



Brussels, 14.3.2018
SWD(2018) 83 final

COMMISSION STAFF WORKING DOCUMENT

Implementation Roadmap for the European Open Science Cloud

Summary

This Staff Working Document (SWD) presents the outcome of the exploration of appropriate governance and financing mechanisms for the European Open Science Cloud (EOSC) in the form of a possible implementation Roadmap, as foreseen by the Communication on the 'European Cloud Initiative'¹ (henceforth 'the Communication').² The document also describes the measures taken under Horizon 2020 Work Programmes to start implementing the EOSC.

The implementation Roadmap draws upon the outcome of an extensive and conclusive consultation with scientific and institutional stakeholders in 2016 and 2017 and builds concretely on the Horizon 2020 Work Programme 2018-2020.³ The consultation confirmed and upheld the intervention logic presented in the Communication, to create a fit for purpose pan-European federation of research data infrastructures, with a view to moving from the current fragmentation to a situation where data is easy to store, find, share and re-use. On this ground, this SWD sets out a comprehensive overview of the implementation of the EOSC, with possible action lines and timelines resulting from the consultation.

The document serves as a basis for further consultation with Member States, the European Parliament and other relevant stakeholders on the next steps to take. It will also help stakeholders to orient their future contributions to the initiative.

¹ COM(2016)178 final.

² http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=15266

³ European Commission Decision C(2017)7124 of 27 October 2017.

Acronyms and shorthand

<i>Associated Countries</i>	Countries associated to the Horizon 2020 Framework Programme for Research and Innovation
<i>EOSC</i>	European Open Science Cloud
<i>ERAC</i>	The European Research Area and Innovation Committee.
<i>ESFRI</i>	European Strategy Forum for Research Infrastructures
<i>FAIR data</i>	Data that is Findable, accessible, interoperable and re-usable
<i>Horizon 2020</i>	The EU Framework Programme for Research and Innovation Horizon 2020
<i>INFRAEOSC</i>	Call dedicated the implementations of the EOSC in the 2018-2020 Work Programme of Horizon 2020
<i>WP</i>	Work Programme of Horizon 2020

Introduction

The Commission presented its vision for the European Open Science Cloud (EOSC) in its April 2016 Communication on the 'European Cloud Initiative',⁴ as a part of the Digital Single Market Strategy. The objective of the EOSC is to give the Union a global lead in research data management and ensure that European scientists reap the full benefits of data-driven science, by offering '1.7 million European researchers and 70 million professionals in science and technology a virtual environment with free at the point of use, open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines'. The European Cloud Initiative also foresees setting up a European Data Infrastructure, underpinning high-capacity cloud solutions with super-computing capacity,⁵ as well as widening the EOSC by gradually opening up its user base to the public sector and industry.

As announced in its Communication on "Building the European Data Economy"⁶, the objective of the Commission is to create a policy framework that enables data to be used throughout the value chain for scientific, societal and industrial purposes. As such, the EOSC will be a fundamental enabler of Open Science and of the digital transformation of science, offering every European researcher the possibility to access and reuse all publically funded research data in Europe, across disciplines and borders. Consultations with stakeholders and Member States reinforce the view that EOSC has the potential to leverage past investment in research data infrastructures to add value in terms of scale, interdisciplinarity, and faster innovation.

The EOSC has emerged as a clear policy priority for European research and innovation. It has been strongly supported by the European scientific community in the *EOSC Summit and Declaration*, by the Council in Council Conclusions (May 2015 and May 2016)⁷ and by the European Parliament in a Resolution (January 2017).⁸ It also received favourable opinions from the Economic and Social Committee⁹ (September 2016) and from the Committee of the Regions (October 2016).¹⁰

The Commission is providing and planning the necessary financial support to implement the EOSC along the policy orientations of the Communication by means of projects under the EU Framework Programme for Research and Innovation (Horizon 2020).¹¹ Actions were included in the Work Programme (WP) 2016-2017,¹² and in the WP 2018-2020,¹³ for an aggregate budget of about €600m. This includes the launch in 2018 of the INFRAEOSC dedicated Call, which will support notably the integration of services and the federation mechanism; the setting/operationalization of the principles of FAIR data (findable, accessible, interoperable and reusable); the development of a FAIR-compliant certification scheme for data infrastructure and the connectivity of the pan-European Research Infrastructures such as the ESFRI projects and landmarks. Moreover, the Commission launched the EuroHPC Joint Undertaking, which will

⁴ COM(2016)178 final

⁵ This document does not report in detail on the implementation of other pillars of the Communication, namely *European Data Infrastructure (EDI)* and *Widening access and building trust*. The Communication foresees to link the three pillars at a second stage, after the set up of clear and stable governance arrangements for the EOSC and EDI.

⁶ COM(2017) 9 final

⁷ <http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>

⁸ <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0052+0+DOC+XML+V0//EN>

⁹ TEN/592-EESC-2016 <http://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/european-cloud-initiative-building-competitive-data-and-knowledge-economy-europe>

¹⁰ CDR 2880/2016 <http://cor.europa.eu/en/activities/opinions/pages/opinion-factsheet.aspx?OpinionNumber=CDR%202880/2016>

¹¹ <https://ec.europa.eu/programmes/horizon2020/>

¹² European Commission Decision C(2017)2468 of 24 April 2017.

¹³ European Commission Decision C(2017)7124 of 27 October 2017.

provide a world-class High-Performance Computing infrastructure to underpin data computation in the EOSC.¹⁴

In parallel, as foreseen by the Communication, the Commission services initiated in 2016 an extensive process of engagement and consultation of stakeholders and Member States, to arrive at a coherent *implementation Roadmap* for the initiative. In the Mid-Term Review of the implementation of the Digital Single Market Strategy,¹⁵ the Commission confirmed its intention to come forward with an implementation Roadmap for the European Open Science Cloud, based on these ongoing consultations.

Research stakeholders provided key inputs for the implementation of the EOSC at the *EOSC Summit* (June 2017). Later they demonstrated their concrete support to the initiative via wide endorsement of the *EOSC Declaration* (October 2017),¹⁶ which sets out the guiding principles for implementation of the EOSC. The consultation showed clearly that the EOSC must be federated - understood both as agreement on standards of operation and on the need for a collective governance - and inclusive. The views of stakeholders and the evidence from the existing practice of similar initiatives (i.e. existing multi-stakeholder infrastructures such as e.g. ELIXIR, ICANN and ITF) point to the need for a stable, inclusive and effective governance mechanism for the implementation of the EOSC.

This SWD summarises the inputs received from the consultation with stakeholders and Member States (Section 1); then, it presents for discussion the steps towards an implementation Roadmap based on a possible EOSC model and governance framework, together with relevant actions currently underway and their respective timelines (Section 2).

The SWD does not prejudice any decisions to be taken by the Commission.

Section 1 – Stakeholder input

The Commission services received inputs both from the Union institutions and from the scientific community. They engaged with Member States and Countries associated to Horizon 2020 regarding options for implementing the EOSC initiative (see Annex 1 for full details of this process). The Commission services also procured studies and reviewed independent advice on the feasibility and appropriateness of different setups for the governance and financing of the EOSC. The main outcomes of these processes are presented here.

1.1 The response of Union institutions

The European Union institutions responded favourably to the vision proposed by the Commission. Both the European Parliament and the Council clearly flagged the need for further specification and for an overview of the activities foreseen by the Communication. The European Parliament *Resolution on the European Cloud Initiative* (16 February 2017), Council conclusions on *The transition towards an Open Science system* (27 May 2016), the Opinion of the Committee of the Regions (CoR) and the Opinion of the Economic and Social Committee (EESC), all welcome and support the initiative and ask the Commission to act quickly and resolutely with an integral, coherent implementation plan.

In summary (details are provided in Annex 1), the Union institutions called on the Commission, supported by an inclusive process involving all relevant stakeholders, to:

- present an implementation roadmap with clear timelines, actions and budget, including the resources available in through Horizon 2020 and within its proposal for FP9;

¹⁴ Proposal for Council Regulation establishing the European High-Performance Computing Joint Undertaking ('EuroHPC'), COM(2018)8 final, http://europa.eu/rapid/press-release_IP-18-64_en.htm

¹⁵ COM(2017)228 final, <https://ec.europa.eu/transparency/regdoc/rep/1/2017/EN/COM-2017-228-F1-EN-MAIN-PART-1.PDF>

¹⁶ https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf

- explore an appropriate governance structure, based on existing initiatives and their sustainability;
- define an architecture that ensures information security and personal data protection.

In particular, the Council agrees that Europe may benefit from a EOSC that federates existing research data infrastructures, today scattered across disciplines and Member States, enabling, safe and long-term storage, efficient analysis, and re-use of research data across borders and disciplines; it endorsed several of the principles underlying the implementation of the EOSC (FAIR data, data stewardship, and cost coverage, see Section 2); and called on the Commission, in cooperation with Member States and stakeholders, to explore appropriate governance and funding frameworks, based on existing initiatives and their sustainability.

The European Parliament called on the Commission to present an *action plan* with clear working packages and timelines, defining the results to be achieved, the sources of financing and the stakeholders involved throughout the process; to reduce the fragmentation of digital infrastructures by establishing a roadmap for actions and a robust governance structure; and to identify appropriate financing mechanisms for the EOSC and to provide sufficient resources for this policy through Horizon 2020 and within its proposal for FP9.

In addition to these calls, the Committee of the Regions and the European Economic and Social Committee stressed the need for an EOSC implementation plan that is inclusive of all relevant stakeholders, including civil society and the industry, to ensure information security and personal data protection, and for the EOSC to be appropriately financed by Member States.

Acting on this positive response, the Commission services further engaged with Member States to prepare the implementation of the EOSC, notably to explore the governance and financing aspects, including a workshop with Member States representatives (29 June 2016)¹⁷. The outcomes converged with the institutional responses described above; they also underlined the need to differentiate clearly between the architecture of the EOSC (e.g. the implementing model), and its governance (e.g. how it is run). Specifically, they suggested that the EOSC should act as a trusted platform between research data needs, traditionally funded by public money, and cloud-based scientific services in the private marketplace. The outcomes of the workshop also underlined that the governance of the EOSC should help prevent duplication of efforts, fragmentation and isolated solutions. Moreover, it was clear that governance should streamline decision-making through a 'one-stop shop' in Europe for decisions related to research data sharing. The workshop highlighted the need to re-focus Horizon 2020 and the next Framework Programme by extending existing research data infrastructures beyond the lifecycle of single projects, and to integrate and federate them in more sustainable facilities. For the same purpose, a combination of European Structural and Investment Funds for the investment in new infrastructures, equipment and connectivity,¹⁸ competitive funds and user fees could support the long-term sustainability of the EOSC.¹⁹ The Commission services further engaged with Member States in discussions, the latest taking place in December 2017, seeking further their views on a possible implementation Roadmap (see Annex 1 for details).

The need for better coordination and a one-stop-shop approach for governance is confirmed by Member States and Associated Countries official documents regarding research data and infrastructures. These include:

¹⁷ Representatives from SK, DE, NL, FR, AT, DK, SE, IT, ES, EE, BE, RO, UK, TR, NO and CH attended; <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud-workshop>

¹⁸ In the current ERDF programmes, over EUR 875 million were earmarked to ICT infrastructure like large-scale computer resources or equipment and over EUR 1.5 billion to a very high-speed broadband connectivity.

¹⁹ Other technical and policy-related topics were discussed at the workshop, including FAIR Data, interoperability, data management plans, connectivity, standards, interdisciplinarity, cloud-based services, joint procurement, rules of engagement, various options for a 'legal vehicle' to implement the roadmap and channel funds, human resources and skills, the need for core data experts, and international cooperation.

