Solutions for a Sustainable EOSC

A strawman report from the Sustainability Working Group

Abstract

This document explores possible means for sustaining the European Open Science Cloud beyond its initial phase which terminates at the end of 2020. The document is intended to lead discussion about the possible options and is not to be seen as a final proposal. It considers the financing model, legal vehicle, governance structure as well as the regulatory and policy environment of the EOSC in a series of iterations as it grows from an initial minimal viable product to enable digital marketplaces and extending the user base to the public sector and industry.

Editors: Bob Jones, Claire Devereux

Members of the Sustainability Working Group

Rupert Lück, Executive Board, EMBL (Co-Chair),
Conor O'Carroll, Executive Board, Science Europe (Co-Chair)
Maarja Adojaan, Estonian Research Council, Estonia
Lajos Bálint, KIFÜ, Hungary
Emanuele Barbarossa, DG RTD, European Commission
Jean-Yves Berthou, INRIA, France
Paolo Budroni, Vienna University Library and Archive Services, Austria
Mirjam van Daalen, PSI, Switzerland
Claire Devereux, UKRI, UK
Franciska de Jong, CLARIN ERIC, Netherlands
Maria Häll, Swedish Research Council, Sweden
John Howard, University College, Ireland
Bob Jones, EOSC Secretariat, CERN (Rapporteur)
Norbert Meyer, Poznan Supercomputing & Networking Center, Poland
Johnny Mogensen, Ministry of Science and Higher Education, Denmark
Panayotis Moschopoulos, DG RTD, European Commission
Liina Munari, DG CNECT, European Commission
Henriikka Mustajoki, Federation of Finnish Learned Societies, Finland
Alberto Pérez, Red.es, Spain
Federico Ruggieri, GARR, Italy
Afrodite Sevasti, GRNET, Greece
Roar Skålin, The Norwegian Meteorological Institute, Norway
Gintare Tautkeviciene, Kaunas University of Technology, Lithuania
Klaus Tochtermann, ZBW - Leibniz Information Centre for Economics, Germany

Foreword

The Sustainability Working Group held its first meeting, face to face, in Frankfurt on 11th July 2019 and followed-up with a telephone conference during the month of August. Thanks to the flexibility and commitment of the members during this holiday period, we are very pleased to be able to distribute what we consider to be a relatively complete first version of this document. We would also like to thank the EOSCsecretariat project and in particular the work package 5 partners for providing detailed input on the legal and regulatory aspects. The document highlights a number of outstanding questions that need to be addressed and makes several recommendations for further work. We consider this a living document which will continue to evolve over the coming months and we look forward to receiving feedback from the governing bodies, contributing projects and working groups so that all the stakeholders can converge on a common way forward.

Rupert Lück, Executive Board, EMBL (Co-Chair),
Conor O'Carroll, Executive Board, Science Europe (Co-Chair)

Acknowledgements

We would like to thank the following people for their contributions to this document:

David Foster, CERN
Jessica Klemeier, EMBL
James Mitchell, Strategic Blue
Silvana Muscella, TRUST-IT
Marie Timmermann, Science Europe

Executive Summary

A key objective of the Sustainability Working Group is to identify the most feasible alternatives for the Financing Models and Legal Vehicle of the EOSC by the second quarter of 2020.

The primary stakeholders of the EOSC are end-users (research communities, long tail of science, business organisations), service providers (including data and compute services), as well as research funders.

The EOSC is considered as a public good that could potentially serve 1.7 million researchers in Europe and progressively expand its user base to include the wider public sector and the private sector (business organisations).

For EOSC to be a success, it must be widely adopted by researchers and that implies it must provide services that allow them to pursue their research activities more effectively. Researchers are practically minded and will only adopt EOSC if it makes their research practices simpler by providing easy to use services that interoperate to support all phases of the research lifecycle. Readily available training and documentation employing the latest e-learning techniques will be needed to reduce the barriers to adoption.

The sustainability of EOSC depends not only on sound financial, legal and governance models that create added-value for the stakeholders but also on the incentive and rewards for researchers that encourage them to participate in a culture of sharing the results of their research. Without such incentives and rewards it is possible that the uptake of the EOSC could be stymied by lack of engagement from researchers.

While the services to be provided via EOSC to end-users are expected to be free of charge at the point of use, they are not without significant cost to build, maintain and operate. There needs to be political will not just towards the notion of open research, but also to fund it in practice, at a reasonable level, and the governments and research funders have to be in a position to supply such funding, in accordance with a jointly agreed cost sharing model, and to coordinate their approach, policies and legislation.

Measures are proposed to attract a large number of data providers to EOSC by sustaining a service-oriented environment where it is possible for innovators to exploit FAIR[1] data using the latest IT technologies and trends.

There are a range of business models which can support the EOSC ecosystem. However, as the literature and the documented use-cases highlight, open access systems rarely achieve self-sustainability. Consequently, long-term funding will be required to ensure the EOSC continues to exist and serve its users.

[1] FAIR - a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. https://www.force11.org/group/fairgroup/fairprinciples

EOSC, as an assumed key contributor to the ERA (European Research and Innovation Area), has been identified as a possible candidate for European Partnerships to accelerate its implementation and the selection of a co-programmed or co-funded partnership will depend on the decision of the Member States and appetite by the current governance for the EOSC post-2020.

The expansion of EOSC takes an iterative approach, starting from a Minimum Viable Product (MVP) that offers added value to researchers, with each iteration progressively bringing more functionality and services for a wider user base. The objective of the first iteration is to boot-strap the EOSC by providing a *Federating Core* as a MVP that will enable the federation of existing and planned research data infrastructures. The *Federating Core* encompasses elements needed to allow research targeted services to operate but does not, itself, provide the means to transfer, store, process or preserve research data, which needs to be undertaken by a federation of research data infrastructures. The sustainability of the EOSC depends on the sustainability of the contributing infrastructures, a factor that needs to be considered at later stages.
The benefit of the *Federating Core* for the stakeholders is to gain intelligent access to a sophisticated resource and tools that deliver more collective value than the aggregate of the individual national and thematic contributions.

The second iteration builds on the *Federating Core* to provide an environment where *Digital Marketplaces* can flourish. A digital marketplace includes data and services from both publicly funded and commercial providers that are accessible against payment. It is expected that rivalrous services, such as those that store or preserve research data as well as those that compute against it, will be made available via *Digital Marketplaces*.

The initial focus of this second iteration will be to grow *Digital Marketplaces* serving the public funded research community by offering services that store, preserve or add value to data (e.g., derived data, tools or analysis) through the payment of fees and/or subscription. Such services would support different aspects of the research activities including those necessary to implement the researchers' data management plans.

The set of *Digital Marketplaces* will then be expanded to cater for the requirements of end-users from the public sector who are not involved in research activities but want to exploit research data. As a subsequent step, *Digital Marketplaces* dedicated to the requirements of end-users from the private sector will be encouraged, without distorting market competition, so that they can exploit the research data and associated services for commercial gain.

Participation in EOSC and contribution to the financing model of the *Federating Core* must be positioned at a national level where governments have a political interest in encouraging open research and the means to define national policies that can support it.

Research population size as reported by Eurostat may be considered as a basis for determining the financial contribution to the Federating Core by participating countries since it is a reasonable measure of the likely usage that will be made of the services provided.

As an initial EOSC Minimal Viable Product (MVP), the *Federating Core* limits the risks of failure by not engaging in the curation of research data which remains the responsibility of the contributing research data infrastructures. Specifically, should the EOSC MVP fail, no research datasets will be lost.

EC directives on GDPR, Copyright, Database and Procurement as well as legislation concerning VAT and employment will apply to the *Federating Core* and consequently there must be a legal entity responsible for the *Federating Core*. There are several possible forms the legal entity could take, such as: Host agreement with an existing public or private organisation; European Research Infrastructure Consortium (ERIC); Joint Undertaking; AISBL or similar.

Additional legal complexity will be added when the EOSC is extended to support *Digital Marketplaces* and the user base expands to include the public sector, industry and possibly citizens not belonging to any organisation (citizen scientists). Thus, the decision on the legal vehicle to be implemented should not be taken on the basis of the *Federating Core* alone.

The concept of *Digital Marketplaces* assumes all services provide fee-based access and service providers are free to define the pricing model for their services which would be visible to end-users via the marketplace. The marketplace announced fees should reflect subsidies and state-aid, where applicable, so that the Member State funding is visible/traceable across the different service layers of the EOSC. The financing model for the operation of a marketplace itself may be based on a combination of a registration charge to be paid by the service provider for each paying service when it enters the marketplace and a transaction charge paid by the service provider based on the total volume of consumption of the service.

The funds to pay for researchers' use of fee-based services may come from multiple sources such as research grants at national or research funder level as well as a centralised funding model whereby multiple countries or agencies collectively fund access to services.

The costs and overheads of operating a marketplace and procuring services can be reduced by providing a procurement service through aggregated procurement. Working examples already exist, such as the commercial cloud service marketplace operated GEANT.

Digital Marketplaces introduce procurement and financial transactions into the EOSC ecosystem. This will impact the obligations and risks for market operators, service providers and service consumers.

The selection of market operators must be done through an open and transparent process. The regulations of the Public Procurement Directive may apply, in particular if incentives are needed to establish *Digital Marketplaces* serving publicly funded researchers. The host of

the *Federating Core* must establish contracts and Service Level Agreements with the market operators.

The exceptions for scientific research in the Copyright and Database directives does not apply to industrial or public use of research data. Thus, compliance with these two directives must be ensured.

In terms of governance, a number of points need to be addressed, including:

- There is a conflict between the breadth of representation needed in an expanding EOSC and the effective size of the executive arm of governance;
- The rules for joining and leaving each board require careful consideration, including in areas such as minimal terms and membership (eligibility, types etc).
- Are the representation channels for key providers of the *Federating Core* appropriate for a mature EOSC?
- To meet EOSCs global ambitions there may come a time when membership of EOSC is requested from parties beyond the current Member States and Associated Countries. Further consideration is needed over the policy for future membership, types of membership and levels of entities represented. All will have implications on governance, such as the composition of boards, and voting rights and majorities.
- The resourcing available to deliver EOSC governance must be realistic and appropriate for the scale of the undertaking and a coordination structure will continue to be required to support the governance.

