European Commission

PROMPTING AN EOSC IN PRACTICE

Research and innovation

EUROPEAN COMMISSION

Prompting an EOSC in practice

Interim report and recommendations of the Commission 2nd High Level Expert Group [2017-2018] on the European Open Science Cloud (EOSC)

DRAFT

Members: Isabel Campos Plasencia, Andreas Mortensen, Silvana Muscella (Chair), Toivo Räim, François Robida, Linda Strick, Klaus Tochtermann, Žiga Turk, George A. Komatsoulis (Observer, USA), Ross Wilkinson (Observer, Australia)

2018

Directorate-General for Research and Innovation
Directorate A – Policy Development and Coordination

Prompting an EOSC in practice

European Commission

Directorate - General for Research and Innovation

Directorate A – Policy Development and Coordination

Unit A2 – Open Data Policy and Science Cloud

Contact Maarja Adojaan

E-mail maarja.adojaan@ec.europa.eu

RTD-PUBLICATIONS@ec.europa.eu

European Commission

B-1049 Brussels

Manuscript completed in June 2018.

Luxembourg: Publications Office of the European Union, 2018

Acknowledgements

The authors are grateful to many experts for their input and private conversations. In particular, they would like to thank:

Carmela Asero	Former Policy analyst at the European Commission - DG Research and Innovation (DG RTD)
Alberto Azevedo	Oceanographer at National Laboratory for Civil Engineering (LNEC) for OPENCoastS
Paolo Balboni	Founding Partner of ICT Legal Consulting and Professor of Privacy, Cybersecurity, and IT Contract Law at the European Centre on Privacy and Cybersecurity within the Maastricht University Faculty of Law
Juan Bicarregui	Head of the Data Services division in the Scientific Computing Department at Science & Technology Facilities Council (STFC)
Paolo Budroni	Head of Department of e-Infrastructure at Vienna University Library and National Delegate at e-IRG
Donatella Castelli	Senior Researcher Consiglio Nazionale delle Ricerche (CNR-ISTI)
Ron Dekker	Chief Executive Officer and legal representative of Consortium of European Social Science Data Archives (CESSDA)
Matthew Dovey	Head of e-Infrastructure Strategy at JISC
Willem Elbers	Software Engineer / Infrastructure Specialist for CLARIN
Tiziana Ferrari	Technical Director at the EGI Foundation and Coordinator of the EOSC-hub project
Jorge Gomes	Jorge Gomes researcher at Laboratory of Instrumentation and Particle Physics (LIP) for OPENCoastS
Silvia Gomez Recio	Secretary General at Young European Research Universities Network (YERUN)
Annabel Grant	Senior Business Development Officer at GÉANT
Simon Hodson	Executive Director of CODATA
Robert Jones	Responsible for Externally Funded Projects at CERN IT Department and Coordinator of Helix Nebula initiative
Sarah Jones	Associate Director of Digital Curation Centre (DCC)
Hermann Lederer	Deputy director, Max Planck Computing & Data Facility
Damien Lecarpentier	Project Director at CSC– IT Centre for Science
Jesús Marco	Research Professor at Instituto de Física de Cantabria (IFCA)
Eva Mendes	Representative of Young European Research Universities Network (YERUN), Deputy Vice-President for Strategy and Digital Education of Universidad Carlos III de Madrid
Natalia Monola	Research Associate in the University of Athens
Barend Mons	Professor in Biosemantics at the Leiden University Medical Center (LUMC) and also a board member of the Leiden Centre of Data Science, Director of the International Support and Coordination office of GO FAIR
Steven Newhouse	Head of Technical Services at European Bioinformatics Institute for ELIXIR
Gareth O'Neill	President of Eurodoc
Anabela Oliveira	Head of Division at National Laboratory for Civil Engineering (LNEC) for OPENCoastS
Per Oster	Director of the Research Infrastructures unit at CSC and Project Director of the EOSC-Hub project
Federico Ruggeri	Director Consortium GARR
Gergely Sipos	Customer and Technical Outreach Manager at Stichting EGI for ELIXIR
Dieter Van Uytvanck	Technical Director of CLARIN ERIC
Susheel Varma	Technical Coordinator at European Bioinformatics Institute for ELIXIR

And, at the European Commission, for their continued guidance and support:
Maarja Adojaan, Emanuele Barbarosa, Patrick Brenier, Jean-Claude Burgelman, Athanasios Karalopoulos, Wainer Lusoli, Corina Pascu, Katarzyna Szkuta, Victoria Tsoukala.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Item	Description
EDI	European Data Infrastructure
eIDAS	electronic Identification, Authentication and trust Services
EIROforum	Combines the resources, facilities and expertise of its member organisations to support European science in reaching its full potential
EOSC	European Open Science Cloud
EOSC Portal	The term applied for the benefit of the current report to imply the EOSC is a common portal giving consolidated access to existing e-infrastructures
ERAC	European Research Area Committee
ESFRI	European Strategy Forum on Research Infrastructures
EURODOC	The European council of Doctoral Candidates and Junior Researchers
FAIR	Findable, Accessible, Interoperable & Reusable
GO FAIR	A bottom-up international approach. for the practical implementation of the European Open Science Cloud (EOSC) as part of a global Internet of FAIR Data & Services
HLEG	High-Level Expert Group
IaaS	Infrastructure as a Service
ICANN	Internet Corporation for Assigned Names and Numbers
IETF	Internet Engineering Task Force
LERU	League of European Research Universities
MVE	Minimum Viable Ecosystem
OSPP	Open Science Policy Platform
PaaS	Platform as a Service
RDA	Research Data Alliance
SaaS	Platform as a Service
SWD	Staff Working Document
WG	Working Group
YERUN	Young European Research Universities

EXECUTIVE SUMMARY

"Science is the captain and practice the soldiers".

Leonardo Da Vinci

With the Staff Working Document (SWD) on the EOSC at its very heart, this report picks up from the recommendations of the first EOSC HLEG report with the aim of providing truly practical considerations and pointers for the timely implementation of the EOSC, based around the concept of a "Minimal Viable Product" i.e. a product with just enough features to satisfy early customers, and to provide feedback for future product development. It also highlights the ground-breaking work taking place in Europe, guided by the increasingly prevalent principles of research data-sharing.

The Commission presented its vision for the European Open Science Cloud (EOSC) in its April 2016 Communication on the 'European Cloud Initiative', [1] as a part of the Digital Single Market Strategy. This will be attained through policy action and financial support to **integrate and consolidate e-infrastructure platforms**, to **federate existing research infrastructures** and scientific clouds and to support the **development of cloud-based services** for Open Science. In close collaboration with Member States to **connect the priority European research infrastructures** to the European Open Science Cloud, the Commission will also work towards an **Action Plan for scientific data interoperability**, including 'meta-data', specifications and certification.

To help drive forward and implement the EOSC, the main thread of the report is to understand how the EOSC can effectively interlink **People, Data, Services and Training, Publications, Projects and Organisations**. Formation of these bonds would not only produce valuable data, but also provide visibility and networking space, with obvious incentive mechanisms for the recognition of the work of scientists and the value of supporting infrastructures. The EOSC should be a user-friendly, collaborative tool for data sharing and re-use.

The group took stock of work that is underway in Europe. Existing use cases could work as **common service working models** illustrating what the European scientific community is already achieving. The latter, indeed, form a sample of twelve **EOSC in Practice Stories** that accompany this report found in ANNEX I – EOSC IN PRACTICE STORIES.

As Vincent Cerf sums up last month: "*I hope our computer science community will find or invent ways to engage, using powerful computing, artificial intelligence, machine learning, and other tools to enable better quality assessment of the* ***ocean of content contained in our growing online universe***[2]".

This, group believes, is what the EOSC is trying to do and reflects the change in the way scientific research is carried out. With a twist: scientists are not alone to navigate in this ocean of data, but rely on an increasingly sophisticated set of tools, practices and networks that enable data sharing and re-use. As a central element of this report the 2nd EOSC HLEG has given set of practical considerations and recommendations for implementation of the EOSC, in the areas of Implementation, Engagement and Steering.

The following recommendations are the result of a combination of extensive discussion exchanges amongst expert findings, the stakeholders cited at the beginning of the report and first-hand experience from some of the Science Demonstrators (SDs) involved in EOSCPilot project[3] and from the EOSC in Practice stories.

Implementation recommendations:

1. The EOSC should implement "whatever works" and do "whatever it takes" to **increase the availability and volume of quality & user-friendly scientific information on-line**.
2. Define **EOSC interoperability standards** so that services can be interconnected and federated to be as effective as possible and be based on existing open standards.
3. Define an **EOSC Quality of Service (QoS) standards**, separate for all elements of the ecosystem (data, data access services, software, etc.), to develop a trustable ecosystem.