- the ERAC Opinion on Open Research Data,²⁰ focusing on the need for open research data infrastructures that are FAIR and financially sustainable;
- the ERA National Action Plans, flagging the great variety of national e-infrastructures and the fragmentation of access to results of publicly funded research and storage of science-related digital content at national level;²¹
- the 2017 reporting of the National Points of Reference for the Recommendation on access to and preservation of scientific information (C(2012)4890 final),²² providing evidence of technical and policy work by Member States on national e-infrastructures.

The Commission services also liaised with the Research Infrastructure Programme Committee over the course of eighteen months (see Section 2).²³ The Commission included initial implementing provisions for the EOSC in the Commission Decision for WP 2016-2017 of Horizon 2020. In WP 2018-2020, the Call INFRAEOSC was discussed and approved by the Programme Committee. It provides about €60m to support implementation and governance of the EOSC (see Section 2). The aforementioned Committee called for establishing a clear link of the activities of the support project INFRAEOSC-05a-2018²⁴ to the implementation Roadmap for the EOSC (Commission Decision C(2017)7124 of 27 October 2017).

1.3 Stakeholder engagement

The Commission services engaged stakeholders in a sustained dialogue on the needs of user communities, on the ideal functional specifications for the EOSC and on the optimal model of implementation, as foreseen by the Communication.

A Commission High Level Expert Group on the EOSC was created²⁵ to provide advice to the Commission on a possible strategy for the implementation of the EOSC initiative. The group convened three large-scale stakeholder meetings and participated in more than thirty global events relevant to the EOSC. The final Report, published in October 2016,²⁶ recommended closing the discussions about the *perceived need* of an open science cloud and taking immediate action in close concert with Member States, building on existing capacity and expertise. It recommended setting clear *Rules of Engagement* for the access to the EOSC and for the provision of services based on research data (e.g. TDM, data analytics). It recommended framing the EOSC as the EU contribution to a future global *Internet of FAIR Data and Services* underpinned by open protocols. Finally, it exhorted the Commission and Member States to set a wide framework for the implementation of the EOSC initiative that extends beyond individual projects well into the global scientific community, to address fundamental issues for the future of EU science.

On 12 June 2017, the Commission services organised the *EOSC Summit*²⁷ to ascertain the commitment of stakeholders to the implementation of the EOSC (see Annex 1). The Summit provided strong support for implementation of the key components of the initiative and resulted in the *EOSC Declaration*.²⁸ The Declaration is composed of 33 high level statements that capture

²⁰ ERAC Opinion on Open Research Data, 1202/16 of 3 February 2016 <https://era.gv.at/object/document/2402>

²¹ Commission Staff Working Document Accompanying the document Report from the Commission to the Council and the European Parliament *The European Research Area: time for implementation and monitoring progress* SWD(2017)21 final <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52017SC0021>

²² The National Points of Reference were established as a Commission Expert Group (no. E03477) by the 2012 Recommendation to report on the advancement of Member States every 18 months. The Commission plans to published the reporting of 2017 in the spring 2018.

²³ Relevant meetings were held in 2017 on 17/03 (first draft of the INFRAEOSC call); 12/05 (presentation of the revised WP following comments); 27/06 (new version of the INFRAEOSC call and debriefing from the EOSC Summit; presentation of new topics on governance and coordination with national and thematic initiatives); and 22/09 (final discussion on the INFRAEOSC call).

²⁴ <http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/infraeosc-05-2018-2019.html>

²⁵ <http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3353>

²⁶ http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf. The Report was downloaded more than 2000 times shortly after publication and it was featured in a number of scientific publications.

²⁷ <http://ec.europa.eu/research/index.cfm?eventcode=44D86060-FBA1-1BD1-9355822B162BB0EE&pg=events>

²⁸ <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

stakeholders' shared understanding of the actions needed on Data culture and FAIR data, on research data services and architecture, on governance and funding, to make the EOSC a reality by 2020. About 70 major scientific stakeholders undersigned the EOSC Declaration following the Summit, out of which 60 committed to undertake specific actions to make the EOSC happen. Furthermore, several key European e-Infrastructures signed position papers²⁹ and a joint e-Infrastructures statement on the EOSC³⁰ that expressed strong support for the principles of EOSC implementation expressed in the Declaration.

Finally, the Commission services examined the relation between the initiative and long-standing reference European Research Infrastructures, such as the European Strategy Forum for Research Infrastructures (ESFRI) infrastructures.³¹ Indeed, the Communication included *Connecting Research Infrastructures to the EOSC* as one of the main actions required, and the legislators mentioned Research Infrastructures explicitly as assets to be federated into the EOSC. The *Consultation on the Long Term Sustainability of the ESFRI Infrastructures* (May 2016)³² identified the better exploitation of data generated by the Research Infrastructures as one of the pre-conditions for long-term sustainability.³³ Respondents identified interoperability, common services, policies and open data obligations as top requirements to improve data management, better exploiting data and facilitating reuse of research data (see Annex 1.4). As a result, the Commission services identified *exploiting the data produced by European Research Infrastructures in a more strategic way* as a key challenge and potential action of a future European action plan on long term sustainability of European Research Infrastructures (SWD(2017) 323 final).³⁴ Specifically, the SWD encourages Research Infrastructures to promote the re-use of their research data for innovation and education purposes by supporting the connectivity of Research Infrastructures to the EOSC.

1.4 Independent advice based on existing evidence and good practices

To complement the inputs from EU institutions and from the scientific community, the Commission services gathered specific independent advice and analysed relevant documents on the appropriate shape of a future governance structure for the initiative.³⁵ This included an analysis of current governance practices of major functioning, large-scale scientific networks including ICANN, IETF, AIOTI, GÉANT and ELIXIR (for an overview of the results see Annex 1.4). The synthesis of the results in this section is based on:

- A study on the appropriate governance of the EOSC;³⁶
- A report on EOSC governance by Science|Business;³⁷

²⁹ https://www.geant.org/News_and_Events/CONNECT/Pages/european_open_science_cloud.aspx
<https://documents.egi.eu/public/RetrieveFile?docid=2417&filename=EGI-Strategy-2015-2020.pdf>
<https://www.eudat.eu/news/the-european-open-science-cloud-for-research>

³⁰ https://zenodo.org/record/32915#.Wjz38_6ouUk

³¹ Research Infrastructures are key actors in open access as they generate, collect and handle significant volumes of data that is used by thousands of researchers and innovators across scientific disciplines.
https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

³² The consultation collected about 200 answers from stakeholders on the pre-conditions of long-term sustainability of Research Infrastructures and on the potential actions/measures to tackle the challenges posed by their implementation. The communities targeted were European Research Area stakeholders, ESFRI projects, European Research Infrastructure Consortia, ESFRI delegations, members of the Programme Committee for the Research Infrastructures part of Horizon 2020, e-Infrastructure Reflection Group delegations, EIROforum members, International Organisations, Research Infrastructure associations, National Contact Points and science attachés from strategic third country partners.

³³ https://ec.europa.eu/research/infrastructures/pdf/lts_report_062016_final.pdf.

³⁴ https://ec.europa.eu/research/infrastructures/pdf/swd-infrastructures_323-2017.pdf

³⁵ Member States and national research funders also produced reports on the governance and funding of national research data infrastructures, including Germany (<http://www.rfii.de/en/category/documents/>) and the Netherlands (<https://www.nwo.nl/en/news-and-events/news/2017/ew/nwo-advocates-permanent-funding-for-national-digital-infrastructure.html>). Reports from France and from the UK are forthcoming in 2018. While these reports were considered as academic inputs for the implementation Roadmap, this editorial process is not reported here.

³⁶ <https://ec.europa.eu/research/openscience/index.cfm?pg=publications>

- A report on the governance aspects of the EOSC by the Open Science Policy Platform;³⁸
- Two OECD reports on coordination and financing of international research data infrastructures;³⁹
- A deliverable of the EOSCpilot on a Draft Governance Framework for the EOSC;⁴⁰
- Review of the governance model proposed for the GO-FAIR initiative.⁴¹

All these sources converge that EOSC requires:

- Strong policy guidance in the initial stage, e.g. a **clear governance framework** for the initiative that makes it predictable; a multi-level and multi-stakeholder governance with clear institutional, executive and advisory roles that empowers the scientific community and evolves with time; the need for long-term public funding for the services needed to enable the integration of and access to the data resources to be federated in the EOSC;
- The definition of **the initial services** that are needed to gather and organise FAIR research data and data-related research products, to be available via a service platform;
- A **clear business model** for research data repositories and networks that mixes sources of revenue for long-term sustainability;
- The facilitation of **access and re-use**;
- **Cost optimisation** (e.g. reduction of duplication, etc.) to be sought via synergies.

It emerged clearly from the consultation of scientific stakeholders and Member States and Associated Countries that the EOSC would need to be both scalable and flexible, adaptable to the emerging needs of the scientific community and able to support the whole research data lifecycle. The implementation roadmap should be iterative and take account the impacts on innovation;⁴² so that the EOSC could respond to changing needs of scientists regarding research data and to strategic EU and national decisions on research data.

Section 2 –EOSC implementation Roadmap: the next steps

The Communication provided that Horizon 2020 would be used to integrate and consolidate e-infrastructure platforms, to federate existing research infrastructures and scientific clouds and to support the development of cloud-based services for Open Science. On this basis, the Commission is supporting the implementation of the EOSC as a federated model that combines effectiveness and flexibility, primarily through the Research Infrastructures (including e-Infrastructures) part of Horizon 2020 Work Programme. Horizon 2020 funding constitutes the core of the actions around which the EOSC implementation Roadmap unfolds (see Annexes 2 and 3).

On the other hand, engagement with stakeholders and Member States helped formulate a set of consensual statements at the *EOSC Summit* in June 2017; these support and further define the concrete elements of the *federated* model of implementation of the EOSC foreseen in the Communication and sketched out in Horizon 2020 Work Programme 2018-2020. Overall, results from the consultation strongly supported the intervention logic proposed by the Commission, as it presents numerous advantages over the current fragmented model of scientific data computing.

³⁷ ; <https://sciencebusiness.net/report/governing-european-open-science-cloud/>; Science|Business is an independent consultation group representing research, industry and policy

³⁸ <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform>

³⁹ Under the aegis of the Global Science Forum, the Commission services participated to two OECD working groups on Open Science, respectively on Co-ordination and Support of International Research Data Networks and on Business Models for Sustainable Research Data Repositories. The Commission hosted the final validation workshops on 28-31 March 2017. The groups provided recommendations to OECD member institutions regarding the challenges and enablers for the effective functioning of international research data networks, and the income streams, costs, value propositions, and business models for research data repositories. <https://www.innovationpolicyplatform.org/open-data-science-oecd-project> and <http://www.codata.org/working-groups/oecd-gsf-sustainable-business-models>

⁴⁰ <https://eoscpilot.eu/content/d22-draft-governance-framework-european-open-science-cloud>

⁴¹ GO-FAIR was first presented at the COMPET Council of 30 May 2017. See <https://www.go-fair.org/news/>

⁴² As mentioned in the European Parliament Report and in the *EOSC Declaration*.

