium.

Science clusters and the long-tail

Ensure that EOSC engages research communities and grows the user base by pursuing cross-cluster use-cases or demonstrators, citizen science initiatives, and engaging non-obvious disciplines such as economics, arts and humanities. Intermediaries like University Associations and Learned Societies will be used to help engage the long-tail.

Quality of research data

Explore what are the most relevant quality dimensions considered in different communities and how they can be incorporated in EOSC by identifying common measures and aligning them with the FAIR principles.

Quality of research repositories

Define indicators for quality research repositories by developing standards for FAIR certification of repositories and the data within them. This will enable research repositories to be recognised as an authoritative source of quality data which is preserved for the long-term.

Long-term access to and preservation of data
Aims to support preservation actions to ensure EOSC content is available in the long-term by promoting a network of trustworthy digital repositories that follow FAIR principles and ensure long-term access.

Quality of software
Improving the quality of software by fostering the development and deployment of infrastructure, tools and services that allow researchers to properly develop, describe with proper metadata, archive, share and reuse research software.

Linking to sensitive data from public authorities
Aims to link data repositories from public authorities to EOSC since valuable large-scale sensitive data collections often reside in repositories which are not well connected to scientific data infrastructure.

Enabling AI and access to un-structured and non-standard data
Aims to address the lack of structure and heterogeneity of non-standard data (e.g. social media) to make it accessible using AI. This activity area also aims to ensure the AI algorithms are themselves managed and made accessible in the same way as data and scientific publications. This will enable the community to implement knowledge graphs to link these entities to one another.

Multilingualism in EOSC
Since EOSC is a multi-country endeavour, the aim is to increase the visibility of content from countries working in their own language by improving use of ontologies, multi-lingual thesauri, controlled vocabularies and language mapping.

4 EOSC Future recommendations

Below is a list of priority areas that the EOSC Future consortium put forward for consideration in the 2023/24 work programme. These can be categorised as focusing on:

- widening EOSC through strategic alignments with the RIs, data spaces and research performing organisations;
- providing better connections between EOSC services and resourcing models to enable more seamless access;
- ensuring trusted digital repositories that support quality data and long-term preservation; addressing the human support and skills needed to facilitate uptake of EOSC; and
- enabling tools that help monitor research activity to act as an incentive.

The main topic areas and inter-relations between them are visualised in the diagram below.