1 COM (2016) 178 final

2 "Unintended Consequences" By Vinton G. Cerf Communications of the ACM, Vol. 61 No. 3, Page 7 doi: 10.1145/3184402

3 http://eoscpilot.eu/science-demonstrators

4. Introduce, as part of EOSC's mission, that a **state of the art analysis is carried out on a national level within the Member States** for assessing statistics and key assets around the composition and relevant clustering of the community of users, with the respective eInfrastructures & research infrastructures & scientific communities.
5. The universal entry point to the EOSC should provide access to a **marketplace of efficient and effective services**, with lightweight integration (authentication and authorization infrastructure, order management, etc...) where service providers can find service users and vice versa. Nothing is wrong with a number of multiple entry points which should be seen as a plus rather than a negative fragmentation.
6. Adopt EOSC standards – from international standards - for information encoding and for protocols for data sharing and publication to **implement FAIR principles over data and services** in a complete & efficient way. And vice versa, offer open EOSC standards to international standardisation bodies.
7. Promote the development of services as **independent, interoperable and exchangeable building blocks** to foster the future accreditation of innovative and/or efficient alternatives.
8. Promote **open software development**, where possible, for all elements of the EOSC.
9. Introduce a **regular assessment of EOSC against other alternatives, including commercial providers.** This could be made to either enhance an EOSC Service, or to support new Services;
10. Simplify early (beta) participation in the EOSC by potential data providers, service providers, and underlying infrastructure providers, by relaxing initial constraints but without relaxing quality standards for data and services.
11. **Resource allocation**, particularly for early stage pilots, needs to be dynamic and amenable to change even at short notice;
12. Build a workforce able to execute the vision of the EOSC by **ensuring data stewards, data and infrastructure technologists and scientific data experts who are trained and supported adequately**.

Engagement recommendations:

1. Create **career-enhancing incentives** for researchers who open the science that they produce as also indicated by the OSPP, e.g., who lodge high quality, curated data in trusted repositories, share data services to their peers, or develop open software and services, and make the EOSC a portal to those incentives.
2. Develop, both at Member State and also EU level, appropriate **engagement schemes** whereby **publicly-funded research infrastructure providers and research communities take part in the EOSC.** EOSC should take **national and international developments into account** and should be connectable to national and international frameworks (e.g. the National Research Data Infrastructure in Germany or the AOSP ICT Infrastructure Framework[4])
3. To **stimulate the "supply side"** of the EOSC, ensure creation of economic incentives for research infrastructure providers to use and co-develop shared facilities and data repositories through the EOSC. This would be supported by the Commission through the European Data Infrastructure (EDI).
4. To **stimulate the "demand side"** of the EOSC, ensure establishment of dedicated funding for demonstrations of the EOSC at EU level that would include researchers and their infrastructure providers (e.g., EOSC in practice stories, cross-disciplinary success of EOSC).

Steering recommendations:

1. Ensure that the WGs and the other advisory structures well cover for the Executive Board **the latest scientific and organisational trends and novel ideas** for the necessary decisions in those areas.
2. Harness **inputs and pledges of the coalition of doers** and support their activities in the context of EOSC implementation.

4 http://africanopenscience.org.za/?p=230

3. Ensure, both at pan-European and international level, that research communities pursue **advanced partnerships**, also supporting with grants and other incentives, so that EOSC's progress is not dependent on the slowest but on the fastest movers.
4. Guidelines and rules should be clearly separated into (i) domains for which stability and trustability are important and (ii) domains for which progress must take place rapidly. The former should have rules and **instructions** that remain stable in time, whereas the latter should be run by **living documents**, facilitating innovation and change. All guidelines and rules should be accounted for in the **establishment and progress of the WGs**.

The2nd EOSC HLEG will conclude its journey in December 2018 with the presentation of the final version of the present report. The final release will have, in addition:

» Results from the Rules of Participation Consultation launched at the EOSC Summit in June 2018;

» Details on the EOSC Governance Structure;

» Options for the to-be EOSC Business Model based around the EOSC in Practice Stories;

» Rules of Participation to the EOSC for the benefit of the various stakeholders based on results from the consultation launched at the EOSC Summit.

1. INTRODUCTION

1.1 Objectives of the 2nd High Level Expert Group

The objective of the group is to advise the Commission on aspects related to practical implementation of the EOSC.

The HLEG EOSC was set up to add value and complement the mission of existing Commission expert groups (e.g. the Horizon 2020 Commission expert group on Turning FAIR data into reality). It has been charged with overseeing the practical, technical implementation of the strategic policy recommendations of 1st HLEG on EOSC, which concluded its activity in February 2017, in the context of Horizon 2020 programming.

Furthermore, the group has taken over the work of the cloud subgroup of the Open Science Policy Platform (OSPP) to ensure continuity on the EOSC topic and to guarantee an impartial, level playing field for all stakeholders. Specifically, the HLEG EOSC will take the recommendations of the OSPP on cloud governance into account and help articulate the Commission's position at the EOSC Summit in both summits in June 2017 and 2018.

The 2nd HLEG EOSC therefore zones in on two crucial aspects of the EOSC:

» governance of the EOSC (including rules of participation and management as such), as well as

» financing of the EOSC (including the development of novel ideas like the 'cloud coins and credits').

The group's goal is to consider these aspects and take a fresh look into the subject; it is composed of experts of varied backgrounds who act in their personal capacity and do not represent vested interests in research infrastructures, have a thorough understanding of the specificities of scientific research, fully understand the value of sharing and re-use of research data and have relevant knowledge on the value and functioning of scientific data/cloud infrastructures.

Additionally, the 2nd HLEG EOSC was asked to:

» Support with the Rules of Participation

» Advise the Commission concerning actions from the implementation roadmap of the EOSC initiative (by end of 2017), notably its governance and financing aspects.

» Elaborate and take forward the stakeholders' agreed conclusions resulting from the EOSC summit held in June 2017.

Collaborate with the project EOSCpilot, by providing views on the project's deliverables and by providing input.

1.2 Scope and structure of the document

This report looks to dig deeper into the EOSC roadmap and serve as delivering insights to help build the EOSC beyond the SWD in light of the imminent MS plus EC mandated board due to be launched in November 2018.

The European Open Science Cloud (EOSC) is a key enabler of continued European excellence in science[5]. The present document builds upon the work carried out by the First EOSC High Level Group (HLEG)[6]. It states the objectives of this 2nd HLEG, and recommends a way forward to implement the EOSC, supplementing the Staff Working Document, SWD[7], which provided the bedrock for this HLEG to build on. The 2nd EOSC HLEG has produced practical considerations and recommendations for implementation of the EOSC. Finally, the document also includes several practical examples selected from the EOSC "coalition of doers".

The approach taken in drafting this report has been that of a "minimum viable ecosystem", suggesting the steps to practically implement the EOSC as the effective product of such an ecosystem. The report identifies actors

5 The EOSC (European Open Science Cloud): https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

6The 1st EOSC HLEG (High-Level Expert Group) – "Realising the European Open Science Cloud": https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none

7 European Commission, "Commission Staff Working Document" – March 14th, 2018, http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

and roles in this minimal viable ecosystem, based on the idea of a minimum viable product; it helps define the EOSC features and governance structures and review the legal context.

The document is structured as follows:

» Section 1 gives an overall introduction on the EOSC HLEG;

» The Policy Landscape for EOSC is briefly summarised in Section 2. As part of the work conducted by the 2nd HLEG EOSC, synergies among EOSC and the Open Science Policy Platform – OSPP8 and the FAIR Data Group9 have been identified.

» Section 3 tackles issues related to viability of the EOSC ecosystem.

» Some options for the EOSC business models are provided in Section 4, with due consideration of their impact on the governance structure.

» Rules of participation to the EOSC are dealt with in Section 5.

» Section 6 looks at the road ahead in 2018, bridging the way into the next phase of EOSC development, also linking with stakeholders and defining the remaining challenges.

» Section 7 summarises all recommendations made, grouped by the stakeholders.

» The EOSC HLEG collected 13 "EOSC in practice stories", whose general message is reported in Section 8, with a detail provided in ANNEX I – EOSC IN PRACTICE STORIES.

» Preliminary conclusions of the HLEG work are reported in Section 9.

Two Annexes to the report are provided. ANNEX I – EOSC IN PRACTICE STORIES, illustrating some EOSC in Practice stories and providing motivation for its development by outlining several success stories and ANNEX II – REQUIREMENTS FOR CONFORMANT PROVIDERS OF CLOUD COINS – A WORKING EXAMPLE describing a more detailed scenario for one possible funding mechanism – the cloud coins.

1.3 Building upon the 1st EOSC HLEG

The first report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud "Realising the European Open Science Cloud"[10] were published on October 11th 2016. They emphasized that immediate action had to be taken to realise a federated, globally accessible environment – where researchers could publish, find and re-use data and tools for research. The 1st EOSC HLEG supported the Commission in outlining a general vision for the EOSC during the timeframe of the preparation and initial follow-up of the Cloud Initiative Communication.

In recognition of the significance of the EOSC Declaration issued in 2017, which delivers key statements on data culture and FAIR data, research data services and architecture, and governance and funding, the HLEG wish to add some recommendations to strengthen the EOSC in the areas of Policy, Governance, and Implementation.