In terms of timing, the EOSC Strategic Implementation plan states that the initial EOSC *Federating Core* should be in place by the end of 2019 and that the connection of most infrastructures and services to the EOSC should be made by the end of the 2nd quarter of 2020. So the end of the 2nd quarter of 2020 would correspond to the transition from *Federating Core* to *Digital Marketplaces*.

The scale and complexity of bringing together all contributors to produce annual work plans and reporting represents a full-time activity. In order to support the work of the Executive Board, the EOSCsecretariat project is requested to commission, via the co-creation budget, the engagement of a professional entity to produce all the necessary deliverables for the planning phase of EOSC. This activity should start in 2019 and produce a first version by the end of the first quarter of 2020.

Outstanding questions that need further investigation, include:

- The means of remunerating publicly funded service providers requires further investigation. The cost models and fees for public service providers need to be considered, visible and traceable.
- The introduction of the role of digital marketplace operator may require some additions to the EOSC Rules of Participation.
- A more detailed requirements analysis would be helpful to determine the needs and added value of EOSC to user groups from the public sector and industry.

Contents

Introduction

As defined in '*Implementation Roadmap for the European Open Science Cloud*'[2], the objective of EOSC is to give the European 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"*

A key objective of the Sustainability Working Group is to identify the most feasible alternatives for the Financing Models and Legal Vehicle of the EOSC by the second quarter of 2020. In order to identify such alternatives, the Sustainability WG decided it was first necessary to clearly define the scope and functions of the EOSC.

The concept of the EOSC as a Minimum Viable Product (MVP) that offers added value to researchers was taken as a starting point with its scope as described in the Strategic Implementation Plan:

> "*Based on the consultation of stakeholders, 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.*"

The Sustainability WG has collectively reviewed the assembled background material, including the results of contributing projects, to assess alternatives in four domains:

- Analysis of financing models taking into account alignment with national investments and roadmaps as well as European Commission programmes and funding instruments.
- Identification of the conditions under which a legal vehicle for EOSC is necessary, and possible structures that are compatible with existing national and European legislation.
- Governance framework options for each possible legal vehicle.
- Regulatory and policy environment relevant to the EOSC.

The assessment takes an iterative approach, starting from the MVP, with each iteration progressively adding more functionality and services for a wider user base and service providers while analysing the impact such an expansion would have on the four domains listed above.

[2] Implementation Roadmap for the European Open Science Cloud, SWD (2018) 83, March 2018, https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

This document starts by identifying the stakeholders involved in EOSC then explores the nature of EOSC and its essential characteristics if it is to become a desirable undertaking that merits being sustained. Possible business models are then discussed and their applicability to EOSC considered together candidate European-level funding instruments. Each iteration in the expansion of the EOSC is then described and the document finishes with a number of points for further consideration.

Stakeholders

The EOSCpilot project report *Shaping the European Open Science Cloud*[3] and the *Turning FAIR into reality* report identified the important stakeholder groups for the EOSC. This set of stakeholders has been extended with information from the EOSC-hub project[4]:

Research communities: practitioners from all research fields, clustered around disciplinary interests, data types or cross-cutting grand challenges.

Long tail of science: Individuals and small teams that collect data for specific projects. This includes citizen scientists. Broadly defined, Citizen Science is "*scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions*"[5].

Business organisations: for-profit organisations that wish to exploit services and data for commercial purposes.

Institutions: universities and research performing organisations.

Publishers: not-for-profit and commercial, Open Access and paywall publishers of research papers and data.

Data/service providers: domain repositories, research infrastructures (e.g. ESFRIs) and e-infrastructures, institutional, community and commercial tools and services.

Data stewards: support staff from research communities and research libraries, and those managing data repositories.

Standards bodies: formal organisations and consortia coordinating data standards and governing procedures relevant to FAIR, e.g. repository certification, curriculum accreditation (e.g. W3C, NIST).

Coordination fora: global and national bodies such as the Research Data Alliance, CODATA, WDS Communities of Excellence, GO FAIR, Dutch Coordination Point (LCRDM) and similar initiatives.

Policymakers: governments, international entities like OECD, research funders, institutions, publishers and others defining data policy.

---

[3] Shaping the European Open Science Cloud, EOSCpilot project, November 2017, https://eoscpilot.eu/sites/default/files/booklet_november2017_web-1_22122017.pdf
[4] EOSC-hub Communications and Stakeholder Engagement Plan, September 2018, https://documents.egi.eu/document/3301 , EOSC-hub D12.1 Procurement requirements and demand assessment, June 2019, https://documents.egi.eu/document/3466
[5] https://ec.europa.eu/research/openscience/pdf/citizen_science_recomendations.pdf

Research funders: the European Commission, national research funders, charitable organisations and foundations, and other funders of research activity.

National Research and Education Networks (NRENs): National Research and Education Network. In most countries a provider of high-performance network connectivity and services, especially focused on trust & identity and collaboration platforms for research communities.

The *Prompting an EOSC in Practice* report[6] highlighted that the **primary stakeholders** are **end-users (research communities, long tail of science[7], business organisations), service providers (including data and compute services),** as well as **research funders**. The iterations in this document focus on the value propositions for these stakeholder groups. It is also recognised that an actor in the EOSC ecosystem may hold multiple roles, for example, an end-user may also be a service provider.

[6] https://publications.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1

[7] when Citizen Scientists are acting as agents of an organisation or Institution, they fit the EOSC End-User definition; when not affiliated with an Institution, Citizen Scientists are defined as 'Consumers'

EOSC as a public good

Cameron Neylon, in his research article entitled '*Sustaining Scholarly Infrastructures through Collective Action: The Lessons that Olson can Teach us*'[8] examines sustainability models to support infrastructures that underpin scholarship and research. Such infrastructures include repositories, curation systems, aggregators, indexes, standards etc. that are effectively public goods. In this strawman document we consider the EOSC as a public good that could serve 1.7 million researchers in Europe and progressively expand its user base to include the wider public sector and the private sector (business organisations) which collectively, represent the 70 million professionals[9] including innovators at large that have the potential to produce economic value via the EOSC.

A key goal of EOSC is to shift the research enterprise in Europe towards an open access model, and many European countries are implementing national programmes that are aligned with the European Commission Recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information[10]. One such example is the National Framework on the Transition to an Open Research Environment published by the Government of Ireland[11] with an implementation plan that is

> "*built on a common understanding that a shift in accepted practices is needed and a genuine willingness to adapt and to do things differently.*"

and the framework has the goal of supporting

> "*the free flow of information across national and international research communities, contributing to research-enabled teaching and learning; citizen science; open innovation; and greater transparency, accountability, and public awareness of the results of publicly funded research. The transition to an open research environment has a key objective of enhancing and supporting research excellence across all disciplines, research integrity, and public trust in research*."

This collective political will for a transition to open research demonstrates a need to consider how communities, platforms, and finances interact and suggests that a political economic analysis has real value. The Science|Business Network report '*Why Open Science is the Future (And how to make it happen)*' explores the case for open science[12] but notes

> *"in reality, progress is patchy. Worryingly, one online survey in the autumn of 2018 found that only 11% of researchers shared data from their last project with people they don't know personally, down from 14% in 2016.*

---

[8] https://doi.org/10.5334/kula.7
[9] See the 'Implementation Roadmap for the European Open Science Cloud' document referenced in the Introduction.
[10] https://www.eosc-portal.eu/sites/default/files/CELEX_32018H0790_EN_TXT.pdf
[11] http://norf-ireland.net/wp-content/uploads/2019/07/NORF_Framework_10_July_2019-2.pdf
[12] https://sciencebusiness.net/report/why-open-science-future-and-how-make-it-happen

With some notable exceptions, the sharing of research data and tools occurs on a piecemeal basis, facilitate by trusted bilateral relationships, rather than through fully open platforms. Moreover, openness is not yet fully designed and embedded into the scientific process. Too often, research tools and data are opened up as an afterthought through a time-consuming retrofit. Another major obstacle is the need to support the long-term preservation of research data and the software code, tools and operating environments required to make sense of the data. Depending on the discipline, important datasets can still yield scientific breakthroughs decades after they were first generated."

There needs to be political will not just towards the notion of open research, but also to fund it in practice, at a reasonable level, and the governments and research funders have to be in a position to supply such funding, in accordance with a jointly agreed cost sharing model, and to coordinate their approach, policies and legislation. Beyond the financing models included in this document, we note that the investments needed to ensure data is FAIR, identified by the EC Expert Group's report on FAIR Data[13], are essential if EOSC is to be able to offer attractive content and we reiterate the groups' key recommendations:

"Funders who issue requirements on FAIR must provide support to ensure the components of the FAIR ecosystem are maintained at a professional service level with sustainable funding. Service providers should explore multiple business models and diverse income streams"

The legal vehicle and governance structure also need to consider the infrastructures supporting EOSC as a public good. A common set of principles for foundational infrastructures has been developed by researchers that rest on three pillars:

- transparency and community governance;
- financial sustainability, efficiency, and commitment to community needs;
- and mechanisms to protect integrity, as well as manage and mitigate the risk of failures.

For EOSC to be a success, it must be widely adopted by researchers and that implies it must provide services that allow them to pursue their research activities more effectively. While the services to be provided via EOSC to researcher are expected to be *free at the point of use*[14], they are not without significant cost to build, maintain and operate. Researchers are practically minded and will only adopt EOSC if it makes their research practices simpler by providing easy to use services that interoperate to support all phases of the research lifecycle. Readily available training and documentation employing the latest e-learning techniques will be needed to reduce the barriers to adoption. Adherence to a service

[13] Turning FAIR Data into Reality, https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf#page=55

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

framework that ensures such characteristics while minimising the support effort will need to be verified for all the services proposed via EOSC.

The approach taken in this strawman document is to place the researcher as an end-user at the centre and determine what added-value the EOSC will bring to them.

The sustainability of EOSC depends not only on sound financial, legal and governance models that create added-value for the stakeholders but also on the incentive and rewards for researchers that encourage them to participate in a culture of sharing the results of their research. Without such incentives and rewards it is possible that the uptake of the EOSC could be stymied by lack of engagement from researchers.

Finally, it is very important that the EOSC initiative, for it to be considered inclusive and supplement the activities of mature and well-structured research communities rather than competing with them, whilst taking on-board and implementing national strategies, business models, relations between different stakeholders (researchers, universities, research funders, NRENs etc) as well.