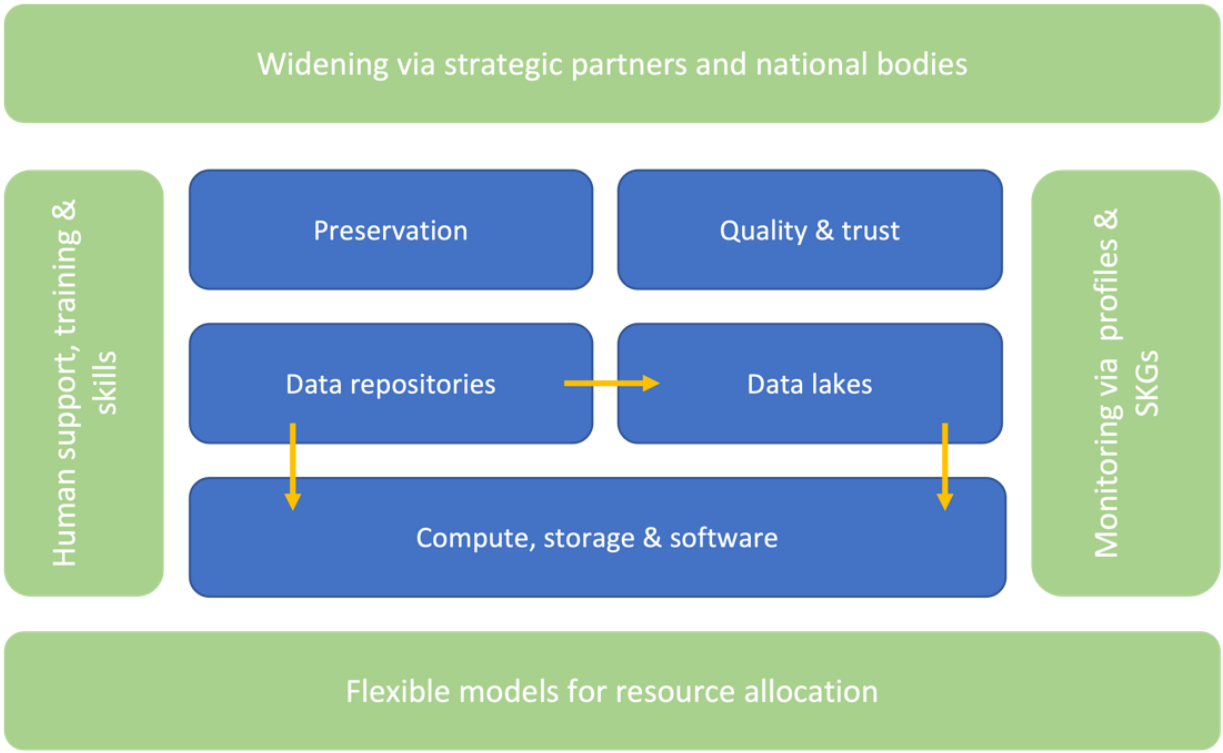

Figure 4.2: EOSC Future priority topics

4.1 Widening the EOSC user base

EOSC needs to grow its user base and connect with other related initiatives to bring together data from different sectors required to address societal challenges. The focus in 2023-24 should be on embedding EOSC into the core operations of Research Infrastructures, Missions and Partnerships. By collaborating with data spaces and other strategic initiatives, EOSC actions can help to address societal challenges.

The existing work to federate e-Infrastructures and Research Infrastructures in EOSC Future should be built on further to pursue a domain or sectoral focus addressing global challenges. Distributed RIs and e-Infrastructures have the potential to contribute to local and regional socio-economic development. There is an opportunity to develop strategies for ESFRI / EOSC cooperation in the creation of local / regional knowledge hubs by aligning advanced facilities, data capabilities and upskilling highly qualified staff.

European Missions and Partnership projects can serve as use cases and proof of concepts to establish best practices to be adopted in the context of integrating heterogeneous datasets for advanced AI aligned with global environmental, social and economic challenges, demonstrating the value of EOSC and FAIR data.

Support should also be offered to develop solutions for international access to public data in trusted repository environments and to foster agreements with industry and data intensive organisations as part of overall

widening efforts. There is currently scope for joint activity between African and European RI nodes on the topic of climate change in the INFRA-2021-DEV-01-02 call, and similar infrastructural collaborations could be pursued in the forthcoming work programme to support the widening of EOSC to other international partners.

4.2 Engaging national bodies and institutions

At this moment, EOSC is still predominated by large scale research infrastructures and horizontal services. With the formation of the EOSC Association and the strong number of members from research performing organisations, an opportunity exists to better connect local support infrastructure into EOSC.

More should be done to seek connections on a national level with the universities, research communities and individual researchers. Leveraging connections with national research funders, service providers such as NRENs, NGIs and NOADs and representative bodies such as the European Universities Alliance would help to engage the long tail of science supported within research performing organisations. The recent study on EOSC National Structures provides many starting points, as do the Mandated Organisations in the EOSC Association membership and fora such as CoNOSC.[2]

By bringing institutional support infrastructure into closer alignment with EOSC, better leverage can be made of institutional support staff such as data stewards to grow capacity for supporting FAIR and Open Science, as noted in the later Human Infrastructure section. Models such as the Dutch network of Digital Competence Centres[3] could be explored to establish closer links between support in research organisations and EOSC. Regional level engagement to support coordination activities will help with practical EOSC onboarding issues.