In particular, from the **Policy** point of view, some Member States have taken immediate affirmative action on the EOSC. Those qualifying as early movers, as is the case for instance for Austria, Belgium, Denmark, Estonia, France, Italy, Malta, Portugal, Spain and the United Kingdom, developed an ERAC SWG OSI (Standing Working Group on Open Science & Innovation) providing their "Draft Opinion on EOSC Governance models and Strategic Implementation Plan"[11].

A number of proposals, closing in April 2018 and June 2018 under H2020, to define an EU framework for FAIR research data (the INFRA-EOSC-2018 calls[12], to develop initial catalogue of datasets accessible via the EOSC

8 OSPP: https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform

9 The FAIR Data Group: http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3464

10 The 1st EOSC HLEG (High-Level Expert Group) – "Realising the European Open Science Cloud": https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none

11 ERAC Secretariat, "Assessment of the Amsterdam Call for Action", February 2018, http://data.consilium.europa.eu/doc/document/ST-1202-2018-INIT/en/pdf

12 https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/calls/h2020-infraeosc-2018-2020.html#c,topics=callIdentifier/t/H2020-INFRAEOSC-2018-2020/1/1/1/default-group&callStatus/t/Forthcoming/1/1/0/default-group&callStatus/t/Open/1/1/0/default-group&callStatus/t/Closed/1/1/0/default-group&+identifier/desc

and develop Rules of Participation in consultation with stakeholders have been issued. Preparation of a FAIR data Action Plan, currently under the remit of the FAIR Data expert group is expected by the end of 2018[13].

As regards **Governance**, EOSCPilot and the 2nd EOSC HLEG are helping provide guidance on the creation of an internationally effective governance, as suggested also by the SWD. Moreover, on the "Amplify good practice" in the EOSC recommendation, the 2nd EOSC HLEG managed to collect and describe a number of "EOSC In Practice" stories, which are reported in ANNEX I – EOSC IN PRACTICE STORIES to the present report.

As to **Implementation**: Rules of Participation are indeed being developed, for subsequent endorsement and implementation. Moreover, regarding the 'funding scheme' recommendation of the 1st HLEG, the present report describes several possible options for practical business models (see Sec. 4). Finally, to 'develop a concrete plan for the architecture of data interoperability in the EOSC', the EOSCPilot initiative is delivering a detailed architecture plan[14].

13 The FAIR Data Group: http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3464
14 EOSC Pilot, http://eoscpilot.eu/.

2. THE POLICY LANDSCAPE: FROM COMMUNICATION (COM/20160178final) TO THE STAFF WORKING DOCUMENT (SWD 2018 83final) Council Conclusions TO LAUNCHING OF THE EOSC

There is increasing agreement among Member States and major research infrastructures and e-infrastructure stakeholders that an integrated approach to research data infrastructures is needed, going beyond layers (computing, data storage, use) and national and disciplinary silos.

In recent years significant development in open science, including open access to research publications and open data, has taken place in Europe. E-infrastructure commons were proposed. The creation of an EOSC as a federated, globally-accessible environment, where researchers, innovators, companies and citizens can publish, find and re-use each other's data and tools for research, innovation and educational purposes under well-defined and trusted conditions has received enthusiastic support from the scientific community. The set-up of the European Open Science Cloud (EOSC) is one of the ambitions announced by the European Commission in the Communication on the European Cloud Initiative launched in April 2016. The European Open Science Cloud vision is: "to give Europe a global lead in scientific data infrastructures and to ensure that European scientists reap the full benefits of data-driven science"[15]. EOSC will operate as a trusted, open environment for the scientific community for storing, sharing and re-using scientific data and results, supported by high-capacity cloud solutions with super-computing capacity via a European Data Infrastructure. The EOSC summit and EOSC Declaration serve as a bottom up component.

On March 14th, 2018, DG RTD and DG CNECT produced the "Commission Staff Working Document" (SWD)[16], an agreed-upon, preliminary concept of the EOSC. On March 14th, 2018, the Commission adopted the "Commission Staff Working Document" (SWD)[17] and transmitted it to the European Parliament and to the EU Council.

The main characteristics of the EOSC and its mission, as planned in the SWD, can be summarized as follows.

» EOSC aims to provide members of Europe's research community with "a virtual environment with free at the point of use, open and seamless services for storage, management, analysis and reuse of research data, across borders and scientific disciplines" European Cloud Initiative COM (2016) 178 final.

» EOSC as a concept builds on the observation that the current rate of progress, in both practices and infrastructure build-up, for the generation, storage, and sharing of scientific data, is very high; yet progress and initiatives are highly fragmented, across several dimensions (practices, disciplines, countries). EOSC therefore has a main goal to federate existing initiatives and structures within a single, consolidated and seamless platform.

» EOSC aims to achieve this through six action lines, comprising:

(i) provision of an architecture for federated infrastructures;

(ii) FAIR data management and tools to ensure data stewardship across borders and disciplines;

(iii) services designed from the user's perspective;

(iv) access mechanisms and interfaces;

(v) rules of participation for its actors, and;

(vi) a governance structure that can pilot EU leadership in data-driven science.

EOSC is to respond to the needs of the community; its construction therefore incorporates the participation of a "coalition of doers" defined and assembled at meetings in June 2017 and subsequently in June 2018; EOSC is to federate existing and planned research data infrastructures, connecting them with a soft overlay and build upon existing large-scale EU scientific networks including ICANN, IETF, AIOTI, GÉANT and ELIXIR.

15 The EOSC (European Open Science Cloud): https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

16 European Commission, "Commission Staff Working Document" – March 14th, 2018, http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

17 European Commission, "Commission Staff Working Document" – March 14th, 2018, http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

2.1Impact of current regulation

EOSC activities will need to comply with the current EU legal framework and Member State laws and take any planned changes into account. The EC is aware of the importance of the free flow of non-personal data for the emergence of the data economy. It claims that "Free flow of non-personal data is a pre-requisite for a competitive data economy within the Digital Single Market. To fully unleash the data economy benefits we need to ensure a free flow of data, allowing companies and public administrations to store and process non-personal data wherever they choose in the EU[18]."

A proposal for regulation was defined[19] in September 2017 "to ensure the free movement of data other than personal data within the Union" and "laying down rules relating to data localisation requirements, the availability of data". It states specifically that "Location of data for storage or other processing within the Union shall not be restricted to the territory of a specific Member State, and storage or other processing in any other Member State shall not be prohibited or restricted, unless it is justified on grounds of public security."

The Regulation is limited to data "other than personal data". Personal data should also not be handled by EOSC, as, for example, in May 2018 the Regulation (EU) 2016/679 of the European Parliament and of Council of 27 April 2016 on the protection of natural persons with regards to the processing of personal data and on the free movement of such data (GDPR) will be in force.

In autumn 2017, the Commission conducted a public consultation in view of reviewing the directive on the re-use of public sector information (PSI Directive). This Directive requires that public sector bodies make data re-usable by third parties but does not currently include research data in its scope. It is still not clear whether bringing research data into the scope of the PSI Directive would lead to more benefits, as opposed to continuing with other 'soft law' measures that are showing some results (e.g. the Open Data pilot under H2020). Initiatives regarding research data, government data and private sector data will all be part of the European Commission's 2018 Data Package, to be delivered in the spring 2018.

Privacy or commercial sensitivity concerns could justify exclusion of the re-use of certain data. In the case of EOSC, the rules of participation will define the rights, obligations and accountability of the EOSC actors including data producers, service providers, data/service users, vis-à-vis the applicable legal frameworks (e.g. GDPR, copyright rules, Data Security and Cybercrime, dispute resolution and redress mechanisms, e-commerce directive).

The above and other regulations, as well as advances made in cloud computing together with Artificial Intelligence has also created uneasiness among some scholars who have misgivings[20] that these developments challenge "traditional legal principles and increase legal uncertainty of various rights protection in the information society". They conclude, and this panel agrees, that legal certainty is of paramount importance if new technologies are to contribute to economic and societal growth.

2.2 THE EUROPEAN OPEN SCIENCE CLOUD, FAIR & OSSP SYNERGIES

The best way to identify synergies between the three advisory bodies of the European Commission OSPP, the FAIR Data Expert Group, and the HLEG EOSC is to categorise them first. Useful aspects to take into consideration in this respect are the individual scope and time frame of each.

The OSPP paints the broader picture of **Open Science** as a goal and gives policy advice to the Commissioner encompassing dimensions such as reward systems, measuring quality and impact (altmetrics), changing business models for publishing, FAIR open data, European Open Science Cloud, research integrity, citizen science and open education and skills.

The FAIR Data Expert Group, initiated by the European Commission, looks closely at the FAIR data principles and formulates recommendations on how to implement them. The focus here lies on cultural change, incentives and metrics, as well as the skills and capacity that need to be built to make **FAIR data** a reality.