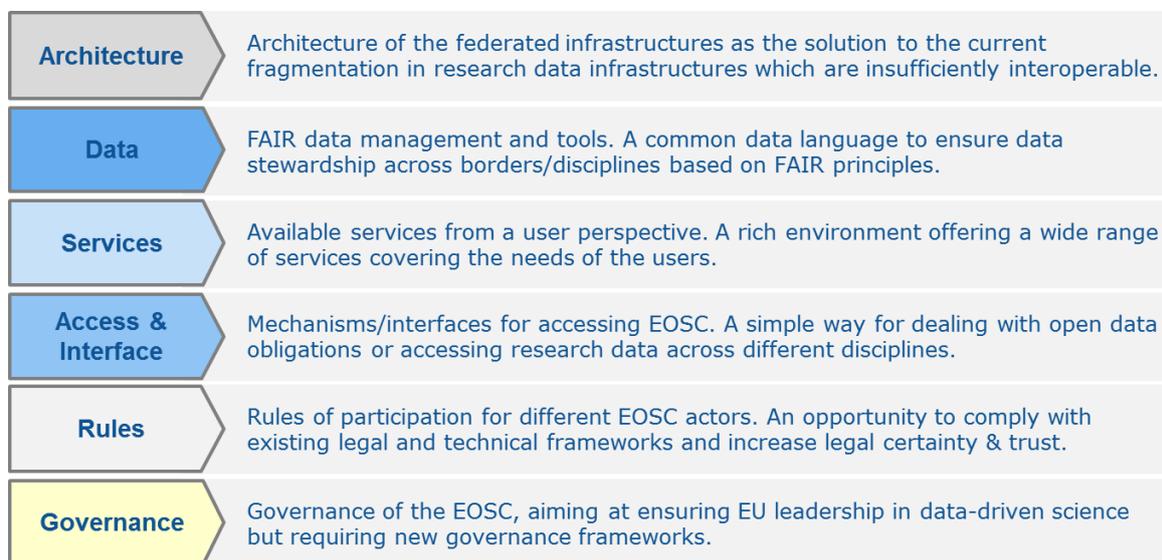
The EOSC would build in stages, progressively moving from the past towards a federation of infrastructures at the European level. The consultation clearly and conclusively ruled out a centralised model of implementation as a valid option for the implementation of the EOSC.

This section presents the main steps towards an implementation Roadmap based on a possible future EOSC model constituted of six action lines. Each action line makes specific reference to the resources committed in Horizon 2020 and to the relevant action timelines, if any. Finally, the section flags the link between the EOSC and the European Data infrastructure initiative (EuroHPC) and provides a preliminary reflection on the widening access and costs and financing of the EOSC implementation.

2.1 A possible EOSC Model

The model describes a pan-European federation of data infrastructures built around a federating core and providing access to a wide range of publicly funded services supplied at national, regional and institutional levels, and to complementary commercial services. The model includes six actions lines: (a) architecture, (b) data, (c) services, (d) access and interfaces, (e) rules and (f) governance (Figure 1, below).

Figure 1 – EOSC Model action lines



(a) Architecture

Based on the consultation, the 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. In terms of architecture, the EOSC would essentially comprise a federating core and a variety of federated research data infrastructures committed to providing services as part of the EOSC. The groundwork for such a federated EOSC architecture was laid by several projects funded under Horizon Work Programme 2016-2017, which aim at federating data infrastructures at the European level and offering shared services (e.g. EGI, EUDAT, ELIXIR, EPOS etc.).⁴³ In addition, resources were committed to examine the EOSC architecture through the EOSCPilot project.

⁴³ The Commission funded the integration, interoperability and federation of data infrastructures in various fields and the development of horizontal data services. These actions delivered generic and thematic data services, workflows, interoperable standards and ontologies, which pave the way toward the establishment of a European integrated environment for research data. In particular, Horizon 2020 supported the development and interoperability of pan-European thematic data infrastructures, through targeted support to the implementation and operation of the ESFRI roadmap projects identified through the ESFRI prioritisation exercise. Among these, two priority ESFRI projects,

The EOSC federating core is understood to be constituted by EOSC shared resources and by a compliance framework including notably the Rules of Participation (see section 2.1.e). The Work Programme foresees developing the initial shared resources around the EOSC-hub project, the EOSC Portal and a catalogue of data infrastructures and services. Therefore, the process of federation entails two inter-related activities:

1. To develop shared resources as part of the *federating core*. In the initial phase, Horizon 2020 projects, notably the EOSC-hub, will provide an access channel complementing the access mechanisms in use at different data infrastructures. A portfolio of projects described later in this SWD (section 2.1.c) will provide horizontal services such as a portal, authentication and authorisation and security services, allowing users to access the computing, data and services of pan-European and disciplinary research data infrastructures, which already federate data infrastructures at the European level. A catalogue of EOSC services, including both thematic and generic services – for data storage, management and analytics, simulation and visualisation, distributed computing, etc. will help researchers to discover, select and use the services they need.

2. To connect to the core *a large number of research data infrastructures* (henceforth *data infrastructures*).⁴⁴ The *hub* would relay the resources and the services of data infrastructures funded at EU, national and regional level. Service and resources might be both generic and thematic-specific. *The progressive federation* over time of existing service providers in the EOSC would provide a single, coherent access channel to EOSC services at European level that meets researchers' needs for data sharing, management and computing.

This process of federation of resources would be implemented gradually, based on simple guidelines consistent with existing good practices:

- data infrastructures would enter the federation on a voluntary basis based on the commitment of resources and on the capacity to comply with its rules; minimum commitments would be set in the Rules of Participation to ensure fairness;
- data infrastructures would define the extent of their own involvement in the federation, in terms of the data sets and services they would contribute to the EOSC; their commitment and rule compliance would be limited to these data sets and services;
- data infrastructures would continue to follow their own rules outside of their specific commitments to the EOSC;
- data infrastructures would operate in the EOSC according to FAIR data principles and seek to become FAIR-accredited/certified entities, meaning that their data services would meet over time certain infrastructural and quality standards under a quality-assurance scheme;
- the structuring of the EOSC federation would occur flexibly, in response to actual needs and requests; i.e. data infrastructures that already have the capacity, commitment and added value to facilitate/coordinate EOSC operations at a geographical or thematic level could seek to become *EOSC federated centres*;
- the federation would entail as few constraints as needed to deliver the expected EOSC services;
- the federating process would aim to achieve economies of scale and scope.

EPOS (European Plate Observing System) and ELIXIR (The European Life-Science Infrastructure for Biological Information), received significant funding in the 2014-2015 WP.

⁴⁴ This terminology is widely understood; it has been discussed for the preparation of the OECD reports mentioned in Section 1. Specifically, *research data infrastructures* refers both to international research data networks and to research data repositories.

Building on past investments, the *EOSC federation* would allow for the further development of shared EOSC resources at the European level,⁴⁵ serving general and cross-disciplinary user needs, and supporting the federation of other research data infrastructures wishing to join. The EOSC would also enable researchers to access services provided by commercial operators that comply with minimum set standards (e.g. see further below, the services provided by Copernicus’s DIAS).

Overall, all federated research data infrastructures and commercial service providers would comply with the Rules of Participation. Importantly, they would agree to implement the FAIR data principles fully. This would contribute to their long-term sustainability, as EU research funders are gradually tying their funding to open access obligations and the use of FAIR-accredited/certified repositories. Over time, as foreseen in the Communication, the EOSC would reach out to resources from EU’s global research partners, to create a level playing field in open data for global scientists.⁴⁶

The timeline below shows how resources of Horizon 2020 would serve this particular action line.

Starting from	Committed resources (non-exhaustive)	Action	Milestones
2018, Q1	<ul style="list-style-type: none"> • EOSC -hub project • OpenAIRE-Advance project • FREYA project 	Develop initial EOSC federating core including the EOSC shared resources	Q4 2019: Initial EOSC federating core in place
2019, Q1	<ul style="list-style-type: none"> • EOSCpilot project INFRASUPP-01-2018-2019 (b3)	Develop catalogue of interested and eligible (per Rules of Participation) data infrastructures to be federated into the EOSC and identify EOSC federate centres	Q4 2019: Registry of data infrastructures of the EOSC (initial)
2018, Q4	<ul style="list-style-type: none"> • INFRAEOSC-04-2018 	Connection the research infrastructures identified in the ESFRI Roadmap to the EOSC. Support to this activity will be provided through cluster projects.	Q2 2020: Preliminary connection of most infrastructures and services to the EOSC

In addition to directly supporting the federation of ESFRI projects in the EOSC (INFRAEOSC-04-2018), WP 2018-2020 of Horizon 2020 funds specific actions in scientific areas with a tradition of research data sharing and services like transport, food, marine, health and earth-observation; this ensures that the EOSC is fully inclusive.

The WP 2018-2020 on food security, sustainable agriculture and forestry, marine, maritime and inland water research and the bioeconomy includes two topics, one each for developing and building cloud services on food data and ocean data, in such a way that they can be eventually federated into the EOSC.⁴⁷ In health, a significant development included in WP 2018-2020 is the Health Research and Innovation Cloud (HRIC), which aims to structure first and later establish a thematic cloud for health-related research, in strict relation with the EOSC (details in Annex 4).

Secondly, the Commission invests heavily in data regarding the planet and the environment in the Copernicus programme, the flagship space programme. Copernicus’s *Data and Information Access Services* (DIAS) provide access, tools and processing capabilities for scientists and innovators to exploit this data. DIAS are operated by the industry and will offer additional services in the EOSC under commercial conditions. Federating Copernicus data and DIAS added-value services into the EOSC will leverage the existing Commission investments for the

⁴⁵ This has been previously defined conceptually as an *e-infrastructure commons*.

⁴⁶ This aim was supported by the Report of the HLEG EOSC and upheld by stakeholders in the *EOSC Declaration*. Here, they call for the EOSC to reach out over time to relevant global research partners. This ‘will increase the global value of open research data [...] It will gradually widen the initiative to federated network of infrastructures and nodes from global research partners. The EOSC Stakeholder Forum will have an important role in this sense’.

⁴⁷ The two calls are: DT-SFS-26-2019: Food Cloud demonstrators and BG-07-2019-2020: The Future of Seas and Oceans Flagship Initiative, <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/food-security-sustainable-agriculture-and-forestry-marine-maritime-and-inland-water>

benefit of multiple science and innovation communities. In line with the intervention logic of the Communication, this will reduce the burden for 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.

International dimension

Finally, the Commission services have pursued the line of action foreseen in the Communication on the global aspects of the EOSC, to ensure that the EOSC contributes to a global playing field regarding open, FAIR research data. The Governance Development Forum of the EOSCPilot project has provided since 2017 a bridge for the contribution of global research institutions from third countries to the EOSC.⁴⁸ Commission services worked in the context of G7, G20 and in collaboration with the OECD Global Science Forum⁴⁹ to ensure that the implementation of the EOSC proceed in line with Commission priorities in the international and development agenda, in the context of the Sustainable Development Goals. Attaining these goals requires global knowledge and FAIR data sharing, including in particular in the areas of science and research. G7 leaders and the Commission called for ‘all researchers [to] be able to deposit, access and analyse scientific data across disciplines and at the global scale, and research data should adhere to the FAIR principles of being findable, accessible, interoperable, and reusable’.⁵⁰ To this purpose, the Call NFRAEOSC-05-2018-2019(c) includes specific support to link FAIR data globally.

(b) Data

It emerged clearly from the consultation - notably from the *EOSC Summit* and from the *EOSC Declaration* - that further measures would be needed to foster the development of professional practices of research data management and stewardship in Europe, specifically:

1. to develop a better culture of research data management and practical skills among EU scientists and innovators, including action on incentives, rewards, skills and curricula related to research data and data science;⁵¹
2. to develop FAIR data tools, specifications, catalogues and standards, and supply-side services to support scientists and innovators, and
3. to stimulate the demand for FAIR data through consistent FAIR data mandates and incentives to open data by research funders and institutions across Europe.

In short, this combined course of actions would provide a set of EOSC shared resources for data management, which can be used by data-savvy researchers, implemented by all data infrastructures, mandated by all research funders and supported locally by host institutions.