4.3 Flexible models for resource allocation

EOSC is presented as an environment in which researchers can access data and associated services, but it is unclear how access to such services will be funded and allocated. There is no model for individual researchers or small teams who may want to look at the data that is accessible via EOSC to be provided with the necessary compute. Clear requirements processes exist for researchers to apply for access to compute resources through large research infrastructures, but there are not yet defined processes for how others could transparently access IT capacity to visualise and process the data within the context of EOSC. This is foremost a problem for small datasets, citizen science or exploratory actions, as those who are processing terabytes of data will follow the existing processes.

Access to services is still managed through individual providers and will not permit one RI or national compute facility to finance the IT resources for processing data from others, or by users who are not in their remit to support. If the end users cannot easily access resources within EOSC, the practice of transferring the data to the local PC/laptop will prevail and the benefits of providing an environment where data can not only be found but also accessed and processed will not be achieved.

Within the INFRAEOSC-07 projects, resourcing models are being tested to provide services free at the point of use, however this can only be offered for services included within those projects. Other providers are not eligible to receive funding so it is not an open model for services and resources made available within the EOSC-Exchange. The procurement model being explored for EOSC-Core is also mostly suited towards commercial service providers, whereas many of those currently operating in EOSC come from the science domain and are not permitted to participate in commercial procurements.

Although further activity on resourcing models is planned in forthcoming projects, ongoing effort is needed to fully implement these. Service providers also need access to expertise to help them understand the business models and legal issues in providing services across borders in the EOSC context. The funding and resource allocation models developed for EOSC must be supported by the EC and be aligned with granting mechanisms from Member states.

[2] https://conosc.org/
[3] https://www.lcrdm.nl/dcc

4.4 Better connections between services for storage, compute and software

Researchers work in a variety of environments to conduct their work and these need to be effectively brought together in EOSC, with clear connections between facilities and institutions to enable researchers to transfer data for analysis to different locations. Many researchers may generate terabytes of data in national research facilities and then lack access to the necessary compute power and tools to process it in their home institution. The EOSC ecosystem should be an interface between Research Infrastructures (RIs), e-Infrastructures, High Performance Computing (HPC), a wide variety of tools and service providers, and the institutions where researchers are based and conduct their work. This will allow data to be easily stored, transferred, combined, processed, published, preserved and shared, and allow researchers to find the necessary resources to do so, independently of where they come from.

An interoperable authentication mechanism is needed so that scientists can access EuroHPC services alongside EOSC provision. EOSC Future has dedicated resources to investigate the links between EuroHPC and EOSC. To this end, several pilots have already started to investigate synergies across multiple scientific domains which could be built on in future programmes.

Europe has significant data and computing resources that are of vital benefit to European research communities, however today's infrastructures need significant new investments to scale up and address the data deluge expected in the coming decade. Member States should expand national investments for technological and research leadership, and with this incentivise transnational access to cloud and high-throughput data processing facilities via EOSC that would be otherwise only accessible to national user communities. As the volume, variety, velocity and volumes of data are increasing, EOSC needs to distribute and decentralise in a secure manner the processing of high-quality research data in a computing continuum offering management of hardware heterogeneity.

EU cross-border federated cloud and edge infrastructure with additional capacity needs to expand national capacity, offer transnational access and be fully integrated with national research clouds, scaling them up with additional nationally procured capacity and offering integrated HTC, HPC, cloud and edge access.

The EOSC compute platform should integrate with the EOSC federated data lakes and data repositories to democratise the analysis of large-scale core data resources and retain European research data sovereignty. Data from multiple distributed sources needs to be brought together at a suitable processing and analysis platform supporting Digital Twin platforms, Virtual Observatories and scientific virtual research environments.