Integrity and ethics

Another aspect that will impact the sustainability of EOSC is how the ethical dimension can best be included within its policies, structures and services. The EOSCpilot project recommended[15] that as a minimum:

- EOSC as an organisation should commit to act, and be seen to act, in an ethical manner, with policies and processes that reflect that commitment,
- EOSC should have structural mechanisms in place to support research integrity, for instance by establishing metadata systems that ensure accurate provenance data and appropriate acknowledgement of previous work.

The EOSCpilot project also strongly recommended the establishment of a coordinating body, an Ethical and Legal Advisory Board, to monitor and report on the ethical practices of EOSC.

[15] EOSCpilot White Paper 4: Ethics Supporting Document to D3.3 Draft Policy Recommendations, 2018, https://eoscpilot.eu/sites/default/files/eoscpilot_d3.3_whitepaper_4_ethics.pdf

Business Models

A challenge for sustaining public goods is to avoid free-loading, where someone who does not contribute to the support of an infrastructure nonetheless gains the benefit of it. Research suggests that there are only three ways to address this for large groups:

- compelling all potential end-users, often through some form of taxation, to support the infrastructure;
- providing non-collective (club) goods to contributors that are created as a side-effect of providing the collective good;
- or implementing mechanisms that lower the effective number of participants in the negotiation (oligopoly).

From analysing many case-studies, Neylon suggests that any successful sustainability model will depend on some mixture of these three approaches for resourcing. Consequently, the funding models driving the expansion of the EOSC as proposed in this strawman document draw on elements from all three approaches to ensure the EOSC ecosystem offers high quality, cost-effective and desirable services.

Funding Models for Open Access Repositories, published by the Royal Irish Academy[16], examined 14 different funding streams, grouped into six classes (institutional, philanthropic, research, audience, service, volunteer), being pursued by open digital repositories to support their endeavours, with a particular focus on academic research data repositories. These models of funding open repositories are relevant for EOSC and each has its own advantages and constraints which have to be taken into account in determining how they could collectively support open research. For example, philanthropy might be a source of funds for

> "*archiving the data from research projects. It might also be the case that philanthropic donations can be used to leverage matched state funds, or alternatively state funding is used to try and leverage philanthropic funding or corporate sponsorship. There are two issues with philanthropic funding. Firstly, it is usually best sourced with respect to specific sub-projects rather than core activities. Secondly, it is cyclical in nature, meaning it is difficult to plan multi-annual budgets given the uncertainties over funds raised*."

Some models may be considered inconsistent with the political will of EOSC and the culture of research, such as those that rely on funding services through advertising revenue (and also selling data about end-users to data brokers), or involve the transfer of intellectual property rights for what the research communities consider among their most valuable assets, the data itself. Similarly, some models can only be applied to selected categories of research data, such as those that have commercial value ("*datasets such as transport, weather, health and map data all have potentially high commercial value. However, cultural*

[16] http://dx.doi.org/10.3318/DRI.2015.4

heritage and data from relatively esoteric research projects have much weaker direct commercial value"[17]).

The article '*Infomediary business models for connecting open data providers and users*'[18] identifies a range of business models making use of open data based on 12 cases taken from the Dutch national open data repository[19]. The two main differentiating variables of these business models are:

- *Level of access to data* - some examples provide a predefined view on the data, providing easier usage at the expense of higher costs and less innovative use, whereas access to raw data often increases the complexity of use but gives more freedom and can result in novel data usages. This strawman document assumes that a range of levels of access to data is required.
- *Level of dialogue* - some examples simply present open data while others provide the opportunity for user generated content or dialogue with other users and providers. Social media are sometimes used for ratings and discussions but rarely to stimulate dialogue or influence policy making. This strawman document assumes user engagement will be encouraged.

The article '*An analysis of business models in Public Service Platforms*'[20] studied the business models of public service platforms in quasi-markets[21] using 14 cases from Sweden. It highlighted that the traditional business models for public sector e-services owned by public agencies were designed to provide neutral information on a non-commercial basis in support of informed choice. However, more recent examples cited were not so clear-cut between public e-service and e-business with a 'hybrid' form of public services support with its associated business models that has the potential to serve users as they search for information. The technologies and sophisticated services of evolving Public Service Platforms drive the evolution from traditional PSP business models to hybrid ones. This strawman document assumes such a hybrid approach since EOSC is to engage a diversity of stakeholders from the public and private sectors.

The European Investment Bank published a report[22] which included a section on the EOSC and found that

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

[17] http://dx.doi.org/10.3318/DRI.2015.4
[18] Infomediary Business Models for Connecting Open Data Providers and Users, September 2014, Social Science Computer Review 32(5),https://doi.org/10.1177/0894439314525902
[19] https://data.overheid.nl/english
[20] Agneta Ranerup, Helle Zinner Henriksen, Jonas Hedman, Government Information Quarterly Volume 33, Issue 1, January 2016, Pages 6-14, https://doi.org/10.1016/j.giq.2016.01.010
[21] https://en.m.wikipedia.org/wiki/Quasi-market
[22] Financing the future of supercomputing How to increase investments in high performance computing in Europe, European Investment Bank, 2018, doi:10.2867/31460

Consequently, this strawman document proposes measures intended to attract a large number of data providers to EOSC by sustaining a service-oriented environment where it is possible for innovators to exploit FAIR data using the latest IT technologies and trends.

To summarise, there are a range of business models which can serve specific areas of the EOSC ecosystem. However, as the literature and the documented use-cases highlight, open access systems rarely achieve self-sustainability. Consequently, long-term funding will be required to ensure the EOSC continues to exist and serve its users. As the Science|Business Network concluded[23], a progressive, long-term approach to funding the EOSC is needed:

> *"Although it will ultimately need to break even, the initial priorities for the EOSC have to be driving participation and usage. Like most private businesses, the EOSC will probably need to operate at a loss (be subsidised) in its early years to ensure its proposition is appealing to both the data providers and the data users."*

European Partnerships

The Shadow Programme Committee for Horizon Europe is currently considering various Partnership models for parts of the research and innovation funding programme. Partnerships provide mechanisms to consistently aggregate research and innovation efforts into more effective responses to the policy needs of the Union, developing close synergies with national and regional programmes, bringing together a broad range of actors to work towards a common goal, and turning research and innovation into socio-economic results. As such, they are powerful instruments to address global challenges by translating common priorities into concrete roadmaps and coordinated activities.

European Partnerships[24] will be established for addressing European or global challenges only in cases where they will achieve objectives of Horizon Europe more effectively than the Union alone and when compared to other forms of support of the framework programme.

EOSC as an assumed key contributor to the ERA (the European Research and Innovation Area), has been identified as a possible candidate for Partnerships[25] and two models are being discussed as possible implementation modes: Co-programmed European Partnership and Co-funded European Partnership.

- Co-programmed European Partnerships: Participation in partnerships set up on the basis of memoranda of understanding and/or contractual arrangements between the Commission and private and/or public partners (such as industry, universities, research organisations, bodies with a public service mission at local, regional, national or international level, or civil society organisations including foundations and

[23] https://sciencebusiness.net/report/why-open-science-future-and-how-make-it-happen
[24] https://www.era-learn.eu/partnerships-in-a-nutshell/r-i-partnerships/transition-to-horizon-europe
[25] As discussed at the meeting of the Shadow Configuration of the Strategic Programme Committee on 27 June 2019

NGOs), specifying the objectives of the partnership, related commitments from all involved sides for financial and/or in-kind contributions of the partners, key performance and impact indicators, outputs to be delivered and reporting modalities. Co-programmed European Partnerships are based on the current contractual public-private partnership (cPPP) model of Public-Private Partnerships in Horizon 2020. Future models will include the identification of complementary research and innovation activities that are implemented by the partners and by the Programme;
- Co-funded European Partnerships: Participation in and financial contribution to a programme of research and innovation activities, specifying the objectives, key performance and impact indicators, and outputs to be delivered, based on the commitment of the partners for financial and/or in-kind contributions and integration of their relevant activities using a Programme co-fund action. They use a more or less centralised blending of EU and national public and/or other funding sources and are based on the ERA-NET Cofund and EJP Cofund model in Horizon 2020.

Partnership approaches could be particularly suited to accelerate the implementation of the EOSC from 2020 onwards. Each of the partnership models would provide some or all of the following benefits:

- Support inclusion of various types of public and private stakeholders;
- Allow different types of actions (CSA, RIA and PI/PCP) to consolidate an inclusive EOSC along the whole research data life cycle;
- Test new, flexible funding schemes and governance practices in partnership with different types of stakeholders.

However, the selection of a co-programmed or co-funded European Partnership will very much depend on the decision of Member States and appetite by the current governance for the EOSC post-2020.

Federating Core

The objective of the first iteration is to boot-strap the EOSC by providing a *Federating Core* as a minimum viable product that will enable the federation of existing and planned research data infrastructures. The minimum viable product represents the smallest common denominator permitting the aggregation around disciplinary poles and national infrastructures. The benefit for the stakeholders is to gain access to a sophisticated shared resource and its tools that deliver more collective value than any single contribution.

A political goal of EOSC is to promote open research across Europe and consequently the *Federating Core* should be as widely used as possible. The *Federating Core* includes a restricted set of basic integration services that enable data to be FAIR. EOSC is interpreted as an environment that simplifies finding and accessing research data and getting results seen by a larger audience, where products and services are searchable and findable.

Value-proposition

For a researcher, the *Federating Core* provides the means to discover, share, access and re-use data and services that can contribute to and support their research activities. It also offers an environment in which a researcher can promote or advertise their own results to a wider audience in a manner that satisfies the open science and open access obligations of the research funders supporting their research.

For a service provider, the *Federating Core* provides a channel through which they can publish their service or bundle their service with other offerings, thus multiplying its value. Interactions with the publicly funded research sector (including research sponsored by private philanthropies, such as the Wellcome Trust as just one example) will allow a service provider to identify end-user needs and make use of the *Federating Core* services.