The development of FAIR data management has already begun in Horizon 2020. Regarding skills, the Commission funded in WP 2014-15 the EDISON project *Data Science Framework*, which helps define curricula and training standards to address the increasing demand for the data professionals required for e-infrastructures and research.⁵² The Commission funded training and resources to enhance skills in Open Science via FOSTER and FOSTERplus, in the WP of *Science*

⁴⁸ <https://eoscpilot.eu/about/governance-framework>

⁴⁹ See details in footnote 39.

⁵⁰ G7 Science Ministers Communiqué, 28 September 2017, Turin. Articles 19 and 20. <http://www.g7italy.it/sites/default/files/documents/G7%20Science%20Communiqu%C3%A9.pdf>; also see the political support for FAIR expressed at the G20 Leaders’ Communiqué Hangzhou Summit http://europa.eu/rapid/press-release.STATEMENT-16-2967_en.htm

⁵¹ For example, the Final report of the High Level Expert Group on the EOSC estimates that half a million core data scientists are needed to make the most of open research data in Europe.

⁵² Horizon 2020 - INFRASUPP-4-2015 - New professions and skills for e-infrastructures.

with and for Society in FP7 and Horizon 2020.⁵³ The WP 2016-2017 then provided resources to address the shortage of data-related skills by training, support for skills development for e-infrastructure providers and users, and support to FAIR data management (e.g. OpenAIRE, RDA Europe, FREYA). Importantly, Marie Skłodowska-Curie actions have been increasingly encouraging the incorporation of open science skills and knowledge training into the 'Innovative Training Networks' action (MSCA-ITN), both for early-stage researchers, and in the 'Co-funding of regional, national and international programmes' action (MSCA-COFUND), for excellent doctoral and postdoctoral R&I programmes across Europe. Such systemic activities foster a culture of data sharing and FAIR data management and help develop the human resource capital required for the EOSC to function.

Starting from	Committed resources (non-exhaustive)	Action	Milestones
2018, Q1	FAIR data Expert Group (E03464), in consultation with stakeholders	Prepare a FAIR data Action Plan	Q3 2018: FAIR data Action Plan published
2018, Q3	<ul style="list-style-type: none"> • DG RTD • RDA Europe 4.0 • INFRAEOSC-05-2018-2019 (b) 	Define a European framework for FAIR research data	Q2 2019: European framework for FAIR research data agreed
2019, Q1	<ul style="list-style-type: none"> • FREYA project • RDA Europe 4.0 	Define a Persistent Unique Identifier policy for FAIR data	Q4 2019: FAIR persistent unique identifier policy defined
2019, Q1	INFRAEOSC-05-2018-2019 (c)	Develop a FAIR data accreditation /certification scheme for repositories	Q4 2019: FAIR certification scheme available.

EOSC shared resources would be developed to cover all the aspects of FAIR data:

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

The consultation and evidence suggest that the process for developing shared resources should be staged, iterative and flexible. It should start by taking stock of the tools and practices in place in different scientific disciplines, before agreeing, cataloguing, certifying and finally implementing them as part of the EOSC shared resources and EOSC rules. Research stakeholders would increasingly rely on standards and recommendations developed by their respective communities under well-established initiatives such as W3C and RDA to ensure service interoperability.

These activities are included specifically in the work description of EOSC-hub, OpenAIRE-Advance, RDA Europe and of several other EOSC-related projects. EOSC-hub will establish competence centres by delivering specialised trainings and co-create technical solutions with users to improve skills and knowledge among researchers and service operators. OpenAIRE-Advance will develop standardised, recognised and accredited Open Science skills training for researchers, data practitioners and citizen scientists. Both projects are expected to put in place a coordinated strategy. Through the implementation Roadmap, such activities to provide shared resources for data management in the context of the EOSC would become more systematic.

Importantly, the FAIR data Action Plan foreseen by the Communication is meant to set out the actions needed to develop EOSC shared resources and define the operational guidance and methodologies for applying the FAIR principles with these shared resources, including through

⁵³ <https://www.fosteropenscience.eu/about> was funded from the Horizon 2020 Call SwafS-07-2016: Training on Open Science in the European Research Area and earlier from FP7 Science in Society Work Programme 2013-1.

FAIR maturity models and FAIR accreditation/certification schemes. The outcome of the action plan would eventually constitute a new operational framework for FAIR research data.⁵⁴

The policy-related work foreseen by the Communication to support the demand for FAIR data has also started. This includes the introduction of the obligation of FAIR data management in Horizon 2020 grant agreements and funding provided to cover the costs of access to data in various research areas of Horizon 2020. Funding for data mandates and incentives to make data open are being addressed at Member State level via the revision the Recommendation on access to and preservation of scientific information,⁵⁵ as foreseen directly in the Communication (action four on the European Science Cloud). The activities of the GO-FAIR initiative, launched jointly by the Dutch, German and French governments also support mainstreaming of FAIR data management. Finally, mainstreaming would build on the *EOSC Declaration*, specifically on the several concrete commitments made by major scientific stakeholders to change their institutional practices (e.g. careers, rewards and incentives) towards open, FAIR data.

(c) Services

The consultation and the available evidence show that EOSC might offer five main types of services for European researchers. While such services are currently being provided to specific scientific communities, they are limited by the contexts of disciplines, by national boundaries or by both. The EOSC would make them all available irrespective of discipline or national boundaries.

These services are:

1. A unique identification and authentication service and an access point and routing system towards the resources of the EOSC.
2. A protected and personalised work environment/space (e.g. logbook, settings, compliance record and pending issues).
3. Access to relevant service information (status of the EOSC, list of federated data infrastructures, policy-related information, description of the compliance framework) and to specific guidelines (how to make data FAIR, to certify a repository or service, to procure joint services).
4. Services to find, access, re-use and analyse research data generated by others, accessible through appropriate catalogues of datasets and data services (e.g. analytics, fusion, mining, processing).
5. Services to make their own data FAIR, to store them and ensure long-term preservation.

The consultation process recommended providing free of charge the services under 1, 2 and 3, as well as under 4 except when the re-use and analysis of data involves big data or large computation power, in particular via a commercial service provider. This would entail co-financing from other sources (e.g. a national or EU grant). The cost model of the services described under 5 would be determined when deciding on the long-term business model for EOSC.

Services as proposed above, that could effectively be provided under the EOSC reflect existing offers by service providers across Europe such as EGI, EUDAT and GEANT, and by existing research data repositories. Work to integrate and federate such services has already begun in

⁵⁴ Similar in nature to the existing European framework for the interoperability of public services. As foreseen by the Communication, this work should aim at maximising re-use of existing practices and standards in the science and ICT field.

⁵⁵ <https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-access-and-preservation-scientific-information>

Horizon 2020 Work Programme 2016-2017, with the EOSC-hub project and other related projects expected to deliver services under the EOSC. The projects will deliver the initial catalogue of services and data to be provided by EOSC and will define the delivery model(s) for the services. Those catalogues would be enriched periodically based on the process of federation.

Starting from	Committed resources (non-exhaustive)	Action	Milestones
2018, Q2	<ul style="list-style-type: none"> • EOSC-hub project • eInfraCentral project • OpenAIRE-Advance project • INFRAEOSC-01-2018 • INFRAEOSC-04-2018 • INFRAEOSC-05-2018-2019 (b) • INFRAEOSC-02-2019 • INFRAEOSC-03-2020 • INFRAEOSC-06-2019-2020 (a) • INFRAEOSC-06-2019-2020 (b) 	Develop initial catalogue of services to be provided via the EOSC (to be enriched periodically) and define delivery model(s)	<p>Q4 2018: Initial EOSC Catalogue of services accessible & prototype EOSC Portal accessible</p> <p>Q4 2019: Updated EOSC Catalogue of services & EOSC Portal</p>
2018, Q2	<ul style="list-style-type: none"> • EOSCpilot project • EOSC-hub project • INFRAEOSC-04-2018 • INFRAEOSC-05-2018-2019 (b) 	Develop initial catalogue of datasets accessible via the EOSC (to be enriched periodically)	Q2 2019: Initial EOSC Catalogue of datasets accessible.

The further support will be provided by the INFRAEOSC-01-2018 - Access to commercial services through the EOSC hub, INFRAEOSC-02-2018 - Prototyping new innovative services and INFRAEOSC-03-2018 calls - Integration and consolidation of pan-European access mechanisms to public e-infrastructures and commercial services through the EOSC hub.

(d) Access and interface

The consultation and evidence gathered indicate the benefits of giving users a choice between different entry points for accessing EOSC services for practical reasons and to ensure a smooth transition from legacy research data systems in contrast to implement a single access point. Work on the EOSC access and interface has already begun under Horizon 2020 Work Programme 2016-2017: the EOSC-hub project will pilot the common platform and the access to EOSC services, while the *eInfra Central* project provides a first catalogue and access to eInfrastructure services.

The entry points to the EOSC would be similar but not equivalent, and typically would consist of a web-based user interface, or front-end, which can be tailored to the specific needs and context of particular user communities. In addition, it would comprise a common platform building on the EOSC-hub project and further developed in the INFRAEOSC-06-2020 call a) and b), that would be accessible to users via machine-to-machine interfaces and which offers access to shared EOSC resources and to the full range of EOSC services.

Services provided under the EOSC would be made accessible via a EOSC portal, based on the work developed by the EOSC-hub and eInfra Central projects and further support planned in Horizon 2020 for the EOSC portal. Acting as a *universal entry point for all potential users*, the portal would have a full-fledged user interface supported by the common platform. A universal entry point usually guarantees that all users have access to the full range of services, irrespective of geographical location or scientific affiliation.

In parallel, established user communities could further develop their existing user interfaces upon the common platform, to offer access to EOSC services within their traditional work environment and give their users a choice as described above.

(e) Rules of participation

In line with good practices of multi-stakeholder governance of large-scale scientific infrastructures, the EOSC would benefit greatly from the development of Rules of Participation.

These rules would set out in a transparent and inclusive manner the rights, obligations and accountability of the different stakeholders taking part in the initiative (e.g. data producers, service providers, data and service users). The groundwork for the design of such rules is being laid primarily by the EOSCPilot project funded by the Work Programme 2016 and the work of the High Level Expert Group on the EOSC, while the EOSC Declaration set the general principles.

The consultation and available evidence suggest that the rules ought to address:

- the use of the tools, specifications, catalogues and standards (EOSC shared resources) and applicable methodologies (framework for FAIR research data);
- the principles for regulating transactions in the EOSC (e.g. financial mechanisms and procedures, agreements/bylaws established by the EOSC governance framework); and
- the applicable legal frameworks (e.g. GDPR, copyright, Data Security and Cybercrime, dispute resolution and redress mechanisms, e-commerce directive).

Moreover, the rules would apply differently to different EOSC participants, depending on their maturity and role (providers vs. users, scientists or innovators), location (EU vs. global research partners), and would need to respect the specificities of different scientific disciplines. Therefore, compliance with the rules could differ based on:

- the current situation and readiness of data infrastructures and services at the level of Member States (research infrastructures, e-Infrastructures) and disciplines (level of standardisation and integration) and the differences in their established rules and processes;
- the actual existence and variety of service providers and the actual needs of users of the EOSC (e.g. public vs private; horizontal vs specialised); or
- evidence of changing needs and practices in relation with the implementation of the rules, in particular as concerns compliance with existing legal frameworks (e.g. GDPR) and emerging ones (e.g. free flow of data).

In short, the rules of participation of the EOSC would need to take into account the established practices and current needs of all researchers and service providers.