Such a platform will need to expand with the contribution of EU13 and EU widening countries. The platform also needs to align with the data spaces, with integrated solutions for data lake management, orchestration, federated AAI, and a portfolio of scientific software, models and Digital Twin engine capabilities as well as EuroHPC systems.

4.5 EOSC Federated Data Lakes and Data Repositories

The goal of the EOSC Future project is to realise an operational platform with an integrated execution environment consisting of data, services, open research products and infrastructures. In order to achieve its mission, EOSC needs to equip researchers with a European network of trustworthy research digital repositories for data storage and analytics with disciplinary and geographical spread. The network needs to offer federated discovery, access and exploitation of data, bringing data next to computing and analytics services with related educational tools and user support. This includes the provisioning of Sensitive Data Processing as a Service and data anonymisation services to ensure EOSC can be a trusted environment offering high-level standards of security and protection to privacy-sensitive data of research relevance.

Alignments can be made here with the objective of widening to public sector and industrial data, agreeing on a data lake reference model, architecture, services and standards for securely accessing and sharing the data. The data repositories should connect to data lakes by delivering data that can be deployed on compute and storage resources as a package for further anaylsis. The data lakes should bridge with repository services to provide the infrastructure for exploitation of the data.

A portfolio of data provisioning agreements with key data providers such as Research Infrastructures, scientific collaborations and research performing organisations, guaranteeing the delivery and support of core data resources and related tools in EOSC, will be paramount to ensure the operationalisation of EOSC.

To expand its data capacity, EOSC needs to facilitate the federation of a growing amount of data and repositories by providing innovative technologies offering built-in EOSC federation capabilities for the on-demand or in-house provisioning of digital repositories as a service. The federated data resources ensure cross-domain collaboration across sectors including industry and SMEs and evidence-based policy makers, and need to integrate with external data spaces from other sectors aligning towards a common data blueprint architecture and reference implementation of data spaces.

4.6 Long term data preservation

As noted in the EOSC Association Long-term Data Preservation Task Force charter, the possibility to reproduce, replicate and re-use scientific results depends on the long-term accessibility and assessibility of the underlying data. The SRIA underlines the importance of long-term data preservation, but an explicit strategy has not been formulated yet for EOSC.

Digital preservation has been a key area of investment in EC programmes for several years, but these activities should now also be directed to the research data made available via EOSC with specific programmes of activity focusing on digital preservation services aligned with research repositories. Digital preservation has been a key area of investments via ESFRI. These activities should now be connected with EOSC to ensure that long term preservation is ensured in all science domains, that it can be applied to new data types, and data are accessible from other domains in order to facilitate and promote re-use of data over domains.

Further research is needed to develop clear processes that identify which data, software and other outputs have long-term value and should be preserved. This can build on work developed in areas such as the life sciences and social sciences. Developing such procedures will help to scale preservation actions accordingly in EOSC going forward. A selection needs to be made since the resources needed to preserve data and other research objects are significant. This selection process will be an ongoing, dynamic task with the decision about what to preserve evolving over time. When there is an initial picture of what needs to be preserved, specialised digital preservation repositories which monitor format obsolesence and apply approriate preservation actions should be applied to this subset of content. A network of preservation repositories which can integrate with other infrastructures to safeguard the data and resources made available in EOSC should be fostered.

4.7 Quality and trustworthiness of data and services

The data within EOSC needs to be of a high quality to be reliable and should be preserved so it can be referenced and reused in the long-term. Research communities differ in the measures they apply to assess data quality, but common characteristics and ratings could be explored to enable an indication of community trust in the resources. Certification models like CoreTrustSeal were recommended for adoption in EOSC by the *Turning FAIR into Reality Expert Group*[4] report and should be pursued further.

We need to be able to trust the services which store our data for the long-term, as well as measuring the quality and FAIRness of the data itself via community review and the application of metrics. Having certified services that guarantee a good quality of FAIR implementation within their area of expertise could be one route to ensuring ongoing trust in the quality of the data and the repository services. The recommendations here are in line with existing priorities flagged by the EOSC Association on the quality of research data and research repositories, but emphasise the trust angle.