For a research funder, the *Federating Core* will provide insights and knowledge about end-users' needs, service/data uptake and re-use as well as the means to assess return on investments and the impact of policy decisions. For example, a government/funding agency receives statistics about their national open science engagement gathered through the monitoring function of the federating services (i.e. the sort of information that can be found in the GEANT compendium[26] but encompassing the metrics to be defined for EOSC) which will help them understand the impact of their open science policies. The importance of the

[26] https://compendium.geant.org/#!/

monitoring must not be underestimated, as highlighted in the report by the European Open Data Portal[27]:

> "*This report finds that, in order to ensure the long-term sustainability of Open Data initiatives, portal owners and data publishers need to be able to effectively understand and communicate the progress and impact of their initiatives. Measuring portal performance and monitoring the use and impact of Open Data are serious challenges that portals across Europe are facing. These challenges are inherent to Open Data given its non-rivalrous nature, the lack of requirements to declare use and the manner in which it generates positive network effects.*"

In short, the value proposition for a participating country or research funder is:

- Support in strategic direction and funding actions through the EOSC governance framework and policies.
- Increase impact of funding instruments and policies through wider access, federation and re-use of research data.
- Align policies to increase excellence and impact of funded research.
- Provide a clear path to increase the number of participating organisations engaged in open research.

For an institution, the *Federating Core* permits its service/data providers to be added to the EOSC catalogs and become visible via the portal; its research population to be authenticated and access the portal.

The EC commissioned a cost-benefit analysis for FAIR research data[28] that found the annual cost of not having FAIR research data costs the European economy at least €10.2bn every year. Additional consequences from not having FAIR were also identified which could not be reliably estimated, such as an impact on research quality, economic turnover, or machine readability of research data. By drawing a rough parallel with the European open data economy, the analysis concluded that these unquantified elements could account for another €16bn annually.

Service Offering

The *Federating Core* includes the technical components which enable the federation, access, ordering and delivery of services as well as the management, monitoring and reporting of services. These elements are technical, cultural, policy and resource related by nature and they must be maintained over the long term. Specifically:

- A standard mechanism for naming and locating data and services

[27] Ensuring the Economic Sustainability of Open Data Portals: Understanding Impact and Financing, The Open Data Institute, 2018, https://www.europeandataportal.eu/sites/default/files/s3wp4_sustainability_recommendations_ii.pdf
[28] Cost-Benefit analysis for FAIR research data - Cost of not having FAIR research data, March 2018, DOI 10.2777/02999

- A mechanism for discovery of and access to data and services in the distributed EOSC ecosystem.
- A common framework for managing user identity and access

The *Federating Core* relies on an underlying high-capacity network, as provided by GEANT, that interconnects research, education and innovation communities worldwide that wish to participate in the EOSC.

The *Federating Core* encompasses elements that are needed to allow research targeted services to operate but does not, itself, provide the means to transfer, store, process or preserve research data, which is seen as being performed by the federated research data infrastructures[29]. The portal will present data and services that are accessible with or without user authentication. It is the responsibility of the service provider to register their service so that is becomes visible via the portal.

Financing Model

The *Federating Core* includes a well-defined set of technical services that are non-rivalrous (i.e. they may be used by one researcher without preventing simultaneous use by others) but will need to be scaled-up as the user base for EOSC grows. The *Federating Core* also includes support services (e.g. helpdesks) requiring human expertise.
Its costs are predictable and, with the exception of training and consultancy services, should remain stable as the user base grows.

There are thousands of institutions across Europe and it is not feasible to develop a financial model for the *Federating Core* that depends on the individual agreement of such a large and diverse set of organisations. Consequently, participation in EOSC and contribution to the financing model of the *Federating Core* must be positioned at a national level where governments have a political interest in encouraging open research and the means to define national policies that can support it.

The figure of 1.7 million researchers representing the user base of EOSC as quoted in various documents is derived from Eurostat data[30]. This data estimates the size of the research population by country. Research population size as reported by Eurostat may be considered as a basis for determining the financial contribution to EOSC by participating countries since it is a reasonable measure of the likely usage that will be made of the

[29] This definition of the Federating Core differs from the definition provided in the EOSC-hub 'Briefing Paper EOSC Federating Core Governance and Sustainability' where *shared resources* capable of storing and processing data are included (see section 2.2.3 of the briefing paper)
[30] The figure of 1.7 Million researchers is taken from the *Realising the European Open Science Cloud* report published in 2016 by the Commission High Level Expert Group on the European Open Science Cloud (doi:10.2777/940154) which referenced Eurostat data on Research and Development (R & D) personnel and researchers in FTEs. The original Eurostat URL is no longer available but a more recent URL is here
https://ec.europa.eu/eurostat/statistics-explained/index.php?title=R_%26_D_personnel#Researchers

services provided (i.e. the more researchers there are in a country then the higher the contribution paid by that country). In this manner, the total cost of operating and maintaining the *Federating Core* could be divided between the participating countries according to the relative size of their research populations as defined by Eurostat data.

As a Minimal Viable Product (MVP), the EOSC *Federating Core* limits the risks of failure by not engaging in the curation of research data which remains the responsibility of the participating research data infrastructures. Specifically, should the EOSC MVP fail, no research datasets will be lost.

For a new undertaking, such as the EOSC MVP, engagement based on a proportional membership fee can be a challenge because participating countries are being asked to invest in a resource under development, rather than at maturity, where the business case has already been demonstrated. To overcome this challenge, it is proposed that the risk for participating countries is shared with the European Commission through a funding mechanism that contributes to the membership fees.

The EOSC-hub project estimated it would require 44 FTEs to operate and maintain the *Federating Core* and a further 40 FTEs to provide consultancy, training and support[31].

Legal Vehicle

The Federating Core will provide

- access to data and services for researchers, with or without user authentication;
- a channel through which service providers can publish their service or bundle their service with other offerings, thus multiplying its value;
- research funders insights and knowledge about end-users' needs, service/data uptake and re-use as well as the means to assess return on investments and the impact of policy decisions.

Furthermore,

- the services to be provided by EOSC to end-users are expected to be free at the point of use;
- the total cost of operating and maintaining the *Federated Core* could be divided between the participating countries;
- the *Federating Core* will include a definition of minimum quality standards of service (based on clear Service Level Agreements (SLAs))[32];
- the EOSC will have a governance structure involving EU Member States and the European Commision at an institutional level and a governance board at an executive level.

Based on this, we can infer that the following legislation will apply to the *Federating Core*

[31] Figures taken from the EOSC-hub project document: 'Briefing paper - EOSC Federating Core Governance and Sustainability' (section 4.4)
[32] https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf

GDPR[33]: In order to provide the services listed above, the *Federating Core* will most likely include Personal Data (PD) of researchers, employees of data service providers and employees in research funders. These groups will require personalized accounts to access the *Federating Core*, thus there is a need for holding PD beyond those held by employers and service providers. The *Federating Core* may also hold PD for data stewards, publishers, institutions and NRENs. It should be noted that the consequences of GDPR breach carry high financial penalties. Consequently, it is important that the EOSC governance includes the role of Data Protection Officer (DPO)[34].

Copyright directive and database directive[35]: Both directives have exceptions that may apply to scientific research. However, it must be ensured that the exceptions apply, and the conditions must be clear for all stakeholders. The responsibility of the Federating Core in case of breach of the conditions must be clarified. The copyright directive must be implemented in the member state legal frameworks by April 2021 and national variations may have an impact on the EOSC.

Procurement: Assuming that the host of the Federating Core is a non-private company, it will be subject to the EC Directive 2014/24[36], or equivalents which obligate certain processes to be followed to evidence value for money. If the *Federating Core* will purchase goods or services on behalf of any of its users, consumption is only considered to be compliant where the tender/procurement under which orders are placed has listed/identified those users when it was placed.

VAT: The VAT status of the host of the *Federating Core* must be clarified.

There may be other legislation that applies to (the host of) the *Federating Core*, depending on the services offered and the governance structure. The staff operating and maintaining the *Federating Core* will have to be employed by an organisation. If SLAs are required for offering services to the *Federating Core*, there must be a legal contract with the service providers. This implies there must be a legal basis for dispute resolution when SLAs are breached and raises the question of the appetite of legal entity to pursue litigation measures.

Taking into account the above assessment, we can conclude that there must be a legal entity responsible for the *Federating Core*. However, there are several possibilities for how this legal entity could be implemented:

- Host agreement with an existing public or private organisation (similar to the model used for research projects)

[33] https://ec.europa.eu/info/law/law-topic/data-protection_en
[34] https://gdpr.eu/data-protection-officer/
[35] https://ec.europa.eu/digital-single-market/en/eu-copyright-legislation
[36] https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1415180510261&uri=CELEX:32014L0024

- European Research Infrastructure Consortium - ERIC[37] (used by many ESFRI Research Infrastructures)
- Joint Undertaking (used for the EuroHPC[38])
- AISBL[39] or similar (AISBLs are international non-profit organisations under Belgian law, used by some research infrastructures and organisations governed by several member states)

An existing organisation having a governing model based on its core mission may not easily be adapted to (include) the governing model of the *Federating Core*. New constructions of alternatives ERIC, Joint Undertaking and AISBL above may be created to fit the requirements for governing the *Federating Core* (taking into account the possibilities and limitations given by these constructs). The time required to establish a legal entity may also be a concern (see Timelines).

Additional legal complexity will be added when the EOSC is extended to include *Digital Marketplaces* and the user base in extended to include the public sector, industry and possibly citizens not belonging to any organisation (citizen scientists). Thus, the decision on the type of legal vehicle to be implemented should not be taken on the basis of the *Federating Core* alone.

Governance Structure

Key principles for governance structure:
The key principles for governance developed in preparation for the establishment of EOSC were[40]:

1. **Governance model** - A long-term, sustainable research infrastructure in Europe requires a strong and flexible governance model based on trust and increasing mutuality. As interdisciplinarity is one of the main objectives of the EOSC, the governance model should be based on representability, proportionality, accountability, inclusiveness and transparency.

2. **Governance framework** - The EOSC governance framework will be co-designed, stakeholder-driven and composed of three main layers: a) institutional, including EU Member States and European Commission b) operational, including a governance board and relevant working committees (e.g. thematic and functional) and c) advisory, including a stakeholder forum.

[37] https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/eric_en
[38] https://ec.europa.eu/digital-single-market/en/eurohpc-joint-undertaking
[39] https://justice.belgium.be/fr/themes_et_dossiers/societes_associations_et_fondations/associations/aisbl
[40] https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf

3. **Governance board** - A governance board will coordinate the efforts of stakeholders endorsing the EOSC Declaration, with the broad mandate to reach practical agreements for the implementation of an EOSC Roadmap by 2020. The board will have an advisory role and an implementing role of the decisions by Member States and European Commission concerning the programming, financing and towards the setting up of a long-term governance and business model for the EOSC. It will make best use of the outcomes of past and current projects (e.g. EOSCpilot, eInfraCentral and EOSC-hub) and independent expert advice and studies.