(f) Governance

The need to establish an appropriate governance framework for the EOSC clearly emerged from the discussion with stakeholders and from the experience of existing multi-stakeholder infrastructures in the ICT domain (such as ICANN and ITF).

The EOSCpilot project has laid the ground for such a governance framework, which has been sketched out in the Horizon 2020 Work Programme 2018-20⁵⁶ under the Call INFRAEOSC-05-2018-2019 (€50m). This Call intends to support the set-up of an operational framework for the overall governance of the EOSC, including the coordination between relevant national initiatives.

The analysis of all inputs received indicates that the EOSC governance framework should support well-defined functions including strategy (e.g. setting the long-term orientation and priorities and deciding on compliance), implementation (e.g. budgetary orientations), monitoring (e.g. setting out key performance indicators) and reporting, exercised within a clear and bounded remit. Such remit would encompass the actions needed for the coordination and the federation of research data infrastructures in Europe as discussed in previous sections, notably the development and implementation of the European framework for FAIR research data, EOSC shared resources, the Rules of Participation and the EOSC portal.

⁵⁶ Commission Decision C(2017)7124 of 27 October 2017

Stakeholders and national experts converge on the need for a two-staged approach in building the EOSC governance. In the first phase, the governance could entail steering and oversight of the initial development of the EOSC, primarily led by the Member States and the Commission. In a second phase, following thorough evaluation of the first phase, the governance could oversee the initial operations and further development of the EOSC. This second phase would become more stakeholder-driven, with Member States and Commission keeping a higher-level oversight role.

As the current Multi-Annual Financial Framework runs until end of 2020 with resources committed through Horizon 2020 for supporting the EOSC's initial development, an evident cut-off date for the first phase could be the end of Horizon 2020. Any further Commission proposal for governance and decision-making beyond 2020 would be part of the Commission's proposal for the next EU R&I Framework Programme and would depend on whether EU resources for the EOSC become available under the next Multi-Annual Financial Framework.

Starting from	Committed resources (non-exhaustive)	Action	Milestones
2018, Q1	EC, with support of EOSC pilot project, High Level Expert Group EOSC, OSPP and other sources	Set up the EOSC governance framework in consultation with MS	Q4 2018: EOSC Governance established
2019, Q1	INFRAEOSC-05-2018-2019 (a)	Prepare legacy for 2 nd implementation phase (post 2020)	Q3 2020: Recommendations on strategic and financing orientations and organisational settings for the future of the EOSC, post 2020

In line with the Horizon 2020 Work Programme 2018-2020, the EOSC governance framework should ensure a proper representation of the main stakeholders and allow for an effective policy steering of the initiative, through a group/body/entity gathering stakeholders, the Member States and the Commission, to advise on effective implementation and ensure input from a wide range of actors.

The consultation process confirmed a number of principles in the EOSC Declaration, including the need for a low intervention cost, the development of light mechanisms and high accountability of the proposed framework. The consultation produced a number of orientations, which are summarised in Annex 5, for the set-up and composition of the EOSC governance framework. The outcomes of the consultation should serve as a basis for further consultation of the Member States and other relevant stakeholders, without prejudging future decisions in this area.

2.2 - Key actions and milestones

Table 2 presents a summary of the main actions in Horizon 2020 presented in the previous section, in relation to the six lines of actions of the implementation Roadmap. This provides a clear link between the above described possible EOSC federated model and Horizon 2020 implementation resources, and a timeline based on project milestones from Descriptions of Action (WP 2016-17) and descriptions (WP 2018-2020).

Table 2. Overview of the implementation Roadmap				
Starting	Action line	Milestones	Resources (non-exhaustive)	Area
2018, Q1	Develop initial EOSC federating core including the EOSC shared resources	Q4 2019: Initial EOSC federating core in place	<ul style="list-style-type: none"> • EOSC-hub project • OpenAIRE-Advance project • FREYA project 	Architecture
2019, Q1	Develop catalogue of interested and eligible (per Rules of Participation) data infrastructures to be federated into the EOSC	Q4 2019: Registry of data infrastructures of the EOSC (initial)	<ul style="list-style-type: none"> • EOSCpilot project • INFRAEOSC-01-2018-2019 (b3) 	Architecture
2018, Q4	Connect the research infrastructures identified in the ESFRI Roadmap to the EOSC	Q2 2020: Preliminary connection of most infrastructures and services to the EOSC	<ul style="list-style-type: none"> • INFRAEOSC-04-2018 	Architecture
2018, Q1	Prepare a FAIR data Action Plan	Q3 2018: FAIR data Action Plan published	<ul style="list-style-type: none"> • FAIR data Expert Group (E03464), in consultation with stakeholders 	FAIR data
2018, Q3	Define a European framework for FAIR research data	Q2 2019: European framework for FAIR research data agreed	<ul style="list-style-type: none"> • DG RTD • RDA Europe 4.0 • INFRAEOSC-05-2018-2019 (b) 	FAIR data
2019, Q1	Define a Persistent Unique Identifier policy for FAIR data	Q4 2019: FAIR persistent unique identifier policy defined	<ul style="list-style-type: none"> • FREYA project • RDA Europe 4.0 	FAIR data
2019, Q1	Develop a FAIR data accreditation /certification scheme for repositories	Q4 2019: FAIR certification scheme available.	<ul style="list-style-type: none"> • INFRAEOSC-05-2018-2019 (c) 	FAIR data
2018, Q2	Develop initial catalogue of services to be provided via the EOSC (to be enriched periodically) and define delivery model(s)	<p>Q4 2018: Initial EOSC Catalogue of services accessible & prototype EOSC Portal accessible</p> <p>Q4 2019: Updated EOSC Catalogue of services & EOSC Portal</p>	<ul style="list-style-type: none"> • EOSC-hub, eInfraCentral and OpenAIRE-Advance projects • INFRAEOSC-01-2018 • INFRAEOSC-04-2018 • INFRAEOSC-05-2018-2019 (b) • INFRAEOSC-02-2019 • INFRAEOSC-03-2020 • INFRAEOSC-06-2019-2020 (a and b) 	Services
2018, Q2	Develop initial catalogue of datasets accessible via the EOSC (to be enriched periodically)	Q2 2019: Initial EOSC Catalogue of datasets accessible.	<ul style="list-style-type: none"> • EOSCpilot and EOSC-hub projects • INFRAEOSC-04-2018 • INFRAEOSC-05-2018-2019 (b) 	Services
2018, Q1	Set up the EOSC governance framework in consultation with MS	Q4 2018: EOSC Governance established	<ul style="list-style-type: none"> • EC, with support of EOSCpilot project, High Level Expert Group EOSC, OSPP and other sources 	Governance
2019, Q1	Prepare legacy for 2 nd implementation phase (post 2020)	Q3 2020: Recommendations on strategic and financing orientations and organisational settings for the future of the EOSC, post 2020	<ul style="list-style-type: none"> • INFRAEOSC-05-2018-2019 (a) 	Governance
2018, Q2	Develop Rules of Participation in consultation with stakeholders	<p>Q1 2019: Initial EOSC Rules of Participation</p> <p>Q4 2019: Final EOSC Rules of Participation</p>	<ul style="list-style-type: none"> • DG RTD • EOSCpilot project • EOSC-hub project • High Level Expert Group EOSC • INFRAEOSC-05-2018-2019 (a) 	Rules of Participation

2.3 Link of the EOSC with the European Data Infrastructure

The European Data Infrastructure (EDI) is the second pillar of the European Cloud Initiative, foreseen to underpin the EOSC by deploying the high-bandwidth networks and the supercomputing capacity necessary to access and process large datasets stored in the EOSC. The EuroHPC Joint Undertaking (JU) is one of the main components of the EDI. The Commission proposal for a Council Regulation of the EuroHPC JU⁵⁷ is now under Council scrutiny. The plans are that the EuroHPC JU will implement, as of 2019, the European strategy on High Performance Computing (HPC), by pooling European and national resources with the aim to develop, acquire and deploy in Europe top-of-the-range supercomputers for processing big data, based on competitive European technology, so they can run ambitious scientific and industrial applications, and enable scientific discoveries and innovation.⁵⁸

Users will have access to pan-European HPC services through the EuroHPC JU/PRACE as a complementary service in the EOSC catalogue of services via GÉANT, the other main component of the EDI. In the longer term, it is expected that part of the capacities of EuroHPC and PRACE machines would be offered as cloud-based/on-demand HPC services. This would contribute to the widening of the user base of HPC, providing easier access via the cloud both to researchers in key scientific disciplines and to the long tail of science. Over time, *federated* HPC services could be then fully integrated in the EOSC catalogue of services.

2.4 Widening the EOSC

As stated in the Communication on ‘Building the European Data Economy’⁵⁹, the user base of the EOSC and of the European Data Infrastructure will expand to the public and private sectors stakeholders, creating solutions and technologies that will benefit all areas of the economy and society. Achieving this will require a collaborative effort open to all those interested in exploiting the data revolution in Europe as an essential component of global growth.

In order to achieve the above, the EOSC implementation roadmap should be iterative and take into account the impacts of widening the user base and the services it connects to, so that the EOSC could not only respond to changing needs of scientists but also to the increasing needs of other highly knowledge-intensive sectors of society that are in need of this data. This effort and the necessary principles for including third party users will also have to be included in the technical and governance discussions of the EOSC. The plan should also propose the right incentives for data sharing and motivating a large number of public and private users to use the EOSC.

2.5 Preliminary reflections on the costs and financing of the EOSC implementation

The consultation and the studies also helped the Commission services to reflect on the costs and on the financing means that are appropriate for the initiative. Overall, it may be argued that the initiative would have a cost that is only marginally higher (e.g. administrative costs) than the resources that have already been earmarked by Member States and the Commission in support of responsible research data management. This is true in particular for the support made available in Horizon 2020 (~ €300m in WP 2018-2020). The initiative should thus be largely budget-neutral at EU-level until the end of the current financial framework.

Conceptually, the cost of the EOSC in real terms ought to be estimated on a proper baseline of current cost of research data infrastructures and research data management at Member States and

⁵⁷ COM(2018) 8 final.

⁵⁸ The JU would acquire high-end supercomputing capabilities, through two pre-exascale systems by 2020 and two exascale systems by 2022/2023, and support the development of a full European HPC ecosystem in terms of technologies, applications and skills that provides Europe with the HPC and data capabilities required to compete globally.

⁵⁹ COM(2017) 9 final

EU level. This baseline includes the current running costs of research data and infrastructure, and the costs for aligning, federating and integrating research data infrastructures at national level and at the level of individual disciplines, and the cost of making all research data FAIR. Therefore, the real cost of EOSC is the cost of ensuring that current operations, upgrades and related investments happen in a coordinated and consistent way at European level. This horizontal cost, committed by the Commission in Horizon 2020, is expected to be an initial investment until 2020. The costs corresponds to establishing and operating the EOSC federating core (see Section 2.1, a) architecture), the EOSC shared resources, the EOSC Portal offering access to them, the compliance framework (including the rules of participation) and the governance framework. The Commission is also driving behavioural change and infrastructure upgrades by making open FAIR data the default option under Horizon 2020 and making related costs eligible for reimbursement.

Yet, irrespective of the EOSC initiative, all EU research systems are bound to make research data FAIR and will need to upgrade research data infrastructures and processes to stay competitive in data-driven, open science. While Member States would ultimately meet such costs at any rate, the intervention logic of the Communication, and of similar initiatives in the science domain,⁶⁰ propose that it is best, as well as most efficient and profitable, to coordinate and partly mutualise this investment by means of actin at UE level.