18 https://ec.europa.eu/digital-single-market/en/free-flow-non-personal-data
19 http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=46830...
20 http://www.kluwerlawonline.com/abstract.php?id=EULR2017029

Finally, the HLEG EOSC is expected to make recommendations on how to shape and implement the **EOSC** as a federated infrastructure for data-driven research, based on open standards and best practices that support the open creation and dissemination of scholarly knowledge and FAIR scientific data. According to the European Commission, the EOSC aims to create a trusted environment to host and process research data in support of EU science in its global leading role[21].

While the OSPP has the mandate to support the development and implementation of open science policy in Europe, the European Commission sets the agenda with the proposal of the EOSC. The FAIR Data Expert Group is one of the first groups to formulate more detailed recommendations on how to translate the FAIR Data Principles into practical guidelines, based on ongoing effort and best practices in the European member states. The broader picture of the OSPP goes hand-in-hand with a long-term perspective, while the FAIR Data Expert Group recommendations are to be seen from a mid-term perspective, as is the case with the work of the HLEG EOSC to build a federated infrastructure.

The EOSC will only be successful if the federated infrastructure is embedded in a cultural change, thus the work of the HLEG EOSC and the FAIR Data Expert Group is recommended to closely link the two together. Only when researchers value curated data as important research results, rather than sticking to publications (of results), will the full potential of the EOSC be unleashed. This is where the OSPP and the other two expert groups come in. The OSPP should support the European Commission and the Member States to provide guidance on open science principles. This is exactly what the OSPP did by publishing eight broad recommendations for the implementation of the EOSC that can be seen as guiding principles for the HLEG EOSC and the FAIR Data Expert Group[22].

Wisely, the OSPP embeds five assumptions related to the federated infrastructure in three principles addressing issues like awareness, skills development and ethics. For example, to optimally use the EOSC in the future, and to implement it accordingly, the skills set related to research data has to be improved in the scientific system to provide FAIR data services. Data administrators could provide mediation between science and the developers of infrastructure. Nevertheless, researchers will at least need some core competencies in data management and will ideally be supported by data stewards and data scientists.

Hence, the eight OSPP principles built the ideal framework for a balanced division of effort between the HLEG EOSC, which addresses the first five and the FAIR Data EG, which focuses on the latter three. However, it is noteworthy that the somewhat top-down approach of these three (Commission initiated) expert groups and the Commission's work have to be supported by bottom-up processes. Some examples, this is a non-exhaustive list are: RDA, the GO-FAIR initiative, H2020 implementation projects and Coordination Actions.

The Research Data Alliance (RDA) is precisely that, a bottom-up initiative that focuses on research data management. It takes a long-term international approach to the topic and aims to provide guidance on research data standards that are helpful not only 'for some', but also 'in sum' for all research communities. While RDA helps underpin the work of the three different working groups related to research data management, effort is still required to support the infrastructural aspects. With the mid- to long-term scope of ESFRI and the various ERICs encompassed in RDA, the European landscape is well-prepared to meet these needs. The discipline-specific ERICs could help to distinguish between the generic and research community-specific infrastructure needs and ESFRI provides the perfect forum to summarise the findings.

Nonetheless, synchronising the European initiatives with global endeavours while maintaining momentum, requires a fast-track implementation process which harmonizes both research data management and infrastructure needs. Therefore, to fully complete the picture with a relatively short-term, bottom up initiative, GO FAIR is worthy of mention. According to a joint position paper, Germany and the Netherlands, together with France, see GO FAIR as a fast-track implementation initiative to boost the further development of the EOSC[23]. It aims to build on existing initiatives with critical mass, at the Member State level, such as the German National Research Data Infrastructure (NFDI) or the Dutch National Open Science Platform, among other science-driven initiatives, and to identify the early implementation needs of existing networks and consortia. GO FAIR adopts an implementation approach to FAIR research data and services. It builds upon the recommendations of the first High Level Expert Group on the European Open Science Cloud with the objective to 'federate the gems' across the Member States[24]. GO FAIR follows a bottom-up, open implementation strategy for the technical

21 https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
22 https://ec.europa.eu/research/openscience/pdf/ospp_euro_open_science_cloud_report-.pdf
23 https://www.dtls.nl/wp-content/uploads/2017/05/DE-NL-Joint-Paper-FINAL.pdf
24 https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud-hleg

governance and funding needed to establish the first phase of the European Open Science Cloud (EOSC), as part of a broader global Internet of FAIR Data & Services. GO FAIR establishes Implementation Networks (IN) of research communities that will bolster and specify the principles and recommendations of the Commission's expert groups from the bottom up.

All activities mentioned above have a stronger focus on research data as opposed to services for research data management. For example, the H2020 implementation projects and coordination actions (e.g., eInfraCentral, EOSCPilot, EOSC Hub, INFRA-EOSC-2018 calls, FREYA, OPENAire-Advance, RDA Europe 4 etc..) offer service platforms with services at different levels (generic vs. community-specific) for different stakeholders involved in research data management. Still, some challenges remain and should be addressed in the near future, as indicated in Sec. 1.6.4. A synthesis table comparing the relevant initiatives with differences and commonalities will be included in the final version of the report.

3. MAKING EOSC A VIABLE ECOSYSTEM

The European Commission and the EU Member States need to put the required mechanisms in place for the EOSC coordinated effort to materialise, from a policy, technical and user perspective. At this initial stage, a significant step forward is a Minimal Viable Ecosystem (MVE) that would enable EOSC to emerge as a collaborative effort, in an iterative way.

While the present report addresses the aspects of implementation, engagement and steering, more along the lines of user engagement (and proposed value proposition for them) will be dealt with in the final release of the present report, including motivational mechanisms and levers to be used to persuade them to change their practice and use the EOSC versus other potential options.

3.1 The EOSC Minimum Viable Ecosystem

The provision of resources to support the EOSC will take place in a very heterogeneous landscape of e-infrastructures and service providers, with dispersed users at best aggregated around disciplinary poles and national infrastructures. Addressing this challenge requires the definition of a smallest common denominator, referred to here as the **Minimum Viable Ecosystem (MVE)** of the **EOSC**.

An MVE will emerge and thrive only if some basic technical, political and human resources conditions are met. From a technical perspective, interoperable services and open data must be guaranteed. On a human resources level, a coordinated effort has to be made to put the right incentives in place for all the actors involved (researchers, software developers and infrastructure managers, Research managers) to design, contribute to and exploit the system. Policies are needed for the technical and human-oriented conditions to emerge. In modern science, recognition comes mostly from scientific outputs, and this is a fundamental factor to consider for the MVE to emerge. Policies have the greatest influence in setting incentive and obligations from publicly funded research. The MVE will possibly evolve according to the future expectations of EOSC, and the related success criteria.

3.2 Identification of actors, roles, and results of their work

The main actors in the EOSC MVE, as outlined by the EOSCPilot (see for instance the EOSCPilot booklet25), are:

» European researchers structured in small or large teams based around institutions;

» Software developers: from individuals, to small to medium teams;

» Infrastructure managers;

The involvement of all the actors heavily depends on the Rules of Participation. We outline here the basic description from the point of view of their role in EOSC, and leave for the final report a more detailed explanation on how the Rules of Participation of the different actors will contribute to incentivize and generate the MVE.

EUROPEAN RESEARCHERS

The researchers' job is based on data and on computational resources. They need to produce or find data relevant to the inquiry, find an appropriate service or hosting for their own data and services, do the necessary transformations, run the analysis, publish the results and make data available to others.

25 EOSC Booklet https://eoscpilot.eu/stakeholder-forum/eoscpilot-booklet , November 2017.

Flexible ways to access and share data and direct access to fast networks to do so are at the top of the agenda for researchers. One of the main problems that the EOSC needs to solve is the fact that researchers in Europe still have insufficient access to e-infrastructures.

Giving access to all the researchers in Europe to the Pan-European Research Infrastructures facilities in a straightforward way would provide sufficient incentive for scientists to adopt EOSC as a platform to access computing and storage services, or share their scientific data.

SOFTWARE DEVELOPERS

Interoperable services and open data rely on the principles of software openness. The software used in EOSC services should guarantee interoperability: adhere to standards, be they de facto or de Jure; data produced and handled with EOSC software services should respect the FAIR principles; services within EOSC should be secure and adhere to the European Authorization and Authentication policies; as a general policy, the software elements are provided upstream to open source projects, to guarantee the required level of sustainability; etc.

For this key activity to be successful in terms of engaging human talent, breakthrough ideas leading to innovation need to be awarded with the proper recognition. Putting in place transparent mechanisms to recognize successful software development, such as creating an ***"EOSC-Ready"* certification for software** products, would have a positive impact on the software development ecosystem in Europe. The successful development of an *"EOSC-Ready"* branded software product, would improve the Reputation of researchers and technologists and dynamically harness the potential of European developers, across academia and industry.

INFRASTRUCTURE MANAGERS

From the incentive point of view, infrastructure managers can perceive EOSC as a big opportunity. Firstly, integration of their resources in the EOSC is a way to achieve a higher, more efficient usage of resources. The coordination effort should translate into economy of scale for infrastructures, but also into a potential expansion in their user base.