4. **Coordination structure** - A coordination structure, funded by Horizon 2020, will help the governance board to manage the implementation, according to agreed rules and methods of stakeholder participation. The structure and its participating entities should be accountable for the responsibilities assumed, based on an objective assessment of their level of readiness in delivering the EOSC main functionalities.

5. **Global aspects** - The EOSC will be European and open to the world, reaching out over time to relevant global research partners. It will increase the global value of open research data and support stakeholder engagement, including researchers and citizens. 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.

Requirements for the governance structure:

EOSC governance should be “a multi-level and multi-stakeholder governance that ensures a representation for the main stakeholder categories and disciplines, integrating both the national and European levels of authority.” (OSPP Recommendations – Section 2.3)[41].

Ensuring transparency and accountability are essential elements of any business model and must be enforced by the governance framework. Transparency dictates that all participants and providers in the EOSC have the same basic access rights to the marketplace and that actual costs are clear to end users, so that they can make appropriate decisions on the use of the various EOSC capabilities. Transparency also demands open interfaces that are specified for all users[42].

The Governance model will need to address how service providers might need to define their Rules of Participation (RoP) when they may appear in many different services catalogues (e.g. at a national or institutional level) with different RoP whilst remaining “compliant” within the EOSC Portal's presentation of services and other offerings[43].

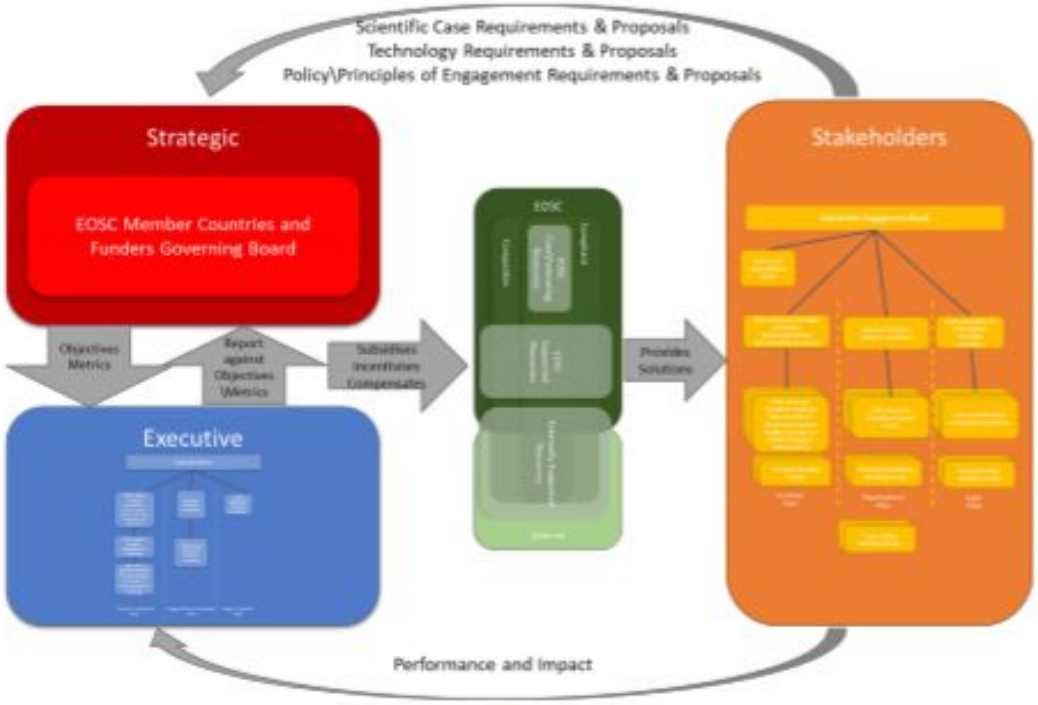
[41] https://eoscpilot.eu/sites/default/files/eoscpilot-d2.2.pdf p.18
[42] https://ec.europa.eu/info/sites/info/files/conferences/eosc_summit_2018/prompting_an_eosc_in_practice_eosc_hleg_interim_report.pdf p.24
[43] https://eoscpilot.eu/sites/default/files/eoscpilot-d2.6-v2.12.pdf p.28

Current approach (2018-2020):

The EOSC Declaration[44] defined three governance layers:

- Institutional – including EU Member States and the European Commission.
- Executive/Operational – including a governance board at the executive level and relevant working committees.
- Advisory/Stakeholder – including a stakeholder forum.

EOSC Pilot document D2.6 outlines potential governance structure: [45]



[46]

The overall decision flow between these layers

- The Stakeholder layer would allow the stakeholders to determine the requirements, policies and Rules of Participation, and make proposals on how these could be met to the Strategic Layer. The Strategic layer would review, agree and prioritise these proposals and requirements to form the strategic vision and objectives of the EOSC.
- The Executive layer would be responsible for ensuring that the EOSC meets this vision and these objectives by: commissioning Core resource as required; commissioning new Supported resources as required; ensuring that Supported resources are properly compensated; and ensuring that the resources within EOSC are both compliant and meet the strategic objectives.

[44] https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf
[45] https://eoscpilot.eu/sites/default/files/eoscpilot-d2.6-v2.12.pdf
[46] https://eoscpilot.eu/sites/default/files/eoscpilot-d2.6-v2.12.pdf p.27

- The Stakeholder layer would also communicate to the Executive how well the EOSC is meeting their requirements at an operational level, and the Executive would report this against the strategic objective to the Strategic layer.

A two phase approach was taken to implement the EOSC. In the first phase (2018-2020) three layer approach was instigated. It consisted of:

- Member State and EC-led Governing Board to supervise the implementation;
- Executive Board composed of 11 members, 8 representing organisations and three appointed as individuals, to ensure effective implementation and accountability;
- Annual Stakeholder Forum events to collect input to feed into the governing bodies.

Additionally, 5 Working Groups of the Executive Board were established to ensure a community-sourced approach to the EOSC priority challenge areas: Landscape, FAIR, Architecture, Rules of Participation and Sustainability. The Executive Board is supported by a Horizon 2020 funded project, EOSC Secretariat.

Phase 2 of EOSCs governance will begin after 2020. Part of the remit of the current Executive Board is to provide recommendations on appropriate mechanisms and possible forms for the future governance of EOSC.

Future approach (beyond 2020) and implications.

There are a number of expectations for the future EOSC governance based on the experiences gained from the current approach, the work of previous bodies in proposing suitable platforms and the EOSC Declaration. These include that there is a strong governance that is flexible, representative and able to respond in a rapid and agile manner. All evidence to date supports the existing three layer model of an institutional/strategic governing board, an executive board and an association of stakeholders.

Based on an expanding EOSC in the future and the early experiences of Phase 1, we can expect the following to impact a sustainable governance model in Phase 2:

Representation

- There is a conflict between the breadth of representation needed in an expanding EOSC and the effective size of the executive arm of governance.
- The rules for joining and leaving each board (executive and institutional/strategic) require careful consideration, including in areas such as minimal terms and membership (eligibility, types etc).
- Are the representation channels for key providers of the federated core appropriate for a mature EOSC (e.g. via the Association of Stakeholders)?
- EOSC should foresee regularly benchmarking activities so that the level of usage and user satisfaction as well as cost effectiveness are regularly reassessed. It is essential that the benchmarking is performed by a third party not engaged in the operation of EOSC

- There are different forms that the Association of Stakeholders could take, including forms that would require their own legal vehicle. Further consideration is required to understand the options available and the merits of each approach.
- The governance structure must include an Ethical and Legal Advisory Board and a Data Protection Officer (DPO).

Resourcing

- The resourcing available to deliver EOSC governance must be realistic. The current rapid pace of development is challenging the resourcing available in the Executive Board and its working groups to deliver a comprehensive and appropriate solution. Consideration should be given to whether such a future body could be constituted from individuals seconded on a voluntary basis or whether the posts would require full time funding.
- The coordination structure, or support project, for EOSC is currently funded through the Commission and Horizon 2020 funding. This is a sensible approach for a body with an evolving governance. A coordination structure will always be required to support the governance of an undertaking of this scale. In the longer term, there should be further investigation into a sustainable mechanism to support this body and consideration given to various funding options for who and how, including in kind contributions, the role of EC funding and member-funded models. In addition, the management overhead of governing the EOSC compared to its total cost must be controlled.

Inclusion

- The ambitions of the EOSC are to be global. In a future situation whereby the EOSC provides access to components from outside its EU members whether by reciprocity or otherwise, how will the governance of these aspects be covered?
- To meet EOSCs global ambitions there may come a time when membership of EOSC is requested from parties beyond the current Member States and Associated Countries. Further consideration is needed over the policy for future membership, types of membership and levels of entities represented. All will have implications on governance, such as the composition of boards, and voting rights and majorities. Participation of non-European countries will have an impact on the applicability of European legislation to the EOSC.

Next steps for planning the Federating Core

This section presents a list of outstanding questions for which further actions/studies are required in order to complete the planning for the implementation of the *Federating Core*.

- The Unique Selling Point of the *Federating Core* need to be further elaborated in terms that appeal to researchers and research performing organisations. In particular, mature and well structured research communities with already

interoperable, federated solutions are invited, in collaboration with research funders, to clearly map how the researchers, especially, the long tail of science of researchers, may benefit from the EOSC without losing the advantages they enjoy today within their aggregated communities.

- The federation of the services and data from the existing infrastructures needs to be planned and depends on the work of other Working Groups such as Landscaping[47]. There are many contributing projects and infrastructures involved representing a major integration activity for which, currently, no-one has responsibility. Will this be a role for the proposal selected for funding via the call INFRAEOSC-03-2020[48]?
- What is the minimum number of confirmed participating countries required to take the decision to establish the EOSC legal entity?
- What happens if a country decides to withdraw from EOSC? The procedures for joining and leaving the EOSC remain to be defined.