The additional costs incurred by Member States are expected to vary greatly, depending on the current level of readiness of their data infrastructures and the existing planning of upgrades. Reporting to the Commission by the National Point of Reference on the *Recommendation on access to and preservation of scientific information* suggests that most Member States are developing policies and planning funding for upgrading data infrastructures and improving research data management. This is further indication of the need to align and coordinate such efforts and funding at the European level to achieve economies of scale that increase performance and efficiency of research in Europe. In addition, Member States would flag the national initiatives that they want to federate into the EOSC (e.g. GO-FAIR, Helmholtz Data Alliance, etc.) and any resources they are willing to provide in kind. Finally, the upcoming update to the Commission Recommendation on access to and preservation of scientific information (C(2012)4890), building concretly on the commitment of national research funders to make research data FAIR, is expected to accelerate the behavioural change and infrastructures upgrades initiated through Horizon 2020.

However, very little to no published information exists on the current level of spending on research data infrastructures and FAIR data management in Member State and this, along with the variable situation across the EU, is why it is not possible to attach concrete figures to these costs consistently across EU28. The Final report of the High Level Expert Group on the EOSC estimated that on average about 5% of total research expenditure should be spent on properly managing and *stewarding* data in an integrated fashion.

Should EU resources become available for a second phase after 2020 (e.g. via FP9), and/or should significant national resources be ear-marked for EOSC, the sustainability of the EOSC would be ensured through a mix of funding streams, including deposit fees flowing from national research funders. A full cost estimate would be needed to be prepared for the further development and operation of the EOSC beyond 2020, addressing scalability and legacy.

Further to this, European Structural and Investment Funds could be used by Member States or regions to support selected data infrastructures in need of specific further development or for research and innovation projects in line with their Smart Specialisation Strategies and operational programmes' priorities. Due consideration could also be given to offering specific research

⁶⁰ According to the ESFRI Roadmap 2016, cutting edge research increasingly requires investments in methods and instruments, and in computing and data, that exceed the capacity of any individual Member State – the total budget absorbed by all European RIs is in the range of €10b per year.

grants, voucher schemes or coins to allow researchers finance the use of advanced EOSC services, as well as to introducing business fees for accessing the EOSC.

Tentatively, the most effective business model for the EOSC is one that links regional, national, European and third party funding to operate a federation of data infrastructures, to align practices and policies regarding research data in such a way as to gain the maximum efficiency for European researchers, the industry and the citizens. In principle, the business models and funding streams of existing data infrastructures should not be affected by the development and operation of the EOSC, as long as they are compatible with the operating principles of the EOSC. In the case of business modelling, EOSC would optimize costs and effectively manage digital assets over time and across Europe, to ensure a thriving data-driven science environment and a sustainable future.

Annex 1 – Stakeholder consultation process

1.1. Detailed response from EU institutions

Annex 1 - Main resolutions on the EOSC from inter-institutional dialogue	
<i>EP resolution of 16 February 2017 on the European Cloud Initiative P8_TA(2017)0052</i>	
EOSC	<ul style="list-style-type: none"> to present an action plan with clear working packages and timelines, defining the results to be achieved, the sources of financing and the stakeholders involved throughout the process to identify appropriate financing mechanisms for the EOSC and the EDI to provide sufficient resources for this policy area in H2020 and in proposal for FP9 to reduce the fragmentation of digital infrastructures by establishing a roadmap for actions and a robust governance structure to engage with Member States and other stakeholders in the design and implementation of the roadmap for governance and funding of EOSC to take the lead in promoting intersectoral, cross-lingual and cross-border interoperability and cloud standards; to take measures to preserve a high-quality standardisation to take into account cyber-security issues from the very first stage of all its IT initiatives to provide more clarity on the definitions used in the communication and, in particular, to create a clear distinction between the European Cloud Initiative and the EOSC
Research data	<ul style="list-style-type: none"> to ensure that all scientific research and data produced by the Horizon 2020 is open by default
<i>The transition towards an Open Science system - Council conclusions (adopted on 27/05/2016)</i>	
Open Science	<ul style="list-style-type: none"> to collaborate on incentives for an internationally accepted system for data citation, making use of initiatives and expertise that already exist to inform the Member States and stakeholders on a regular basis on the ongoing developments and outputs of the OSPP at least twice a year
Research data	<ul style="list-style-type: none"> to explore appropriate governance and funding frameworks for European Cloud Initiative to promote data stewardship and to implement Data Management Plans as an integral part of the research process increase opportunities for the optimal reuse of research data can only be realised if data are consistent with the FAIR principles (findable, accessible, interoperable and re-usable) within a secure and trustworthy environment; importance of storage, long term preservation and curation of research data, taking into consideration the capacity of the research group or organisation, as well as ensuring the existence of metadata based on international standards; to continue to make the costs incurred for both data management and preparation of research data eligible for funding in Horizon 2020, linking it to FAIR;
<i>Opinions from the European Economic and Social Committee and the Committee of the Regions</i>	
	<p>CoR</p> <ul style="list-style-type: none"> to ensure that relevant national players in all the Member States promote open science cloud services as well as knowledge-sharing more generally

	<ul style="list-style-type: none"> • to take measures to help develop cybersecurity know-how at every level necessary <i>EESC</i> • to clarify the future timetable and arrangements for widening the user base • to launch wide-ranging consultations on governance, directly involving the scientific community • to clarify the share of different financial resources (various WPs, European Structural and Investment Funds, CEF and EFSI) in the EOSC
--	---

1.2 Engagement with Member States

The Commission engaged with the Competitiveness Council, the European Research Area and Innovation Committee (ERAC),⁶¹ the ERAC Standing Working Group on Open Science and Innovation of the ERAC, and the Horizon 2020 Programme Committee Configuration 'Research Infrastructures'. Commissioner Moedas presented three six-monthly updates on the progress on the EOSC at the Competitiveness Council. The Commission services presented four updates at the ERAC Committee and at three dedicated meetings of the ERAC OSI. The EOSC was discussed at four consecutive meeting of the 'Research Infrastructures' Programme Committee in 2017.

Additionally, the Commission services analysed several institutional inputs from Member States, to ensure that the drafting of the implementation Roadmap presented in this SWD is in line with national developments.

First, the ERAC Task Force on Open Access and Innovation drafted an Opinion on Open Research Data, and adopted the Opinion on 3 February 2016. In the Opinion, Member States and Associated Countries stress the need to have open research data infrastructures that are FAIR (findable, accessible, interoperable and reusable), diverse and trusted; it flagged the need to ensure sufficient funding for open research data and for data sharing activities, as the structure of costs is complex (human resources, infrastructure, overhead, legal and use costs in a single model). The Opinion identified the European Open Science Cloud, architecture for open research data, as one the five future steps that could help progress open research data.⁶²

Second, the ERA National Action Plans pay significant attention to concrete national actions to promote open access, including a variety of actions supporting the EOSC such as creating e-infrastructures to enable access to the results of publicly funded research and storage of science-related digital content at national level.⁶³ There is significant and increasing activity, and diversity, of national implementations of open research data infrastructures.

Third, the 2017 reporting of the National Points of Reference for the Recommendation on access to and preservation of scientific information (C(2012)4890 final)⁶⁴ provides insights related to national advances in research data and infrastructures. Most Member States and Associated Countries report activities in developing policies for research data, but the variable intensity and the heterogeneity of such activities suggests the need for Europe-wide coordination to create scale and scope, and drive progress in less advanced EOSC-related areas, such as research data policy alignment, certification of repositories, inter alia.

⁶¹ <http://www.consilium.europa.eu/en/council-eu/preparatory-bodies/european-research-area-innovation-committee/>

⁶² ERAC Opinion on Open Research Data, 1202/16 3 February 2016 <https://era.gv.at/object/document/2402>

⁶³ COMMISSION STAFF WORKING DOCUMENT Accompanying the document REPORT FROM THE COMMISSION TO THE COUNCIL AND THE EUROPEAN PARLIAMENT The European Research Area: time for implementation and monitoring progress SWD(2017) 21 final <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52017SC0021>

⁶⁴ The National Points of Reference were established in the 2012 recommendation to report on the advancement of Member States with regard to it every 18 months. They are a registered EC expert group in the EC's Expert Group Register, no. E03477. The reporting of 2017 will be published in the spring 2018 by the EC.

1.3 Engagement with stakeholders: the EOSC Summit and the EOSC Declaration

The Summit brought together 110 players from all over Europe that are key for the implementation of the EOSC. Participation included all scientific fields, national scientific infrastructures, research funders and ministries of Member States and Associated Countries (via the ERAC OSI). Eighty key scientific stakeholders attended, representing all categories and scientific fields, including representatives of 15 national scientific infrastructures. Thirteen research funders and twenty officials attended from ministries of Member States and Associated Countries. Overall, there were representatives from twenty-three Member States and Associated Countries attending. The Summit reached much beyond Brussels: about 1800 people from all over the globe watched the web stream of the event.

The Summit provided strong support for the implementation of the EOSC. There was strong agreement on data culture, on the need for FAIR data and on the need to develop and gear supporting infrastructure of services; there were agreements and no fundamental objections on the much-debated issues of services, governance and financing. Participants demonstrated a strong sense of commitment towards the implementation of the EOSC. They supported the need to act immediately and swiftly in the next few months to keep the momentum achieved by a 'coalition of doers'. They agreed that the EOSC is a truly common European project which will ensure long term sustainability and support Europe to become a key player in research data. At the Summit, many participants took upon themselves to make the EOSC a reality by 2020, proposing practical commitments to its implementation as described in the Declaration (see below). In that respect, the summit proved the willingness to build the EOSC as an inclusive (of all disciplines, countries and researchers), sustainable open research data commons for Europe's research and innovation system.

The Summit resulted in the EOSC Declaration, which is composed of 33 high level statements meant to capture stakeholders' shared understanding of the action needed on the Data culture and FAIR data, Research data services and architecture, Governance and funding, to make the EOSC a reality by 2020. About 70 scientific stakeholders undersigned the EOSC Declaration following the Summit. Many of the signatories, sixty as of today, also committed to one or more actions to implement the principles of the EOSC Declaration. By endorsing the principles of the Declaration, stakeholders signalled their intention to be involved in the making of the EOSC, for instance by taking specific action, by joining the Executive Board, by providing inputs via the annual stakeholder forum, or again by joining consortia to implement the EOSC via Horizon 2020. These commitments signal the tangible intention of organizations to support the implementation of the different elements of the European Open Science Cloud. The Commission services regularly review the inputs, and has used them extensively to align the implementation Roadmap, to actual stakeholder activities, needs and requirement.

1.4 Summary of relevant studies

The Commission tendered a study on the EOSC Governance model (December 2017), to analyse empirical cases of current governance practices of major functioning, large-scale scientific networks including ICANN, IETF, AIOTI, GÉANT and ELIXIR and provide pertinent advice on the EOSC. It concludes that the EOSC requires 1) strong policy-led and funding-enabled action 2) to create a service platform, 3) to set its mission and values and 4) to define the initial services needed to attract and organise data and related research products, and 5) to facilitate access and re-use. The study recommends that the EOSC Governance model is *evolutionary*, in that it takes all stakeholders' views into account to the greatest extent possible. It recommends that the governance system is built around an *empowered community*, working with Working Groups with a clear mandate, to develop data policies in a transparent way, alongside traditional and formal governance structures.

The Science|Business report on governing the EOSC (October 2017) identifies the principles that should underpin the governance of the EOSC: trust, transparency, accountability, inclusiveness,

flexibility, pragmatism, efficiency, a global perspective and a strong focus on the needs of science. It outlines three different implementation options for the governance structure, each involving different trade-offs between simplicity/efficiency and breadth of representation. While all models seek to ensure that key stakeholders are represented, they propose different ways to streamline the governance structure to ensure that the EOSC can move at the same pace as advances in data science.