4.8 Human support infrastructure

All actors need to have the appropriate skills and competencies needed to interact with EOSC and increase uptake of resources. This should be done in a coordinated way to ensure the less technical communities or no

[4] https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283

countries are left behind, levelling the playing field. The human infrastructure that supports the use of technology is critical and often overlooked or under-resourced. The purpose of this action is to widen and deepen the EOSC user base by running a European coordinated technical support programme to users. Offering coordinated 'last mile' support to help research groups put services in place is incredibly important to integrate multiple resources from EOSC with existing and future user environments. This requires expertise to integrate resources and services, involving data and service experts from Research Infrastructures, e-Infrastructures, emerging national EOSC nodes and other EOSC providers.

Tailored programmes will be necessary to work with users from different groups including research collaborations, long tail of science, citizen science, private entitites and evidence-based policy makers. The support infrastructure needs to integrate national human networks participating in e-Infrastructures and Research Infrastructures and expand their capacity to support users and their digital needs, provide bespoke solutions, develop and uptake innovation, and support the user uptake of standards, products, and services. Human support infrastructures are already available from individual infrastructure providers, but support will need to be provided in an integrated, coordinated way for EOSC to help put services in place as well as offer support for users needing and combining services from across individual infrastructures.

We cannot simply list services and assume users know what they need and how to combine them. Facilitation through a coordinated or federated helpdesk, data services providers, and via intermediaries such as data stewards and lab managers who work closely with research groups and could act as a broker or liaison with EOSC infrastructure providers is recommended.

The responsibilities for such provisioning touch on all three levels applied in the Multi-Annual Roadmap: the European Commission supports the human support needed to engage and interact with EOSC-Core services, while Member States, national funders and institutions are responsible for the provision of local support. Models like the Flemish Open Science Board (FOSB) annual investment in Open Science, a portion of which is dedicated to funding data stewards in universities, is recommended for increasing human capacity. Connecting between the local-national-European support layers is also key to promote alignment of curricula and skills sharing.

4.9 Training and capacity building

One of the key recommendations of the UNESCO Draft Recommendation on Open Science is 'Investing in human resources, training, education, digital literacy and capacity building for open science'. To this end, and complementary to the previous subsection, there is a need to ensure researchers have the basic skillsets to create a level playing field across borders and disciplines. Moreover, in relation to the subsection on widening and deepening EOSC's user base, it is critical for the future of EOSC to propagate the core values of open science and FAIR. For all these factors, easily accessible and sustainable quality training is essential.

Infrastructures consist of people, processes and technology that care for digital objects, enabling seamless access to data, publications, and software. Every science cluster has concerted training efforts but an opportunity exists to create joint training materials and run cross-discipline and cross-cluster events. There is a need for cross-community knowledge sharing in order to establish best practices, and subsequent training to ensure compliance and uptake of the best practices.

We could start from a common core set of best practices and materials and adapt to each setting rather than everyone needing to implement everything from zero. It would also promote the use of cross-cluster use cases, supported by data stewards and data scientists.

It is recommended that projects focusing on the development of data stewardship activities, competences and best practices find adequate support and funding in the upcoming INFRAEOSC calls in 2023-2024, to stimulate and foster cross-country, cross-discipline collaboration and alignment of practices. Recognising these data stewardship roles and professionalising career paths is also key to incentivise further development.

A limited set of core training infrastructure also needs to be supported with permanent sustainable funding to enable the pooling of training content and learning platforms. This could be advanced on a consortium basis of overarching disciplinary infrastructure rather than an EC funded activity, building on the Knowledge Hub advanced through EOSC Future.

4.10 Monitoring and assessing Open Science via Scientific Knowledge Graphs

Monitoring and tracking the evolution of science is essential to ensure its transparency, which leads to proper attribution, reproducibility, assessment and evaluation.