More importantly, at the level of increasing competitiveness, integrating in EOSC opens the opportunity to gather experience in running innovative services needed by cutting-edge research projects, supporting multi-disciplinary teams, etc.

Both managerially speaking and from a human point of view, the generation of expertise in deploying and running advanced services to support frontier research creates *know-how* in the resource centres. Often this type of service is only available in Research Infrastructures as prototype, long before it becomes commercially viable or profitable, if ever. When fed back into industry, in the form of trained people, it is this *know-how* that deliver the added value necessary for economic growth.

OVERALL USER ANALYSIS

A schematic user analysis for EOSC is summarised in the table below (and it will be further expanded in the final release of the present document).

User / Provider	Actions	Story ending
End User	Register for use Discover service Find data Transform data Run analysis Store results Pay for service Sponsored to use a service	Evidence based on research accomplished, followed and cited

User / Provider	Actions	Story ending
Service developer	Identify user needs Create services Publish service Provide consulting about service Charge for service	Investment into development of service returned
Research funding organisation	Identify user needs Recognition opportunities Aggregation services	Acknowledgement of the EOSC as central reference in research funding themes
Core Infrastructure provider	Attract service hosting Charge for hardware resource use	Well exploited, secure, interoperable and searchable infrastructure

Table 1 – Actor analysis

3.3 Features of the Minimum Viable Ecosystem (MVE)

In the final version of the report, after the list of user actions and associated pains and gains has been completed, some features will be selected to constitute the proposed MVE using a matrix, such as the one outlined in the figure below.

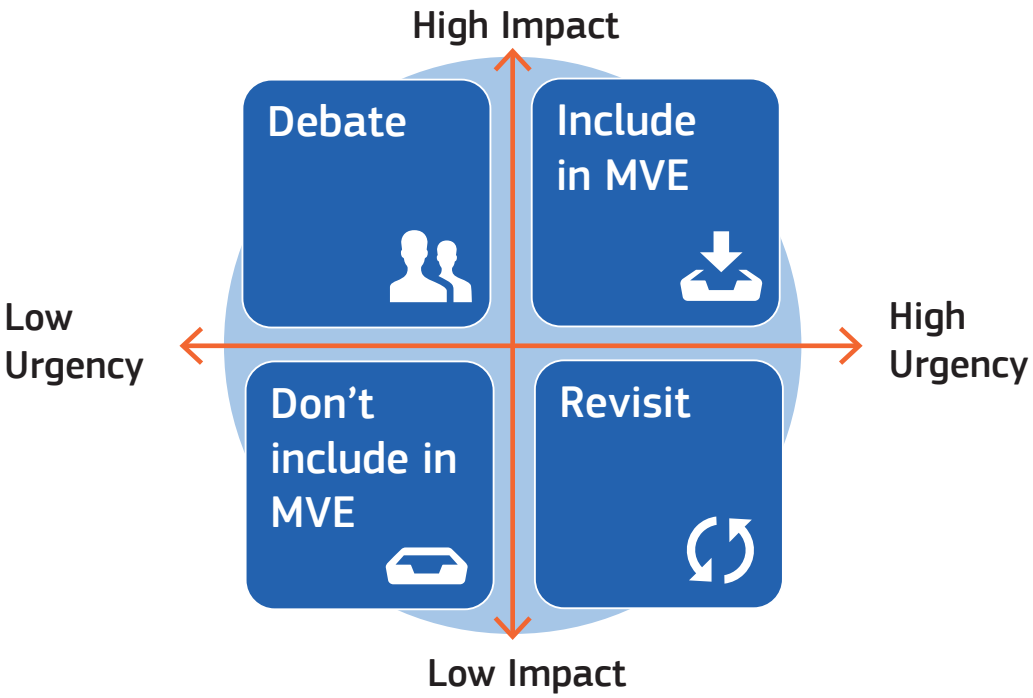

Figure 1 – Selection matrix for the MVE

3.4 Governance

To commence with, the governance of EOSC should enable the MVE process as defined in 5.1. Furthermore, the proposed structure needs to be capable of running a fully-fledged EOSC. As mentioned above, the users have to play a central role in the design and implementation of the EOSC. Requirements, standards, operating procedures, etc., should be defined through close collaboration between:

» the end user - scientific community;

» the service providers – developers, intermediaries and operators;

» funding agencies and scientific policy makers.

Rapid realisation of an EOSC Governance is needed to move from vision to implementation. The two-stage approach for the implementation of EOSC has received general support, the first stage being the process of developing the EOSC and the second stage its management, operation and evolution. EOSC is a concept of an important core scientific infrastructure for European science that will need to lead the future development of infrastructures far into the future.

Having said all of the above, and in full conformance of the SWD, a three-layer governance model, based on the EOSC Declaration is proposed, as depicted in the following Figure. The three layers are:

1. **Strategic Layer** in the form of an EOSC Board to combine state-of-the-art expertise on scientific cloud infrastructures with the Funders and Policy Makers. The Board will therefore include EU Member States and Associated Countries representatives. The EOSC Board would mainly make strategic decisions on the development and evolution of the EOSC.
2. **Executive Layer** in the form of an Executive Board to manage day-to-day operation of the EOSC and procurers designing and planning work-related future developments. This would be the only full-time staffed layer, will be supported by Working Groups, and would have the responsibility of ensuring that user needs are met and strategic requirements addressed.
3. **Stakeholder layer** organised in the form of a Stakeholders Forum to provide a medium for stakeholders: Users (Consumers), Providers and Intermediaries of EOSC Resources. This would have the main role to discuss, supervise and channel communication between the EOSC and the communities across all three layers. The Stakeholders Forum advices the Strategic and Executive Layers through its Presidency.

A **CSA Structure** is providing additional support to the Executive layer.

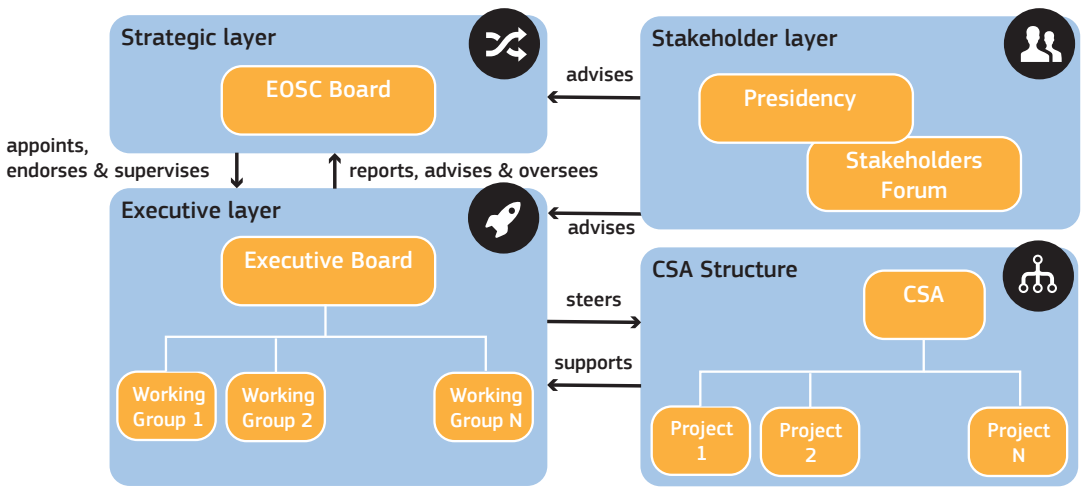

Figure 2 – Three-layer governance model

The three-layer model above provides a good basis for discussion and decision-making between Member States at Council level, enabling the EOSC to move forward. The structure must ensure an adequate flow of information between the different governance layers. It should also accommodate the transparent sharing of responsibilities, as well as their supervision.

The **Stakeholders Forum** has to grant all stakeholder groups the possibility to determine the requirements, policies and principles of participation. The Stakeholders Forum should on the one hand advise both the EOSC Board and the Executive Board, while appraising the work of the two Boards on the other. The Stakeholders Forum has to have an organisational structure that enables consumers, providers and intermediaries of all sizes to participate and it will be interacting with the Boards by means of its Presidency. The Stakeholders Forum Presidential board proposes the rules of membership and participation and processes and acts as the key contact point for the Strategic and Executive Layers.

A number of different **Working Groups (WGs)**, formed on the basis of interoperability contexts, the stakeholder roles, or broad scientific or infrastructure domains, could be created. The WGs should be time-based (e.g., up to 2-year duration) and should work on specific areas, whereby the priorities are set by the Stakeholders Forum, in conjunction with the Strategic and Executive Layers. The Forum has to be open to discussing new activities, which may, in turn, lead to the creation of new groups. Some ideas already for Working Groups which merit further analysis and discussion are:

1) The Rules of Participation WG
2) Architecture WG
3) Open Standards in Service Development and Seamless deployment WG
4) Resource Allocation WG
5) EOSC Business Models WG
6) FAIR Principles over Data & Services WG
7) Data Management Policies WG
8) International Liaisons WG
9) Open Science Policies WG
10) Data Security & Compliance WG

The **EOSC Board** should involve Member States and Associated Countries and Commission representatives and receive mandate from the Stakeholders Forum and the Competitive Council. The EOSC Board should review, agree and prioritise the EOSC proposals and requirements from the strategic vision point of view. The Board should monitor and assess the achievements of EOSC, approve the list of the Executive Board members and Executive Board work plan. The Strategic Board has to ensure the coordination of Member States and Commission initiatives.