- What threat or opportunities could alternative systems pose to EOSC? For example, in 2018 Google launched[49] a dataset search engine called Dataset Search[50] to enable end-users to find datasets stored across the Web through a simple keyword search. The tool searches for information about datasets hosted in thousands of repositories across the Web, making these datasets universally accessible and useful. Will EOSC collaborate with organisations and companies providing such services or will it compete with them?
- The establishment of EOSC What happens if the services of the *Federating Core* were to fail? How would the EOSC ecosystem be recovered? Could a data escrow mechanism provide a business continuity plan? For example, the Registrar Data Escrow[51] program was initiated by ICANN to protect data associated with registered domain names.
- How will EOSC interact with EuroHPC? The European Investment Bank report states "*According to HLEG EOSC, HPC and the EOSC are regarded as two interdependent concepts and shall form future European data infrastructure. As the EOSC is aimed at facilitating access to scientific data in an open format, these significant amounts of data will increasingly require more HPC capacity. Consequently, the provision of more data in the system could fuel the use of HPC and also open up opportunities for innovation and new businesses.*"

[47] https://www.eoscsecretariat.eu/working-groups/landscape-working-group
[48] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/infraeosc-03-2020
[49] https://www.blog.google/products/search/making-it-easier-discover-datasets/
[50] https://toolbox.google.com/datasetsearch
[51] https://www.icann.org/news/announcement-2-2007-11-09-en

Digital Marketplaces

The second iteration extends the *Federating Core* beyond a minimum viable product so that it is capable of supporting multiple digital marketplaces of services and data.

The section below explains the concept of digital marketplaces and explores potential financing models as well as the regulatory and policy environment in order to support such marketplaces. This is followed by more detailed explanations of establishing a succession of marketplaces dedicated to satisfying the needs of identified user groups.

Value proposition

To encourage data re-use and ensure a rich set of services exploiting FAIR data that are available to end-users, the second iteration extends the *Federating Core* so that it can support multiple digital marketplaces.

A digital marketplace builds on the *Federating Core* to provide additional non-collective benefits by offering access to a set of services for an identified membership.

The EOSC-hub project recently documented[52] a demand-side market research to understand the need for and level of demand for digital services for research in the context of EOSC. The market research confirmed that there is a growing demand for digital services for research. From all the services that are being used by researchers, the most popular are data repositories and data registries, followed by a large variety of other services (e.g. analytics, data management tools, collaboration services and general computational services). Among the challenges of using digital services, researchers raised ease of use and ease of access as main areas for improvement (e.g., availability of a single sign-on facility). Also, the configuration of digital services to be used with own data sources was a major concern, which indicates challenges with the interoperability and integration of services.
Researchers also expect support for data management planning with trusted repositories, collaboration tools, together with training and support, when necessary.
The introduction of digital marketplaces offers *freedom of choice* to end-users so that they can select the most appropriate services for their research.

In the case of researchers, such a digital marketplace may offer services that store, preserve or add value to data (e.g., derived data, tools or analysis) through the payment of fees. Such services would support different aspects of the research activities including those necessary to implement their data management plans. The services would have been tested to verify that they conform with the Rules of Participation that are defined for EOSC service providers.

[52] EOSC-hub D12.1 deliverable, Procurement requirements and demand assessment, June 2019, https://www.eosc-hub.eu/deliverable/d121-procurement-requirements-and-demand-assessment

For service providers, participation in a digital marketplace offers the opportunity to:

- Enlarge market, visibility and broaden user base thanks to service interoperability.
- Increase service value through interconnection of data and services.
- Improve end-user experience through a federated offer.
- Provide a revenue channel via fee based access to services.

Service Offering

A digital marketplace builds on the *Federating Core* to offer a range of services.
A digital marketplace includes data and services that are accessible against some form of payment from both publicly funded and commercial providers. It is expected that the rivalrous services, such as those that store or preserve research data as well as those than compute against it, will be made available against payment via a marketplace.

The list of digital marketplaces registered with EOSC will be accessible via the portal, the number of marketplaces and the volume of activity they generate will be important measures of the impact of EOSC.

While a single marketplace could handle the volume of activity expected for EOSC, the differences in the regulation and policies applicable to the diverse range of service providers and user groups engaged with EOSC (as explained later in this section), suggests it more appropriate to handle them separately with different marketplace operators. EOSC could create an environment in which multiple digital marketplace operators develop their own business models and attract the participation of service providers and users.
The article '*The Organization of Digital Marketplaces: Unmasking the Role of Internet Platforms in the Sharing Economy*'[53] highlights the important role of a market operator[54] and five key organisational elements necessary for digital marketplaces:

- Membership - market operator requires members (sellers and buyers) to join the marketplace
- Rules - market operator decides rules which members have to comply with. These include the types of services that can be offered as well as the general conditions for marketplace transactions
- Monitoring - the market operator monitors members' compliance with marketplace rules
- Sanctions - the market operator can sanction members to enforce marketplace rules
- Hierarchy - the market operator exercises the right to make binding decisions

Note that participation in a digital marketplace as a service provider or user is voluntary.

[53] Kirchner, S.; Schüßler, E. (2018, forthcoming) The Organization of Digital Marketplaces. In: G. Ahrne & N. Brunsson (Eds.), Organization outside organization, Cambridge: Cambridge University Press.
[54] The market operator facilitates market transactions by creating and operating a market order for external sellers and buyer

The focus of this iteration is to extend the *Federating Core* to include functionality to support these key organisational elements that will make it simple for market operators to create digital marketplaces. For example, extending the common identity framework to support marketplace membership, extending the service and data monitoring functions to simplify verification of compliance and enable billing, etc. It will also be necessary to define specific rules of participation for organizations that wish to take on the role of a market operator. Note that is likely that the same service or dataset may be made available by a service provider in multiple digital marketplaces with differing terms and conditions (e.g. more stringent SLAs) or pricing models (e.g. academic vs. commercial license fees).

Financing Model

The financial model for Digital Marketplaces assumes the *Federating Core* is operational and sustained through its financing model. The Digital Marketplaces financial model addresses those aspects concerned with extending the *Federating Core* and registering and maintaining a list of digital marketplaces.

The funding required to support a single digital marketplace were included in the estimate made by the EOSC-hub project. The operational cost of extending the *Federating Core* to support multiple marketplaces could be addressed by participation fees for market operators including a one-time registration fee and a recurring annual fee based on the volume of transactions executed by a digital marketplace.

The financial model for operating an individual digital marketplace and remunerating the service providers is the responsibility of the market operator.

Regulatory and Policy Environment

Digital Marketplaces introduce procurement and financial transactions into the EOSC ecosystem. EOSCsecretariat WP5 has itemised a series of factors influencing the design of the EOSC business model once procurement is introduced which need to be considered[55]. This will impact the obligations and risks for market operators, service providers and service consumers.

A digital marketplace may be considered as an online platform subject to the EU Regulation on platform-to-business relations that was adopted on 20 June 2019 aimed at creating a fair, transparent and predictable business environment for smaller businesses and traders when using online platforms[56].

[55] Innovative business models for the EOSC - background material, EOSC secretariat, 2nd August 2019
[56] https://ec.europa.eu/digital-single-market/en/business-business-trading-practices

The EOSC-hub project documented[57] a business model analysis in the EOSC context, with the assumption that the end-user always operates in a business-to-business (B2B) type of relationship and not as a consumer (as defined by the EU Consumer Rights Directive) for personal purposes. This assumption is important, as it rules out the need to comply with consumer rights regulation.

As the breadth of services offered via digital marketplaces increases, there will be sector-by-sector differences in the regulatory and policy environment. For example, when artificial intelligence services are made available it will be necessary to take into account the Ethics guidelines for trustworthy AI[58].

The selection of market operators must be done through an open and transparent process. The Cloudwatch project report '*Roadmap to a cloud market structure encouraging transparent cloud pricing*'[59] describes how to manage systemic risks and shape policies for a fair and transparent cloud market.

The regulations of the Public Procurement Directive may apply, in particular if incentives are needed to establish a marketplace. This could be the case for the *Digital Marketplaces* serving publicly funded researchers. The host of the *Federating Core* must establish contracts and Service Level Agreements with the market operators.

Digital Marketplaces for publicly funded researchers

GEANT[60] operates a digital marketplace[61] with 36 NRENs to offer a range of Infrastructure as a Service (IaaS) cloud computing solutions from a variety of commercial providers[62] for academia. Since 2016, approximately 12M€ of procurement has been facilitated via the marketplace. This digital marketplace will be extended and replenished via the tender to be published during 2019 by the OCRE project[63]. The costs of developing and operating this marketplace are currently covered by EC project funding[64].

A prototype marketplace for researchers has also been developed by EOSC-hub[65] and is primarily stocked with services supported by publicly funded organisations participating in the EOSC-hub project. The marketplace is being used as a vehicle to assess the applicability of the Virtual Access[66] financial instrument as a means to reimburse the access

[57] EOSC-hub D12.1 deliverable, Procurement requirements and demand assessment, June 2019, https://www.eosc-hub.eu/deliverable/d121-procurement-requirements-and-demand-assessment
[58] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
[59] http://www.cloudwatchhub.eu/Roadmap-for-transparent-pricing
[60] https://clouds.geant.org/support-for-institutions/
[61] https://clouds.geant.org/support-for-institutions/
[62] https://clouds.geant.org/geant-cloud-catalogue/geant-cloud-catalogue-iaas/
[63] https://www.ocre-project.eu/sign-tender
[64] Marketplace funded via GN4-2 then GN4-3 and OCRE projects.
[65] https://marketplace.eosc-portal.eu/
[66] The Virtual Access (VA) instrument is provided by the European Commission to increase the sharing of research infrastructures and services that otherwise would not be available to international user groups. In VA, the services – also called "installations" – have to be made available 'free of

provisioning costs to service providers. This instrument is provided by the European Commission to increase the sharing of research infrastructures and services that otherwise would not be available to international user groups.

In the USA, the NSF[67] is funding the creation of CloudBank[68] as a digital marketplace serving the research community. CloudBank will provide financial engineering options that will give researchers more flexible cloud terms for multiple cloud vendors tailored for their needs and contribute to the sustainability of CloudBank operations. It will be complemented by a cloud usage monitoring system that gives NSF-funded researchers the ability to grant permissions to research group members and students, set spending limits, and recover unused cloud credits.

Funding Model

Below is an example financing model for an individual digital marketplace serving the needs of publicly funded researchers. It assumes all services provide fee-based access and service providers are free to define the pricing model for their services which would be visible to end-users via the marketplace. The financing model for the operation of the marketplace itself would be based on a combination of a registration charge to be paid by the service provider for each paying service when it enters the marketplace and a transaction charge paid by the service provider based on the total volume of consumption of the service.