Scientific Knowledge Graphs (SKGs)[5] are taking over as a flexible tool to track facts about science, from the publishing of an article and research data, to the semantic links between the two, reaching out to the contextual information on researchers, organisations, funders, and services to offer an interlinked graph of research activity. SKGs may have a regional, cross-disciplinary, or disciplinary character, in order to serve the applications for which they were engineered. These may range from research impact and assessment to discovery and reproducibility.

Specific activity areas proposed for the 2023-24 work programme are:

- Construction of domain-specific and national (or regional) SKGs;
- EOSC catalogues and SKGs;
- Interoperability of SKGs:
  - *Common models:* agreed metadata formats, protocols to share metadata about research-related entities and scientific organisations e.g. publications, data, software, workflows, computing, storage, organizations, projects, funders, services, researchers, facilities, companies;
  - *Common tools:* Artificial Intelligence, Full-Text Mining and Natural Language Processing for SKGs: workflows, models, software, tools, provenance, and reproducibility.
- SKG-based applications:
  - Monitoring and accounting, tracking events of science, publishing, citation, data processing, data and software usage, service consumption;
  - Impact services on top of SKGs: Open Science indicators and bibliometrics, quality, performance, impact, popularity, etc.

4.11 EOSC Researcher profile

This activity is suggested to develop profiles for researchers in the EOSC, connecting to national and institutional identifiers or ORCID, through which researchers can login across services (EOSC AAI), track and expose their scientific career and results, and participate to projects at EC and country level. The profile can be used to apply for funding via the EC Participant Portal, or vice versa be contacted for expertise.

It should support an infrastructure where added-value services can be built (e.g. tools to receive notifications, and to which national or research infrastructures can connect) to assign profiling rights and support assessment and evaluation. Similar to the Knowledge Exchange Openness Profile,[6] an EOSC researcher profile will also assist in attributing credit and building the recognition and rewards systems which are essential to motivate engagement in Open Science and FAIR. The profile can offer a portfolio of a researchers' outputs and activities, compiled through the various associated identities, helping to recognise contributions and support wider collaborations and engagement.

[5] For a definition of knowledge graphs see: https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph
[6] https://www.knowledge-exchange.info/event/openness-profile

5 Conclusions

The EOSC Future project is establishing the MVE of the EOSC-Core, EOSC-Exchange and Interoperability Framework. The development of EOSC however needs to advance beyond the MVE, hence the identification of several priority areas for the forthcoming Work Programme.

Several priorities noted by the EOSC Future consortium focus on widening EOSC via strategic partnerships. The Research Infrastructures are put forward to play a key role here since they have the potential to contribute to local and regional socio-economic development, while the European Missions and Partnership projects are envisaged as use cases to demonstrate the value of EOSC and FAIR data. Engaging national bodies and research performing organsiations is also prioritised as a way to better connect to the long-tail of science.

Activities on developing skills and human infrastructure were seen as a gap in the current planning so two topic areas addressed this. One relates to ensuring support is in place for the uptake of services, coordinating that support between the different levels of provision with EOSC and in research institutions. The other covers training and the need to continue to build capacity and professionalise career paths to ensure the necessary support is available to researchers.

Other priority areas focus on advancing the services by better connecting between research facilities, HPC and home institutions, incentivising transnational access to cloud and high-throughput data processing facilities and deploying Scientific Knowledge Graphs and researcher profiles to better monitor and track research activities to ensure attribution and credit. The need for a network of trusted digital repositories to ensure quality and trust in the data is also prioritised, alongside support for long-term preservation.

An updated version of this deliverable is due in month 28 (August 2023) at which point we will reflect on how the recommendations were adopted and what priorities continue to exist.

We wish to acknowledge all the consortium members who helped to shape the recommendations put forward for consideration.