Members of the first EOSC Executive Board could be selected through a similar procedure used in the case of the first European Research Council Scientific Council. The Commission could nominate a four-person Identification Committee from the representatives of the Open Science Policy Platform, European Strategy Forum for Research Infrastructures and e-Infrastructure Reflection Group, who could select the first ten EOSC Executive Board members from the candidates proposed by some of the main European branch organisations and initiatives (e.g., ESFRI Infrastructures, eInfrastructures, scientific organisations, networks and university associations). The Identification Committee could then submit the list of members of the Executive Board, as well as the names of the Chair and Vice Chair, to the EOSC Board for approval. The term of office of Executive Board members could be for a maximum of 4 years. The procedure for the selection of the following set of EOSC Board members could be to have a list directly proposed by the Executive Board and approved by the EOSC Board. Following each 2-year partial renewal of the Executive Board could take place.

In the Executive Layer of governance, the **Executive Board** designated by the EOSC Board should include representatives of the stakeholders from European branch organisations. The Executive Board would provide the EOSC Board with background and options for the way forward and propose its work plan for approval. To support the implementation of the workplan the Executive Board could work with stakeholders or set up ad hoc Working Groups. The Executive Board could then select Working Group members from the Stakeholders Forum or map its Working Groups to Stakeholder Groups.

During the first phase of the EOSC, the Executive Board will receive support from the CSA Structure (effort coordinated through a specific H2020 CSA project). The CSA Project(s), funded by Horizon 2020, will help the Executive Board coordinate all relevant EC-funded projects and support the implementation of the work plan and act as a secretariat to deliver the EOSC main functionalities.

4 EOSC BUSINESS MODEL: FINANCING THE EOSC

As highlighted in the Draft Council conclusions of the EOSC May 2018 this chapter provides some considerations around developing a sustainable business model and the council invites the Commission to elaborate, in consultation with the Member States, on the future financing of the EOSC. This section elaborates the business model, considers the Governance implications gives examples of different business models, a hybrid business model that uses a mix of more conventional grant/contract acquisitions and a 'cloud coin'-like environment.

Conventional grants and contracts could be a logical way to manage the fixed costs of an EOSC, for providing resources to academic institutions and the private sector to also include models such as pre-commercial procurement for the development of new services for development of capabilities or making large public data sets available on widely accessible infrastructure. Cloud coins are most useful for managing the marginal (or operating) costs of the EOSC, providing individual investigators with the ability to effectively move among the variety of resources that constitute the EOSC. They are also convenient when it comes to monitoring.

4.2 Business model

The EOSC Business model is a critical non-technical element that will determine the success of the EOSC vision. Given the dispersed nature of scientific research and the variety of tools and processes required by scientists in different fields and locations, a federated environment requires a similarly decentralized business model to support the technical environment that will be developed. Such a business model must:

1. Support the rapid acquisition and delivery of a variety of cloud services and other technologies needed by investigators
2. Allow International, National and private funding entities to maintain key electronic data and software resources for the benefit of the scientific community
3. Enable reuse *in situ* of high value scientific digital objects, by which we mean data, software, metadata, workflows and other digital artefacts of scientific research.
4. Ensure sufficient interoperability to consent movement of digital objects between environments for reuse elsewhere, where unique capabilities exist elsewhere
5. Maintain appropriate requirements on providers (of both digital objects and services) so that it is simpler for investigators to utilize FAIR principles for digital objects and
6. Limit the scope of the federation of scientific clouds to a reasonable number to help ensure that critical masses of digital objects exist in locations where they can be aggregated and reused
7. Ensure that the private sector re-invests in R&D to stimulate innovation and create new markets
8. Provide the necessary human support for the technology delivered to scientific users.

The final report will analyse the externalities for research funders in terms of decreasing overheads and increasing aggregation of user demands, what human resources would be necessary, the potential of re-use across fields and others.

The currents model for provisioning access to Research Infrastructures is based on the guidelines contained in the Charter for Access, where three main models are described:

A. Excellence-Driven Access: exclusively dependent on the scientific excellence, originality, quality and technical and ethical feasibility of an application, evaluated through peer review conducted by internal or external experts. This enables Users to gain access to the best facilities, resources and services wherever located. This mode enables collaborative research and technological development efforts across geographical and disciplinary boundaries.

B. Market-Driven Access: is defined through an agreement between the User and the e-Infrastructure that will lead to a fee for the Access and that may remain confidential. This is on a grant-based option.

C. Wide Access mode: guarantees the broadest possible gateway to scientific data and digital services provided by the e-Infrastructure to Users, wherever they are based. Adopting this mode maximises availability and visibility of the data and services provided.

In practical terms, a model based on the Wide Access mode modulated by a negotiated, agreeable Access restriction, is the pragmatic way to start moving with the EOSC. Private providers willing to provide resources within the EOSC framework will envision a Market-Driven approach to support users.

4.2 Governance, transparency and accountability

The EOSC board will have the ultimate responsibility for defining the business model and the requirements for participation, as well as oversee that EOSC service providers operate according to the rules of participation. The key is complete transparency of that the marketplace for cloud services for all stakeholders: CSPs, academic providers and users of the EOSC.

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 must be 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. The EOSC Executive Board should provide an annual report that provides insights into the relative use of various CSPs and services.

All data and service providers need to comply with the applicable EU data protection regulations and other requirements in terms of technical and organisational measures and ensure the protection of the rights of the data subject. Compliance with some of these regulations may be subject to certification, whether this is to be self-certification, or an accredited certificate, lies with the governing board and depend on the level of security.

4.3 Funding Model and Payment Mechanisms

Regarding funding models (how money transit from public and private coffers to ensure that data is open (opened up) and shared (and re-used), & payment models A payment model is how this money actually is distributed in the ecosystem, e.g. how in what modes it can be transferred to ultimate data and service providers. Several possibilities need to be envisaged in this heterogeneous landscape.

The most obvious model involves member states and multistate entities (e.g. the EC, EMBL, etc.) directly supporting elements of the EOSC that exist locally or in areas of scientific expertise, via institutional funding, direct grants and contract funding.

In order for such support to create an EOSC MVE, rather than simply another set of digital silos, recipients of support would need to meet a set of technical and operational standards that would ensure that these resources (storage, computing, and higher order services such as software, pre-defined workflows, etc.) are accessible to scientists outside of the host institution and across member states. Computing assets and research infrastructures such as ESRF[26] in Europe and at many US National Laboratories have been successfully funded using similar models. In those cases, the institute receives a grant from the funding entity (the US Department of Energy in this example) to build/operate the resource and make it available to other grantees of the funding entity. These grantees apply for capacity at the resource, which is tasked to identify high value research projects, and indeed, the resource operator is evaluated on the quality of the science that it provides resources to. The limitations of such a payment model is that the amount of resources available is by definition pre-determined; that is, there is often little ability to increase or decrease capacity in the event that the resource is incorrectly sized. In addition, restrictions placed by law or policy at funding agencies can restrict the ability of certain researchers to access these researchers – in the example cited above, non-Department of Energy grantees cannot get access to the resource.

26http://www.esrf.eu/

Another option is a funding and payment model that is based on a certification programme for commercial and non-commercial providers of computing services that can provide scientifically useful services, that will meet EOSC-defined standards to ensure minimum levels of access and interoperability and that will accept specific, EOSC-defined financial transactions in payment for these services ("cloud coins", see below). Ideally, this will create a competitive marketplace for these services that could, with appropriate governance, become an implemented EOSC. The basic outline of such a model for the EOSC is relatively straightforward.

The EOSC Board defines a series of **minimum requirements for commercial or non-commercial entities** that wish to participate in the EOSC as service providers. These requirements have been preliminarily discussed in the "Principles of Engagement" draft document elaborated by EOSCPilot[27] and shall be further specified in the upcoming months by the various EOSC initiatives. The EOSC should be designed to maximize the ability of data and other digital objects to meet FAIR criteria. At a minimum, these requirements would likely include compute/storage/network capacity, accessibility, interfaces (these could be as simple as basic upload/download/execute commands or value-added Application Programming Interfaces), Identifiers and Metadata, Authentication/Authorization, Information Assurance, compatibility with various regulatory requirements where necessary[28] and a willingness to accept euro-denominated EOSC vouchers, distributed to investigators through grants, or supplementary awards, that can be used to purchase services from these compliant providers. This model will dictate that a series of providers of services will have to appeal to thousands of individual scientists and research groups, incentivizing competition among them, which should in turn yield better service levels at lower prices. To coordinate acquirement, the EOSC and member states would also certify one or more brokers to manage the acquisition, distribution and payment for EOSC vouchers. These brokers could be government agencies in member states, entities within member states, transnational governments or private firms.