The funds to pay for researchers' use of fee-based services may come from multiple sources including:

- Research grants at national or research funder level. Many research funders now expect the data from funded projects to be deposited and preserved in an open access manner to ensure potential future re-use and to ensure research validation and integrity. The built-in costs at source model, used by UK research grant agencies and elsewhere, requires that archiving costs are factored into the original grant application by the researcher for the services they consume. This implies that the researcher and/or their institute would consult service offerings in the marketplace to estimate the price they would have to pay for the services they require before submitting their grant request. Research funders at a European level, such as the European Research Council (ERC)[69], Marie Sklodowska Curie[70] and European Cooperation in Science and Technology (COST)[71], could also be engaged. For example, costs for the deposit of research data in an open access data repository are considered an eligible cost of ERC grants. Similarly, costs associated with open

charge at the point of use' for European or International researchers. VA access is open and free access to services through communication networks to resources needed for research, without selecting the researchers to whom access is provided.

[67] https://nsf.gov/awardsearch/showAward?AWD_ID=1925001&HistoricalAwards=false
[68] https://cloudbank.org/
[69] https://erc.europa.eu/
[70] https://ec.europa.eu/research/mariecurieactions/node_en
[71] https://www.cost.eu/

access to research data, can be claimed as eligible costs of any Horizon 2020 grant[72]. If this eligibility included all the services in the digital marketplace for publicly funded researchers then researchers could use a fraction of their grants to fund their needs, and the overhead of planning subscriptions to services would be limited.

- Centralised funding model whereby multiple countries or agencies collectively fund access to services. The access to the services would be excellence-driven and could be organised as a succession of calls in a similar manner to that organised by PRACE for access to HPC resources[73].

Note that payment for service consumption by a user may contribute to recuperating operational costs of a service but it is unlikely to cover the development costs of a service available via a marketplace.

Similarly, a means of remunerating public sector service providers for consumption may contribute to the optimisation of sharing of existing public sector capacity across participating countries and communities, but significant additional capacity will be required to cope with the demands of 1.7 million researchers. This is an area where private sector investment can potentially contribute by financing the development of innovative services to be made available via a marketplace. With the existence of functioning digital marketplaces and a growing user base, the business case for private sector service providers to engage becomes more convincing. Joint investment instruments, such as Pre-Commercial Procurement (PCP)[74] and Public Procurement of Innovative solutions (PPI)[75], have been shown to reduce the R&D risk for procurers and vendors in the commercialisation of innovative services and products. The ARCHIVER project[76], building on the lessons learnt from the Helix Nebula Science Cloud[77], is pursuing the pre-commercial procurement of archiving and preservation services with the intention of introducing them in the EOSC context. The Pre-Commercial Procurement financial instrument is also being employed by the contributing research infrastructures that will connect to the EOSC[78].

The costs and overheads of operating a marketplace and procuring services can be reduced by providing a procurement service through aggregated procurement as recommended by

---

72 https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

73 http://www.prace-ri.eu/excellence-in-science-drives-prace-16th-call-for-project-access/

74 https://ec.europa.eu/digital-single-market/en/news/impacts-eu-funded-pre-commercial-procurements

75 https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/innovation-procurement_en.htm

76 https://www.archiver-project.eu/

77 Integrating commercial cloud services into the European Open Science Cloud, March 2019, https://doi.org/10.5281/zenodo.2598039

78 https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/infrainnov-04-2020

the EOSCpilot project[79]. Unlike PCP and PPI, such a procurement service would be targeted at already commercialised services that can be procured by individual research funders or collectively as buyer groups via a framework agreement model. The OCRE project[80] is preparing a tender to procure general cloud services and earth observation services in the context of EOSC. The tender specifications for commodity services as well as those resulting from PCP & PPI instruments can including criteria to ensure the EOSC Rules of Participation for service providers are respected. The economies of scale provided by EOSC to such procurement activities will encourage preferable terms and conditions for EOSC users.
The EOSC-hub project is working closely with the ARCHIVER and OCRE projects and has recommended[81] that the 'Cloud Coin' or Voucher as well as 'Sponsored Use' market-driven access models are supported by EOSC.

Digital Marketplaces for the public sector

The set of *Digital Marketplaces* building on the *Federating Core* can be expanded with additional marketplaces dedicated to the requirements of end-users from the public sector[82] who are not involved in research activities but want to exploit open access data.

As stated in the POPSIS report[83]

> "*The public sector collects, creates, produces and disseminates a wide variety of information ranging from legal and administrative information, business and economic data, to geographic and meteorological information. Public sector information (PSI) constitutes a valuable raw material which can be re-used by third parties in added-value information products and services*."

Digital marketplaces dedicated to the public sector exist in some member states. For example, the UK Digital Marketplace[84] is transforming the way the UK government buys technology and digital services by opening the market up to small and medium-sized enterprise (SME) suppliers. A total of £3.2 billion has been spent through the UK Digital Marketplace in just under 6 years. Of that total, 48% is spent with SMEs.

Being able to link such public sector digital marketplaces to EOSC would:

---

[79] EOSCpilot White Paper 3: Procurement Supporting Document to D3.3 Draft Policy Recommendations
https://eoscpilot.eu/sites/default/files/eoscpilot_d3.3_whitepaper_3_procurement.pdf
[80] https://www.ocre-project.eu/
[81] EOSC-hub D12.1 deliverable, Procurement requirements and demand assessment, June 2019, https://www.eosc-hub.eu/deliverable/d121-procurement-requirements-and-demand-assessment
[82] In this document the term 'public sector' refers to all bodies governed by public law as defined in Public procurement of services: Council Directive 92/50/EEC
[83] Pricing Of Public Sector Information Study (POPSIS) Models of Supply and Charging for Public Sector Information (ABC), Final Report, October 2011,
https://ec.europa.eu/digital-single-market/en/news/pricing-public-sector-information-study-popsis-models-supply-and-charging-public-sector
[84] https://www.digitalmarketplace.service.gov.uk/

- Increase re-use of FAIR research data
- Increase the range of services available to EOSC users

The European Data Portal[85] harvests the metadata of Public Sector Information available on public data portals across European countries. Information regarding the provision of data and the benefits of re-using data is also included. Currently the portal contains references to almost 900,000 datasets contained in 78 catalogues from 34 countries. It is also harvesting the metadata of European Union Institutions and bodies disseminated via the EU Open Data Portal[86], as well as EUMETSAT and ESA data collections. It is available in 24 EU languages. Activities supporting data production and data take-up have also taken place (e.g. training material and guidelines for data producers, studies on Creating Value through Open Data and on Open Data Maturity in Europe). The software code of the portal and all its modules has been made freely available as Open Source Software. The European Data Portal is hosted by DG DIGIT and funded by the Connecting Europe Facility (CEF) Telecommunications sector.

Establishing a connection between EOSC and the European Data Portal would make it simpler to combine public data published by the EU institutions and agencies with FAIR research data. Beyond providing a working example of how the EOSC portal could develop, the European Data Portal provides insights into the cost of operational and long-term sustainability of Open Data portals.

Financing Model

To accommodate a wider public sector user base which may have an impact on the *Federating Core*, the underlying membership fees for participating countries may need to be revised to take into consideration not only the size of a country's publicly funded research population but that of the wider public sector.

Digital Marketplaces for industry

The set of *Digital Marketplaces* building on the *Federating Core* can be expanded with additional marketplaces dedicated to the requirements of end-users from the private sector so that they can exploit the open data and associated services for commercial gain without distorting market competition.

Value-proposition

- Access to Open Science data, publications and software;
- Facilitated knowledge and technical skills transfer;

[85] https://www.europeandataportal.eu/en/homepage
[86] http://data.europa.eu/euodp

- New business opportunities on top of open data and services

The European Investment Bank report included a study of the value of the EOSC to the private sector, and noted:

> "*Being bold and visionary, the European Open Science Cloud initiative could provide a substantive positive impulse for the further development of HPC and Cloud Computing in Europe.*
>
> *So far, however, the EOSC initiative is little known and understood by private businesses. Hands-on use cases and dedicated content marketing activities are necessary to stimulate substantive engagement by commercial players in the EOSC.*
>
> *Access to relevant, commercially interesting data is needed to engage businesses in the EOSC. In contrast to that, there is little need for additional hardware capacity in the market.*
>
> *In order to become relevant for commercial activity, industry data security standards and service levels need to be met*."

The Big Data Value Association published a paper in April 2019 *Towards a European Data Sharing Space - Enabling data exchange and unlocking AI potential*[87] in which the EOSC is mentioned as a 'Prominent Data Sharing Ecosystem for Science'. The paper also outlines the role that next-generation data platforms, as data-driven complex systems, perform in driving the European Data Economy and enabling full adoption by both business and private consumers' markets, while implementing EU policies and upholding values regarding democracy, privacy protection and equality. It also identifies opportunities for science with potential benefits for academia including:

> *Monetisation opportunities brought about by emerging data-driven business models. Providing controlled access to research data will enable scientists, universities and research institutes to have the opportunity to exchange or monetise their research data by making it available in a controlled way to other institutes and companies. This will strengthen cooperation between research and industry, enable different data to be integrated and analysed and thus introduce new revenue opportunities for academia*.

As stated in the European Data Market Study report[88]:

> *"the real game changer is the connection between Big Data and Artificial Intelligence. Big Data is the fuel powering the emerging AI innovation wave, driven by breakthroughs in machine learning and deep learning technologies. Without massive*

[87] http://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpace%20PositionPaper_April2019_V1.pdf
[88] Data as the Engine of Europe's Digital Future: Second Report on Policy Conclusions, http://datalandscape.eu/study-reports/data-engine-europes-digital-future-second-report-policy-conclusions

datasets to train neural networks, the current generation of AI systems and applications would not exist."

Providing access to FAIR research data via EOSC can help form research data markets of value to business organisations. Building a European data economy is part of the Digital Single Market strategy[89]. The initiative aims at enabling the best possible use of the potential of digital data to benefit the economy and society. Some EU Member States have already started to specify and/or even establish national data markets, either through the involvement of policy makers or driven by the industry. Beside national data markets providing national data infrastructure, industry specific data markets appear providing data and services for specific industries.

In Austria there is Data Market Austria[90], in Germany the Industrial Data Space[91], in France Dawex[92], in Italy SpazioDati[93], in Switzerland the Swiss Data Alliance[94] has just been established. In the case of Dawex, it is an IT start-up company providing a data marketplace where organizations meet, buy and sell data, directly and securely. The company makes profit from the data transactions between data suppliers and users – it can request a per-transaction percentage commission, a subscription, or a set of variable fees based on optional services. Since its foundation, Dawex has progressively enlarged its data offer in distinct sectors and recruited 2,000+ companies on board of their platform, of which 45 % are based in Europe (with a majority of them in France), 38 % in the United States, and the remaining 17 % in other countries (mostly based in Asia).