The report of the Open Science Policy Platform on the Governance and financial schemes for the EOSC provides several specific recommendations (May 2017).⁶⁵ It recommends that the EOSC should rely on multi-level and multi-stakeholder governance, ensuring representation of main stakeholder categories and disciplines, integrating national and European funders. It notes that European countries and the Commission should ensure long-term funding of services needed to enable the integration of and access to the resources that will be federated in the EOSC. Different and innovative funding schemes should be investigated to support users to use services from EOSC-certified providers, approved based on a commonly-agreed European certification scheme. Finally, it noted that the kick-off of the EOSC ecosystem needs enough coordinated financial support from a sufficiently large set of European countries and the Commission.

The OECD studies⁶⁶ provide useful insights for the governance and financing of the EOSC. The studies aim to provide a framework for developing sustainable business models and to assist policy makers and funders in supporting data repositories with a balance of policy regulation and incentives, based on extensive empirical analysis, and to help build a shared understanding necessary to develop effective and sustainable international research data networks.

Regarding governance, the reports stress the importance of policy action to steer change: authorities should work toward common definitions of, and agreements on, open data, and work toward commonly agreed and enforced legal and ethical frameworks for the sharing of different types of public research data. Responsible national and international authorities must include data networks such as the EOSC in long-term strategic planning and support processes for research infrastructure. *Processes* including standardisation, governance mechanisms and quality control are all crucial to the success of federation of initiatives. Moreover, all parts of the ecosystems and not only the *centre*, should be included as research data repositories and international research data networks are an essential part of the infrastructure for open science.

Regarding financing, the reports stress that research data repositories and networks need a clearly articulated business model aligned with policy, regulation (mandates) and incentives (including funding). Funders and host institutions should view internationally co-ordinated data networks as a long-term strategic investment and support them and engage with them accordingly. The reports stress the importance of flexibility and mixed sources of revenue over time for long-term sustainability. Furthermore, cost optimisation (e.g. reduction of duplication, synergies, etc.) should be explored in order to effectively manage digital assets over time.

The EOSCpilot project (deliverable D2.2) also provided guidance on an appropriate governance framework for the initiative. The deliverable was based on interactions between the EOSCPilot, the EC, scientific stakeholders and Member States. The deliverable outlines a three-layer governance model consisting of Strategic, Executive and Steering layers, and details the interactions and decision flows between these layers. It also designs a resource model for the EOSC. The Strategic/ institutional layer should include EU Member States and the Commission

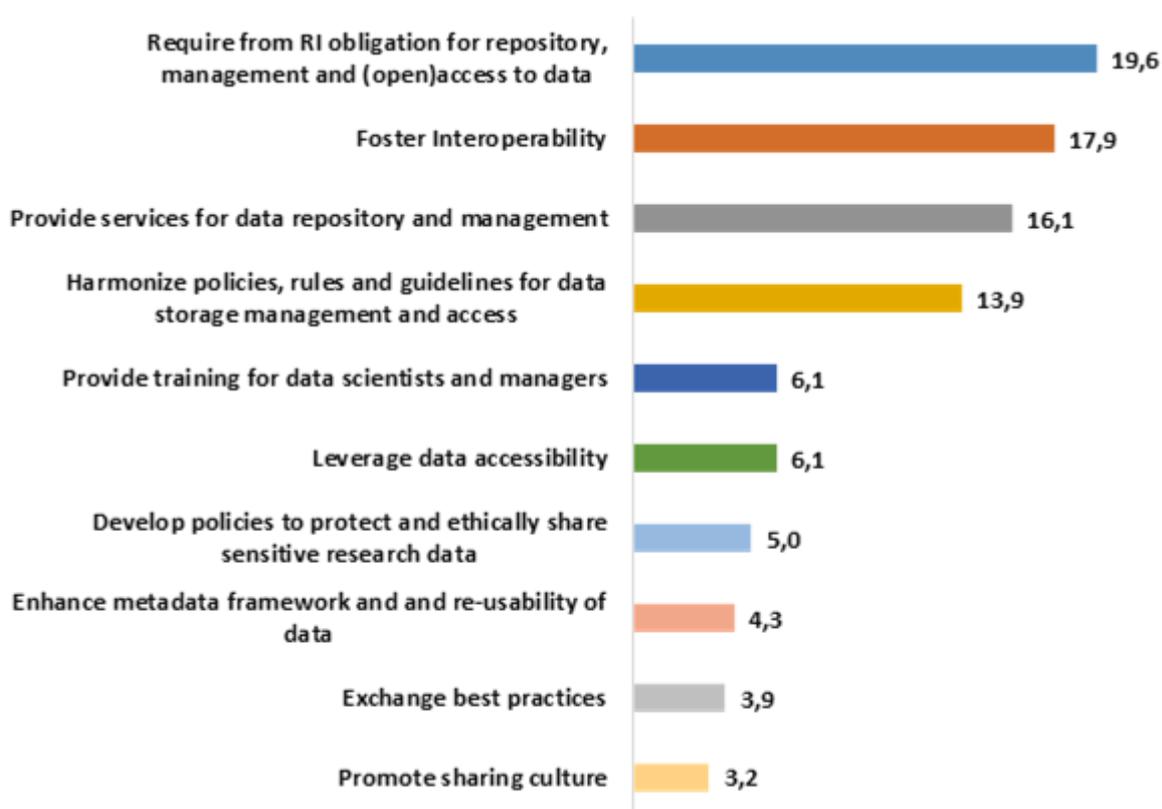
⁶⁵ The report focuses specifically on: 1) Governance structure and principles: to identify the distribution of rights and responsibilities among the different entities in the EOSC ecosystem and rules for making decisions; 2) Financial schemes: to shape the best financial mechanisms that can enable the EOSC ecosystem to flourish and deliver value in an efficient way; 3) Other relevant areas: awareness, skills development and ethics. http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

⁶⁶ DSTI/STP/GSF(2017)1/REV1/FINAL http://www.oecd-ilibrary.org/science-and-technology/business-models-for-sustainable-research-data-repositories_302b12bb-en and DSTI/STP/GSF(2017)5/FINAL http://www.oecd-ilibrary.org/science-and-technology/co-ordination-and-support-of-international-research-data-networks_e92fa89e-en

and define the strategic objectives, measuring the impact and effectiveness of EOSC against these objectives. The Executive layer should work with the Strategic layer to determine how the EOSC is provisioned and commissioned to meet these objectives and ensure that the development and operation meet the needs of the stakeholders. The Steering layer should have an advisory role and include a Stakeholder Forum.

Finally, the Commission services examined the proposed governance structure for the German-Dutch-French initiative GO-FAIR. GO-FAIR is foreseen as a contribution to the EOSC and to a global Internet of FAIR Data and Services. The governance boundaries have been initially determined by the founding Member States and will be further developed as a function of the expressed needs of the implementation networks and the participating countries. The GO-FAIR initiative appears to be in a similar and complementary position with respect to the EOSC, specifically regarding governance. The initiative has a distributed and federated architecture, built on national and international implementing networks. Four governance bodies have been proposed: a Steering Committee of Member States; an International Support and Coordination Office, for operational issues; a Stakeholder Forum representing community inputs, and an Executive Board (EB) composed of 7-10 people from the implementing networks. Initially, the initiative is funded by the founding Member States. Overall, the proposal is in line with the general principles of the EOSC Declaration regarding governance and financing of the EOSC also found in the studies reviewed in the consultation.

1.5 Priority measures to improve data management policies at national, European, International level (% of responses)



1.6 Further inputs from the stakeholder consultation

The position papers and joint e- Infrastructures statement on the EOSC identify eight elements for the success of the EOSC, that it be: 1) open in design, participation and use, 2) publicly

funded and governed with the 'commons approach', 3) research-centric with an agile co-design with researchers and research communities, 4) comprehensive in terms of universality and inclusiveness of all disciplines, 5) diverse and distributed empowering network effects, 6) interoperable with common standards for resources and services, 7) service oriented and protocol-centric, and 8) social connecting diverse communities. They also noted that many of the resources and services already exist and that most of the barriers are related to policy and concern funding, lack of interoperability, access policies and coordinated provisioning. These principles and analysis are fully in line with the e-Infrastructure Commons vision put forward by e-IRG.

Annex 2 – Actions from WP 2016-2017 of the Research Infrastructure Work Programme paving the way for the establishment of the EOSC

EOSCPilot⁶⁷

EOSCpilot (€10m) is a pilot action demonstrating how wide availability of research data and data-analysis services can be ensured through a cloud infrastructure, funded under the Research Infrastructure Part of WP 2016-2017 of Horizon 2020. EOSCpilot lasts for two years from on 1 January 2017. It includes 33 partners from 11 EU countries, which represent main stakeholder categories: domain specific research infrastructures providers, horizontal e-Infrastructure providers, research performing organisations and research funding organisations. The project is actively working towards its objectives, in particular the designs and trial of stakeholder-driven governance framework and service architecture for the EOSC, and the developments of demonstrators of integrated services and infrastructures in a number of scientific domains. It is also engaging with a broad range of stakeholders, crossing borders and communities, to build trust and skills and create consensus on the EOSC.

EOSC-Hub⁶⁸

EOSC-HUB started in 2017 to directly contribute to the EOSC implementation. The project (€30m) will integrate and consolidate services, software and data from the key existing e-Infrastructures EGI Federation, EUDAT CDI, INDIGO-DataCloud and major research e-Infrastructures through a pan-European access mechanism, providing an integrated entry point to both generic and thematic services for the scientific community. EOSC-Hub builds on mature processes, policies and tools from the leading European federated e-Infrastructures to cover the whole life-cycle of services, from planning to delivery. It aggregates services from local, regional and national e-Infrastructures in Europe and worldwide and builds a comprehensive Catalogue of Services based on the results of the eInfraCentral project (see below). Through the project's virtual access mechanism, more scientific communities and users will have access to services supporting scientific discovery and collaboration across disciplinary and geographical boundaries. The project will improve skills and knowledge among researchers and service operators by delivering specialised trainings and by establishing competence centres to co-create solutions.

OpenAIRE-Advance⁶⁹

OpenAIRE-Advance (€10m) addresses key aspects and challenges of the currently transforming scholarly communication landscape in terms of quality assurance, communication of scientific outputs with a focus on EOSC developments. It is based on the OpenAIRE network that supports, accelerates and monitors the implementation of Open Science policies, including Open Access to publications and research data. The project lays the groundwork for OpenAIRE to play a central role in the European Open Science Cloud (EOSC), enabling greater integration with Research Infrastructures and developing a catalogue of services that are inherently interoperable with and complementary to other EOSC services. It will work in close collaboration with the EOSC-Hub project.

Freya

FREYA (€5m) will provide a robust environment for a range of Persistent Identifiers (PIDs), an essential component of the EOSC. A universal and persistent mechanism will be developed for discovering elements in the EOSC interoperable research environment, through richer linking of research entities, metadata enrichment and improved machine actionability. The project will integrate and connect existing PID systems into a federated Graph, and provide a PID Services Registry. FREYA and RDA will collaborate on PID requirements, standards and protocols to

⁶⁷ <https://eoscpilot.eu/>

⁶⁸ Horizon 2020 Grant no 777536 – 2018-2020.