Such a model would have several advantages. First, it incentivizes both providers and investigators to converge to highest value services at the lowest possible price, and for providers to compete to develop new services that will be of interest to the scientific community. Secondly, by requiring that providers meet certain requirements for broad access to the resources being provided and using relatively standard transaction types, this model promotes easy reuse of data and other digital objects by other investigators. To the extent that is appropriate, cloud technology is utilized individual resources can be scaled up or down according to usage. If the transaction method is created appropriately, it provides funding agencies with a unique level of insight into the utilization of various types of scientific computing and particular data and software assets via the pre-existing reporting capabilities of the global financial network. Finally, such a model simplifies acquisition by a variety of categories of entity. As long as the brokers have appropriate agreements in force with the full provider network (which should be mandatory to become a broker), the various member states can acquire, distribute and pay for cloud coins under their own acquisition regulations, with no additional complexity burdening the individual investigator.

The National Institutes of Health in the United States has conducted a cloud credit experiment. Eight service providers (e.g. Infrastructure, Platform, Software), became conformant providers. Approximately $3m were distributed as cloud credits. This project will complete at the end of the 2018 Fiscal Year. The requirements for conformant providers are provided as an Annex to this document.

How the transaction is carried out, e.g. the payment model proper, is largely up to the entity funding the EOSC vouchers. Options could include pre-purchase of services, escrow of funds with investigators given withdrawal rights up to certain amounts, or distribution of credits via a pre-paid debit system. In addition, in this model, funding agencies or participating states could choose to directly fund the costs of maintaining/archiving key scientific data sets or other electronic resources in environments that scientists regularly select for research purposes. Moreover, this model foresees a dispute resolution scheme. For example, if a user pays with EOSC vouchers but does not obtain the quality of service foreseen (i.e. the service is 'down' for a period of time in breach of the Service Level Agreement) a mitigation is introduced.

The question remains as to the management of a coordination centre/gateway/marketplace. Such a site(s) would be helpful to assist in the integration of the disparate pieces of the EOSC federated infrastructure. Several sources of revenue could be envisaged to support the management of these resources. One possibility could be to require conformant commercial service providers to pay an annual subscription fee to offer their services in the marketplace. Another option could be a transaction tax on vouchers (i.e. a service provider is taxed a small

27"Principles of Engagement" EOSCPilot, 2018 - https://docs.google.com/document/d/1JBd30DMG_KhG8vIh1lrOIM4VpYFjOnk4sS59whjaIEI/edit

28Potentially optional in certain contexts.

percentage on each voucher that is redeemed against their services), perhaps combined with a transaction tax on private sector users. The investment recovered could then **contribute to funding the operational costs of the EOSC Gateway/ marketplace** and the provision of key open data sets. Ideally, the marketplace would keep track of how frequently a dataset is used and the provider of that data set compensated accordingly, similarly to how YouTube pays people who upload videos based on how many times they are viewed.

The following table provides a pro vs. con - based comparison of the three major options: the conventional 'Direct Support' funding and payment model, a pure EOSC vouchers model and a hybrid model where some resources are supported via direct funding and other resources are accessed with cloud coins.

Model	PROS	CONS
Direct Support: Elements of EOSC receive direct payments from funding agencies	1. Simple to operate and implement, based on current funding mechanisms 2. Cost effective and tested mechanism to maintain core data assets for broad use 3. Provides EOSC nodes with core revenue to develop servßices that might not be economical otherwise	1. Frequent capacity mismatch (either too large or too small) leading to cost inefficiencies or access issues 1. Resources can have internal foci, reducing access from outside stakeholders 2. Constant funding stream reduces the incentive to innovate to attract users and ensure adequate support 3. Burdensome for commercial entities, even where they could provide significant cost savings and be incentivized to innovate.
Pure EOSC vouchers**:** Researchers use 'cloud coins' to support EOSC services/data	1. Enables maximum choice for researchers by operating in a competitive marketplace for services 2. Enforces innovation by requiring services to support themselves via 'cloud coin'-based cost recovery 3. Provides simplified access for commercial providers to the marketplace 4. Expected to simplify acquisition of 'just enough' services and place significant pressure on costs leading to a migration of data management costs from infrastructure to ensuring FAIR principles	1. Much more complicated to implement, requiring the creation of significant trust and payment management networks. 2. More difficult for non-commercial organizations to directly interface with the EOSC, due to a lack of venture capital in such environments 3. Poor model for maintenance of crucial public data sets – would require that end users of data to 'vote' with credits to maintain data. 4. External focus (to attract investment via 'cloud coin' users could reduce the ability of these resources to support internal stakeholders)
Hybrid: EOSC vouchers and direct support	1. Provides necessary resources for the management of well defined, but extremely high value data sets. 2. Supports access by both commercial and non-commercial organizations into market, allowing both groups to innovate in areas where they have specialized abilities 3 Provides clear opportunities for innovation and rapid development via the creation of a competitive market in scientific information technology services.	1. Increased complexity due to the nature of a mixed market 2. Implementing the 'cloud coin' market will remain complex but can be built organically even as other resources utilize more conventional funding.

Model	PROS	CONS
Hybrid: EOSC vouchers and direct support	4. Provides maximum choice to researchers, who can utilize multiple categories of resources 5. Expected to simplify acquisition of 'just enough' services, thereby ensuring needed capacity and reducing expenditure on unnecessary services.	

Table 2 – Comparison among possible funding/revenue models

5. RULES OF PARTICIPATION

At the bottom-line, the EOSC aims to support three objectives: (1) to increase value of scientific data assets by making them easily available to a greater number of researchers, across disciplines (interdisciplinarity) and borders (EU added value) and (2) to reduce the costs of scientific data management, while (3) ensuring adequate protection of information/personal data according to applicable EU rules (e.g., REGULATION (EU) 2016/679). Indeed, cost reduction is a key element to ensure that resources remain available to carry out the first objective without requiring reductions in the resources necessary to carry out cutting edge research that will generate the next generation of scientific data. Moreover, an EOSC needs to be developed with the fundamental principles of data protection by design and information security in mind. Further "ancillary" objectives of the EOSC are inclusiveness, bearing a clear value proposition, simple to adopt, among others.

A key part of the governance framework is a set of rules of participation that support these objectives by setting the rights and responsibilities of participants to the EOSC. Collectively, they will help ensure that the objectives of the EOSC, described above, are met.

For the final report, an analysis may be carried out to map the FAIR chain into different parts of the EOSC MVE, in relation to the rules of participation. E.g. **Findability** may be a function that is best dealt with centrally, by the 'core' of the EOSC, and responsibility at the same time distributed to data producers via the adoption of DOI that should be developed in parallel. **Accessibility**, what is the best 'locus' to embed it in the MVE?

5.1 Federating the existing infrastructures

A key element of an EOSC MVE is the incorporation of resources that have been used, widely, that have demonstrated value for the scientific community and that provide adequate guarantees in terms of data protection and information security. It is impossible to imagine a reduction of costs if existing capabilities are nor included or are re-developed as part of an EOSC. That is not to say that the federation be limited to existing actors. Robust ecosystems require a constant influx of new resources and, likewise, the departure of those that are no longer widely useful. Existing providers that wish that some of their offering become part of the EOSC will need to follow requirements as per the eligibility criteria for actors but their presence will be essential to creating the critical mass of data and services needed to drive users into the EOSC.

How simple or laborious it will be for existing entities to join the EOSC will depend on a series of choices made during its implementation. For example, if it is determined that a service provider has to provide computing infrastructure in addition to data, sites that are primarily data repositories would most probably have to investment in additional infrastructure, or migrate to a commercial cloud infrastructure in order to participate. Similarly, rules and requirements around Application Programming Interface (API) access and information assurance, particularly liability for data breach, might restrict the participation of a certain type of entity in the EOSC due to the costs associated with API development, identity and access management or security audit. One option is to phase-in certain parts of the eligibility criteria, to enable changes to take place as part of a normal system development life cycle, or to provide funding for conversion of certain high value assets to more scalable infrastructures.

Additionally, the EOSC should serve all possible users and not just academic users. Private sector users should be considered stakeholders in the EOSC as well as participants from the start, not added after the fact as is pointed out in Sec. 2.4 "Widening the EOSC" of the SWD. By participating, private sector may want to invest in the long-term development and sustainability of the EOSC, along with the public sector and not just serve to exploit public data for free. Considerations such as access for non-scientific actors as well as non-EU/associated countries researchers are not addressed herein but will be included in the final report. The other category of participants are the funding entities and the brokers utilized by the funding entity to distribute cloud coins. For both of these groups, it is essential to ensure that they follow practices that enforce fair market conditions. For funders, this means treating all providers as equal partners and allowing the users of the EOSC maximum discretion to select resources. Brokers wold be obliged to behave in a disinterested fashion with all providers. Entities that establish brokers must require that the broker does not establish a monopoly, or fall under the control of a service provider that then uses their influence to exclude other service providers from the marketplace.