Financing Model

The European Data Market Study report states:

"While in the early phase of development of the data market enterprises focused on the collection and management of data, now business users prioritize strongly the extraction of value from data, which includes also data monetization."

The report goes on to identify three primary paths towards realizing that value for data:

1. Direct revenue from data sale/licensing;
2. Additional revenue from bundling data with other services or products;
3. Exchange premiums/trade advantages or discounts.

But also states that a clear-cut form of data monetisation is hard to find and that data exchange resulting in the creation of straight revenue streams for data holders as a result of a direct sale (Type 1), or in the form of additional revenue in conjunction with the offering of

[89] https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy
[90] Data Market Austria https://datamarket.at/
[91] Industrial Data Space http://www.industrialdataspace.de
[92] Dawex https://www.dawex.com/en/
[93] SpazioDati https://spaziodati.eu/en/
[94] Swiss Data Alliance https://www.swissdataalliance.ch/

other services (Type 2), does not appear to be very common in Europe today. Indeed, the case studies cited in the report reveal the prominence of indirect benefits such as operational efficiency, cost optimization, enhanced quality obtained through the data sharing rather than the generation of direct revenue streams.

To create a data marketplace requires[95]:

- A way to make the data marketplace as open as possible — to source as much data as possible — yet protect buyers and sellers from bad actors.
- A way to guarantee data sellers get paid — and get paid the right amount — each time their data is purchased.
- A fast, secure and scalable micropayment infrastructure to allow freedom of use: buyers should only have to pay for the data they consume.
- A way to guarantee data provenance to ensure purchased data hasn't been tampered with and actually comes from the alleged seller.

In a marketplace setting, buyers only pay for the data they consume, this effectively creates a strong incentive for data providers to offer the highest-quality, most sought-after data possible as this maximizes their revenue. Similarly, the infrastructure and inherent data exchange standard created by the marketplace removes friction between data buyers and sellers which lowers the cost of acquiring data. By allowing data providers to set their own data prices and enabling data consumers to choose who they purchase data from, not only do data marketplaces allow consumers to signal which data/sellers provide value, but they also solve the data pricing conundrum by taking a free market approach.

Suggested business models for the co-existence of freely available open data and fee-based access[96] include:

- **Metered consumption**. In this model, a certain amount of data is free, and when you need more you must pay for it. Within this model there are a couple of options: one where rates are measured on a per-dataset basis, and one where rates are measured site-wide, regardless of the dataset(s) being accessed. Although they don't use it for payments, one of the major open data platforms already has this capability built in, and it's likely others do as well. An example is Google Maps[97].
- **Consumer classification**. Charging different rates for various types of customers is a well-known business model. For publicly funded research data, business consumers are most likely to be charged, since they are probably using the data for revenue-generating purposes. Similar to other freemium models, the business consumers are paying customers that effectively subsidize free services for non-business customers.
- **Premium datasets**. Similar to consumer classification, except instead of charging based upon who is accessing the data, costs are determined by what data they

[95] Based on material presented in the article Data Marketplaces: The Holy Grail of our Information Age, Jeremiah Smith, July 24th 2018,
[96] Extracted from an article by Andrew Nicklin, 'How free open government data can coexist with fee-based access', 19 April 2017, https://medium.com/@technickle/how-free-open-government-data-can-coexist-with-fee-based-access-7e2719be1199
[97] https://developers.google.com/maps/premium/

access. For example, data about crimes might be free, but data about licensed business owners might not be.

- **Premium data columns**. A more refined version of premium datasets, this approach allows a set of records to be freely available to anyone, but provides more record details to those who are willing to pay. For example, a free version of real property sales data might contain an address, property type, boundaries, assessed value, and transfer date, but the premium version might contain buyer/seller details, sale amount, taxes paid, lien information, and so on.
- **Data exchange**. A data provider supplies data to a third party in exchange for the ability to consume some other data from them. The Waze Connected Citizens[98] programme is an example.
- **Access methods**. Bulk data could be made available for free while application programming interfaces (APIs) could be fee-based. Charging to push data to a subscriber as it becomes available (rather than pulling it using bulk downloads or APIs) is another option.
- **Raw or processed data**. This isn't about summaries, statistics, or visuals. Instead, it is the distinction between a deluge of data from large numbers of sensors and the observations that may result. For example, electric meters continuously measure the rate of power consumption, but that data might be processed to create new records reflecting daily usage, or simply events like an unusual change in consumption. This raw data generally requires more investment to process, so it may be reasonable to expect consumers to pay for access to it as part of that investment.
- **Realtime or delayed access**. This scenario is useful for transactional data where extremely recent data has greater value, but value decreases over time until it's free. An example from the finance sector is stock market data, where companies pay a premium for real time information (and invest millions of dollars to gain a millisecond lead over the competition), but within 15–20 minutes the data is free and publicly available.

Publicly-operated data markets have potential benefits beyond making FAIR data accessible. With this approach, a government offers a public data market as a platform on which it and third-parties make a variety of data available, some for free and some at a premium. Because they are operating it, the government gains the ability to apply taxes or fees to data-access transactions (and this could be through any or all of the models suggested above), but it also gets an opportunity to regulate the market itself by establishing ground rules to protect privacy, public interest, and so on. Smart Copenhagen[99] is an example.

Developments in the area of smart contracts[100] are relevant for the support of trusted data marketplaces providing secure transactions, data integrity and quality.

---

[98] https://www.waze.com/ccp

[99] City Data Exchange – Lessons Learned From A Public/Private Data Collaboration, March 2018, https://cphsolutionslab.dk/content/2-what-we-do/3-data-platforms/3-city-data-exchange/1-learnings-from-the-city-data-exchange-project/city-data-exchange-cde-lessons-learned-from-a-public-private-data-collaboration.pdf?1527149474

[100] https://en.wikipedia.org/wiki/Smart_contract

Regulatory and Policy Environment

The exceptions for scientific research in the Copyright directive and Database directive does not apply to industrial or public use of the data. Thus, compliance with these two directives must be ensured.

All funders of the EOSC have a responsibility to comply with state aid regulations and they need to ensure that the beneficiaries comply as well. Accordingly, the host of the *Federating Core* will have responsibility and must ensure that the terms and conditions for market operators and service providers are compliant with the state aid regulations.

Next steps for the planning of Digital Marketplaces

This section presents a list of outstanding questions for which further actions/studies are required in order to complete the planning for the implementation of digital marketplaces:

- The means of remunerating publicly funded service providers requires further investigation. The EOSC-hub '*Briefing Paper - EOSC Federating Core Governance and Sustainability*' highlights that a mechanism to compensate publicly funded service providers is needed but the existing Virtual Access (VA) scheme is not suitable for rivalrous services.

- The introduction of the role of operator of a digital marketplace may require some additions to the EOSC Rules of Participation. A registration process with a formal agreement would probably be required for market operators in a similar style to ICANN's Registrar Accreditation[101].

- What happens if a digital marketplace fails? The collapse of a marketplace can have important consequences for service providers and users. There is a risk of loss of data and knowledge which must be taken into account. An exit strategy for market operators could be a formal requirement as part of the Rules of Participation. While it is assumed the the market operator has financial and legal responsibility for the operation of its marketplace, what can EOSC do to minimise the impact of closure on the marketplace members? Providing a business continuity plan for data management services that are no longer available via EOSC would bring an added-value by increasing end-user trust in the whole ecosystem. This is an area where data preservation services, such as those being tendered via the ARCHIVER project, could provide an essential stakeholder protection mechanism.

[101] https://www.icann.org/resources/pages/accreditation-2012-02-25-en

- We were unable to identify more detailed information concerning the extension of the EOSC user base to include the public sector and industry. A more detailed requirements analysis would be helpful to determine the needs and added value of EOSC in these contexts. For the private sector, lessons can be drawn from the Copernicus programme which offers data and information services access[102]. It is also proposed to also collaborate with the Big Data Value Association to explore the most effective means of engaging the private sector as users of the open data and associated services that will be made available via the EOSC.

Further considerations

Timelines

Section 4 'From Launch To Sustainability (2019-2020)' of the EOSC Strategic Implementation plan states that the initial EOSC *Federating Core* should be in place by the end of 2019 and that the connection of most infrastructures and services to the EOSC should be made by the end of the 2nd quarter of 2020. So the end of the 2nd quarter of 2020 would correspond to the transition from *Federating Core* to *Digital Marketplaces*. The decision to make this transition would depend on the proven accomplishment of many steps, including:

- Have the ESFRI cluster projects[103] been able to register and promote their FAIR datasets using the *Federating Core* services?
- Have independent end-users[104] been able to search for, find, access and exploit such datasets?
- Have candidate market operators been identified?
- Has a legal entity for EOSC been established that can engage with market operators?

Managing Complexity

More than 25 EC funded projects at different stages of their lifecycles and with their own set of objectives and key exploitable results are contributing to the EOSC implementation and have been invited to the joint EOSC project meeting scheduled for 9-10 September 2019 in Brussels. The EOSC Executive Board has the task of producing "*a strategic implementation plan and annual work plans, and of a proposed mechanism for overseeing and steering the implementation of the strategic and annual work plans, and for monitoring and reporting on progress*"[105].

The scale and complexity of bringing together all contributors (not just the EC funded projects mentioned above) to produce annual work plans and reporting represents a full-time

[102] https://www.copernicus.eu/en/access-data
[103] https://www.eosc-portal.eu/news/five-new-esfri-cluster-projects-eosc-panorama
[104] End-users not employed by beneficiaries of the EOSC-hub and ESFRI cluster projects
[105] https://www.eosc-portal.eu/governance/executive-board

activity. In order to support the work of the Executive Board, we recommend that the EOSCsecretariat project is requested to commission, via the co-creation budget, the engagement of a professional entity that will use a recognised project management method [106] to produce all the necessary deliverables for the planning phase of EOSC. This activity should start in 2019 and produce a first version by the end of the first quarter of 2020.

[106] Such as OpenPM[2]
https://publications.europa.eu/en/publication-detail/-/publication/ac3e118a-cb6e-11e8-9424-01aa75ed71a1/language-en/format-PDF/source-83307127