⁶⁹ Horizon 2020 Grant no 777541 – 2018-2020.

support interoperability. An open, sustainable framework for collaborative self-governance of PIDs and innovative services built on them will be established.

eInfraCentral⁷⁰

The project (€1.5m) builds and provides access to the catalogue of e-Infrastructure services which will feed the EOSC-Hub. The overall aim of the project is to ensure that a broader and more varied set of users (including industry) benefits from European infrastructures. The catalogue is the single point of reference for researchers and the broad community to discover and compare services and resources, as well as to monitor the performance and quality across multiple service providers. The eInfraCentral catalogue resulted from an open dialogue between e-Infrastructures to consensually define, monitor, improve and increase the uptake of their services. A beta version of the portal is already accessible.⁷¹

RDA Europe 4.0

RDA Europe 4.0 (€3.5m) addresses the need for open and interoperable sharing of research data and the need to build social, technical and cross-disciplinary links to enable such sharing on a global scale. To do so, it builds on its community-driven and bottom-up approach, which has been operational since 2012. The project takes forward the current RDA Europe effort, and brings in and structures the organisations that implemented RDA Europe since 2012. The scope of RDA Europe 4.0 is to become the centrepiece for an EU Open Science Strategy through a consolidated European network of National Nodes, bringing forward an RDA legacy in Europe, providing skilled, voluntary resources from the EU investment to address DSM issues, by means also of an open cascading grant process.

GÉANT

GÉANT (€64m in WP 2016-2017) serves as the network access provider for the EOSC, contributing, together with the National Research and Education Networks (NRENs), secure seamless high-speed multi-domain networking and wide peering together with federated identity services delivering appropriate access to cloud services, data, research infrastructures and the many other components and resources of the EOSC. In addition, GÉANT has a role in a coordinated data management framework where the network, compute and storage are all working together to serve the needs of researchers. Furthermore, GÉANT contributes to end-to-end performance optimisation and user support services including data planning consultancy, troubleshooting, training, service marketing and public procurement.

HNSciCloud

Helix Nebula Science Cloud, HNSciCloud, is a European pre-commercial procurement (PCP) initiative co-funded by Horizon 2020. Scientific research in many different domains generates massive amounts of data, creating enormous challenges for data capturing, management and processing. Today commercial cloud services do not play a significant role in the production computing environments for the publicly funded research sector in Europe. Using the approach of Pre-Commercial Procurement (PCP) leading research organisations from 7 countries have joint to pull together commercial cloud service providers, publicly funded e-Infrastructures and in-house resources to build a hybrid cloud platform on top of which a competitive marketplace of European cloud players can develop their own services for a wide range of users. The project brings together Europe's technical development, policy and procurement activities to remove fragmentation and maximise exploitation. HNSciCloud develops requirements for cloud services addressing the needs of research institutes from across Europe for multiple data-intensive research communities.

⁷⁰ Horizon 2020 Grant no 731049 – 2017-2019.

⁷¹ <http://beta.einfracentral.eu/home>

**Annex 3 - Work Programme Research Infrastructures (including e-Infrastructures)
2018-2020 -- EOSC relevant topics**

	TOPIC	Title	Type of Action	Open Date	Deadline	Budget
Work Programme Research Infrastructures (including e-Infrastructures) 2018-2020	INFRAEOSC-01-2018	Access to commercial services through the EOSC hub	RIA	05/12/17	22/03/18	€12m
	INFRAEOSC-02-2019	Prototyping new innovative services	RIA	16/10/18	29/01/19	€28.5m
	INFRAEOSC-03-2020	Integration and consolidation of pan-European access mechanisms to public e-infrastructures and commercial services through the EOSC hub	RIA	tbd	tbd	€79m
	INFRAEOSC-04-2018	Connecting ESFRI Infrastructures through cluster projects	RIA	05/12/17	22/03/18	€95m
	INFRAEOSC-05-2018-2019	Support to the EOSC governance (a) Setup of an EOSC coordination structure	CSA	10/01/18	19/04/18	€10m
		(b) Coordination of EOSC-relevant national initiatives across Europe and support to prospective EOSC service providers	RIA	26/07/18	21/11/18	€ 30m
		(c) FAIR data uptake and compliance in all scientific communities	CSA	10/01/18	19/04/18	€10m
	INFRAEOSC-06-2020:	Enhancing the EOSC portal and connecting thematic clouds (a) Support to the EOSC portal	RIA	14/11/18	20/03/19	€2m
		(b) Connecting thematic clouds into the EOSC	RIA	tbd	tbd	€5m
	INFRASUPP-01-2018-2019	Policy and international cooperation measures for research infrastructures (b5) Support to the e-Infrastructure Reflection Group (e-IRG)	CSA	05/12/17	22/03/18	€ 0.6m
Other actions GÉANT Partnership projects	Two actions covering the support to the pan-European data network for research and education, including increase of the backbone capacity, which underpin EOSC	SGA	Q3 2018	Q4 2018	€128m	

Annex 4 - The Health Research and Innovation Cloud (HRIC)

Several Horizon 2020 Calls in the field of health contribute to the establishment of a federated EOSC. The Health Research and Innovation Cloud (HRIC) aims to structure a thematic cloud for health-related research data to contribute to the European Open Science Cloud, capitalize on existing EU data infrastructures and link with existing European networks of scientific expertise and with the European High-Performance Computing initiative. Several projects would contribute to the establishment of a HRIC and in turn of the EOSC. The following topics are presented as an example:

1. SC1-BHC-03-2018 – Exploiting research outcomes and application potential of the human microbiome for personalised prediction, prevention and treatment of disease (EUR 50m);
2. SC1-BHC-04-2018 – Rare Disease European Joint Programme Cofund (EUR 55m);
3. SC1-BHC-05-2018 – International flagship collaboration with Canada for human data storage, integration and sharing to enable personalised medicine approaches (EUR 40m);
4. SC1-DTH-07-2018 – Exploiting the full potential of in-silico medicine research for personalised diagnostics and therapies in cloud-based environments (EUR 35m);
5. SC1-DTH-08-2018 – Prototyping a European interoperable Electronic Health Record (EHR) exchange (EUR 30m);
6. SC1-HCC-05-2018 – Support to a Digital Health and Care Innovation initiative in the context of Digital Single Market strategy (EUR 4m);
7. DT-TDS-01-2019 – Smart and healthy living at home (EUR 60m);
8. SU-TDS-02-2018 – Toolkit for assessing and reducing cyber risks in hospitals and care centres to protect privacy/data/infrastructures (EUR 35m).

These topics encourage the use of multidisciplinary data, high quality clinical data, experimental and in-silico methods, the sharing of data and infrastructures, to develop and validate health solutions. The Calls openly encourage projects to link with the EOSC federated architecture, towards a multidisciplinary outcomes in the context of translational medicine and personalize medicine that preserve the patients' privacy.

As a first step, HRIC aims to demonstrate the benefit of data sharing, integration and analytics in various pilot areas in the health domain. When mature, it would serve as an entry point to health-related datasets for research purposes and to a collection of research outputs and recommendations for the whole health systems/services. The HRIC would take stock of existing data infrastructures and initiatives – such as the Human Cell Atlas, Human Brain Project and the Virtual Physiological Human. It would identify good practices, lessons learned, solutions, challenges and outline future requirements of health research data (i.e. technological, governance, management and ethical), based on the concrete needs of the health community.

The aims of the HRIC are to:

- Optimise the outcomes of health research, facilitate early diagnostics and personalised treatments;
- Integrate big, real world, multi-disciplinary and simulation data, ideally in real time (e.g. real time simulations during interventional medical procedures; comparing live millions of data from analysis of whole genomes);
- Obtain faster health research results for systemic health threats (e.g. real time information sharing and data mining for early detection and control of emergent infectious diseases);
- Provide a performant, cost effective, secure and privacy-preserving infrastructure to users (e.g. single entry point/data portal for health research data; access to supercomputers and supercomputing services);
- Share resources within the whole community, enabling health researchers, industry and SMEs to have a common space where health research data can be stored, accessed, managed, re-used for mutual benefit.

Annex 5 - Summary of the results of the consultation on a possible EOSC Governance Framework

The stakeholders who participated in the consultation consider that the EOSC Governance Framework could comprise three main layers (see Figure 2):

- A structure to help ensure proper implementation and accountability (the ‘*Executive Board*’);
- An institutional group/body/entity (the ‘*EOSC Board*’) gathering the Member States and the Commission to ensure effective supervision of the implementation;
- A stakeholder representation to provide input from a wider range of actors (the ‘*Stakeholders Forum*’).

Executive layer

The executive arm of the governance – the ‘Executive Board’ – could in the short term advise on the EOSC implementation and assist with the transition beyond 2020.

The remit of the Executive Board could encompass for example:

- proposing the strategy and the work plan;
- drafting Rules of participation to guide service provision and use;
- helping oversee and steer the implementation of the agreed work plan with the help of working groups;⁷²
- monitoring the progress of the EOSC implementation;
- proposing how the user base could be broadened to the public sector and industry.

The consultation identified European stakeholder organisations that could be represented, such as ESFRI infrastructures, eInfrastructures and other research infrastructures in Europe, scientific organisations, university associations. This broad representation would allow for channelling input and advice from stakeholders and ensure coordination with relevant stakeholder initiatives.

It was considered that evidence from similar initiatives would plead in favour of a limited number of members (e.g. 10) including one Chair and one Vice-Chair.⁷³

The Executive Board could be supported - including administratively - by the Coordination Structure to be set up through the Call INFRAEOSC-05-2018-2019: *Support to the EOSC Governance*. It could also draw, on a voluntary basis, on the competences and resources of the stakeholders who committed to developing the EOSC (through the Executive Board, its working groups and the Stakeholders Forum) and on the inputs provided by relevant projects funded through Horizon 2020 and national research programmes.

For practical reasons, the set-up of the Executive Board should coincide with the launch of the Coordination Structure supported through Horizon 2020, so as to ensure full administrative and financial support from the start.

Institutional layer

Regarding the institutional and political oversight of the initiative, it was invoked that current practices in scientific coordination suggest a need for a group/body comprising delegates of

⁷² Indicatively, WGs could be tasked to map the needs of scientific communities; to channel insight from other research data management initiatives; to advise on the elaboration and implementation of the FAIR data action plan; to identify emerging risks and advise on mitigation measures; to address education and outreach; to assist the coordination of the EOSC with similar international initiatives.

⁷³ As per established Commission rules on transparency, members would be selected following a transparent procedure based on objective criteria, in compliance with, for an initial term of two years, until completion of the first phase in 2020. The consultation pointed also to the importance of the proceedings of the Executive Board to be fully transparent to ensure appropriate management of the risk of conflicts of interest.

Member States and Associated Countries and representatives of Commission services (e.g. DG RTD and DG CNECT). This layer would ensure strategic orientation, effective supervision, commitment and financial support, and coordination with national initiatives. It could be co-chaired by the Commission and a Member States delegate.

The consultation suggested a broad steering mandate for this institutional layer of governance (further named for the purpose of this document the EOSC Board), namely:

- approving the list of the Executive Board members;
- deciding the strategic orientations for the EOSC (based on the advice of the Executive Board) and approving an annual work plan;
- assessing the progress of the EOSC implementation;
- ensuring coordination with relevant Member States/Commission initiatives;
- discussing new activities in support of the EOSC, including its long-term sustainability.

It was suggested that the decisions of the EOSC Board should not be binding on its members (Member States, Associated Countries and Commission) nor engage their financial responsibility unless this is explicitly agreed.

Stakeholder layer

To ensure adherence to an inclusive process for the research data management communities, it was proposed that a broader *Stakeholders Forum* with an intelligence gathering and consultative role could act as a 'sounding board', bringing together scientific/user communities, research institutions, research infrastructures and e-Infrastructures, and specialised EU agencies.

The participation in the Stakeholders Forum could require adherence to the guiding principles of the EOSC, for example commitment to the principles of the EOSC Declaration that underpin the work presented here.

Figure 2. A Possible EOSC Governance framework as it emerged from the consultation