5.2 Eligibility criteria for actors

There are two sets of eligibility criteria relevant to delivering an EOSC MVE: the first for data and service providers, the second for users. This report focusses primarily on users. These criteria need to meet two objectives: ensure that the data respects the policy goals, specifically FAIR guidelines and information assurance/personal data protection guidelines/regulations, and further, create an EOSC in practice that is appropriately sized and defined to increase data sharing and reuse.

Key rules for participants therefore will include:

1. **Capacity**: The ability to reuse data *in situ* (as opposed to making additional copies of data), and the need to create an appropriately sized EOSC suggests that participating services must possess an appropriate amount of computational, storage and network capacity available to external users. The exact minimums may depend on the nature of the node, the types of analysis likely to be required and the size(s) of data sets likely to be deposited. Such capacity needs to be available to external users, capacity that is only available to local users (for example, members of a specific university) should not be counted. Such capacity could be made available by implementing nodes in commercial or non-commercial clouds or providing access to dedicated computing resources associated with a particular node in the environment.
2. **Accessibility**: In line with work underway by the FAIR data group, a key element of FAIR principles is that data must be accessible, and participating entities will need to have appropriate mechanisms to ensure that this is so. This will likely include a minimum set of interfaces for data deposit and download, as well as capabilities to launch analytic tools against data deposited at the site. These requirements should not require complex APIs or simple Graphical User Interfaces (GUIs), although providers could certainly choose to provide them. Over time, an EOSC MVE may deliver such capabilities, but the initial requirements should be more basic; premised on ensuring that there are understood technical means of interaction with the resource. Further, these requirements are not an assertion that all data and services must be directly available for all at all times, several categories of resource will be subject to data, and the resource must provide means to implement relevant restrictions on access and reuse.
3. **Identifiers and Metadata**: The ability to find, interoperate and reuse scientific content is dependent on the ability to understand the data, software or workflow that is being evaluated for reuse. While maintenance of this metadata is fundamentally the responsibility of the submitter of data or other digital objects. Resources must cooperate with such indexing capabilities (which themselves are likely to be part of the EOSC) ultimately developed under the auspices of the programme.
4. **Information assurance and data protection by design**: Resources will need to support appropriate information assurance activities in cooperation with providers of content. Given that liability for improper access will probably reside primarily with the user/organisation that submits the data or that maintains access controls, a shared security model will be essential. Moreover, compliance with the fundamental principle of data protection by design will be needed. In this respect, appropriate technical and organisational measures, designed to implement data-protection principles (such as data minimisation and protection of data subjects' rights) should be applied in an effective manner to integrate the necessary safeguards in the processing activities.

Participants that access the EOSC to deposit or use data will mainly need to agree through the Rules of Participation of the use of the data or other digital objects, and to FAIR principles. In practice this likely means that data that is deposited in the EOSC must meet minimum metadata requirements. These requirements will vary depending on the scientific discipline and will need to be defined by experts in that field. The role of nodes in the EOSC will also need to be clarified here – should the node that receives the data pass metadata to indices or should it be delivered directly by the depositor? Efficiency would seem to demand for the former, but this remains an open question.

5.3 Participation according to the business model

The development of novel capabilities, long-term storage/maintenance of data resources and fixed cost capabilities are likely to be provided using direct payments to organisations setting up nodes in the EOSC. By contrast, numerous research activities by individual investigators may be supported via EOSC vouchers. Nodes in the EOSC will have to be able to engage with the business model. This will probably imply a business arrangement with the brokers set up by funding agencies in order to accept these vouchers as payment.

5.4 Liability related to service provision

The rules of participation will need to set general principles and define minimum standard procedures with respect to possible liabilities resulting from the provision and the use of EOSC services.

To minimise such occurrences, the EOSC MVE should support identity management and access controls on digital objects (see Section 7), that enable appropriate access to data. In the case of breaches of the terms of use, liability is shared between the submitter of data (who controls access) and the provider for improper release of digital objects.

Providers would be liable for failures in areas under their responsibility, such as compliance with data protection rules, data security, security overall and employees with disproportionate access rights. As the submitters control access, they retain liability for data leakage and to ensure that relevant individuals accessing information meet the necessary requirements.

As regards to data quality and warranties as to fitness for purpose, the EOSC MVE would need to operate under the principal of *caveat emptor*. That is, while submitters may be liable for outright fraudulent data, the nature of scientific research data determines that EOSC data should probably be provided with no warranties for any particular purpose, although Section 5.5 section below, on assessing data quality, should be also taken into consideration.

Data should be:

» processed lawfully, fairly and in a transparent manner in relation to the data subject (principle of 'lawfulness, fairness and transparency');

» collected for specified, explicit and legitimate purposes;

» adequate, relevant and limited to what is strictly necessary in relation to the purposes for which it is processed;

» accurate and, where necessary, kept up to date.

5.5 Data quality

Data quality is likely to be the most difficult element to standardize in any given set of rules of participation, considering that the usual standard of "fit for purpose" varies so much from use case to use case. There are two mechanisms to ensure appropriate data quality. The first derives from FAIR principles, as interoperable and reusable data implies that the data set has a given minimum amount of metadata. Defining an appropriate standard for metadata that can be efficiently defined by data depositors and implemented by repositories and index/search services, will be key to implementation of the EOSC.

The second mechanism is that of peer-review and collective filtering, e.g. Yelp/TripAdvisor-type reviews provided by users. As data become more accessible, it may be useful to provide mechanisms in search systems/indices for users to provide reviews that could be used to supplement citation counts. These reviews could make it easier for data sets that were found to be insufficient for the particular needs of one project or publication to be reused elsewhere. This would also reduce the time taken by users looking for data to examine a data set before deciding to take a deeper look or look elsewhere. Such reviews would also provide an opportunity for groups to indicate issues with data analysis performed in the initial publication(s) and provide an important part of the replication infrastructure currently needed in science today.

5.6 Data security

Data that will be distributed via the EOSC will have different levels of access control depending on Intellectual Property (IP) issues, embargoes prior to publication and personal data protection considerations. In addition, certain types of research may have National Security implications that require additional levels of access control. The only model viable in such an environment is one whereby data security, or more accurately, access control remains with the entity that is ultimately legally responsible for ensuring that the data is properly restricted. This implies a very flexible access control regime, as some data (such as, for example the information underpinning a conventional research publication that does not involve human subjects or touch on National Security issues) should be made open after publication, while information such as human subject research data may need to be explicitly controlled by a data access committee at the organisation that carried out the research. In other cases, a holding entity (for example a data repository) could assume the legal burden for ensuring appropriate access control.

From a more conventional information security standpoint, the EOSC, like most clouds, will need to operate within the framework of a shared security model. That is, the provider of IaaS, PaaS and SaaS, (See Glossary for acronyms) will share the information security duties with the entity that is utilizing the service(s). Again, this could be an individual or a research institution using a generic storage service, or a much more sophisticated environment, where a repository provides value-added services on top of basic IaaS provided by a third party. In these cases, the EOSC entity should follow appropriate best practices for cloud security, including a risk assessment, a data protection impact assessment (as the case may be), implementation of appropriate controls and auditing commensurate with the risk assessment[29].

29View ENISA frameworks at: https://www.enisa.europa.eu/topics/threat-risk-management/risk-management

6. THE ROAD AHEAD TO COMPLETION OF THE 2nd HLEG

The 2nd HLEG in the following months of 2018 will address a more detailed description of the Rules of Participation, it will support the preparation of the June 2018 EOSC summit. Moreover, the implementation roadmap shall be incorporated, and the Coalition of Doers will be stimulated to serve a role in the governance structure. Finally, the HLEG II will address, in the final version of the present report, some additional "long-term challenges", as described below.

6.1 Preparing EOSC Coalition of Doers Summit 2018

One year on from the EOSC Summit 2017, this one-day event is proposed as an opportunity to build upon the achievements and progress being made towards a European Open Science Cloud, and to continue to build momentum.

The event will be organised as a Summit in the morning part a Rules of Participation workshop in the afternoon, covering the following topics:

1. Presentation of EOSC in practice stories (presented by members of the 'coalition of doers')
2. Presentation of the EOSC HLEG II Interim report (by EOSC HLEG members)
3. EOSC Pilot Considerations on the Rules of Participation (by members of the EOSC Pilot project)
4. Panel on the Rules of Participation (with panellists from the 'coalition of doers')
5. Panel on EOSC Business Models, Data Management Policies, Data Security & Legal Issues (with panellists from the 'coalition of doers')

The EOSC Summit 2018 will also serve as an opportunity to **launch a stakeholder consultation** on the draft 'Rules of Participation of EOSC', as a key input for the future EOSC governance. Participation will be by invitation only and is primarily intended for the members of the 'coalition of doers' that emerged from the successful first EOSC Summit held in June 2017 and the wide support to the ensuing EOSC Declaration.

6.2 Support to the Implementation Plan and roadmap for the EOSC

The EOSC is expected to grow into an eco-system providing an open environment for the scientific community to use, provide and share scientific data and results. Basic components of the open cloud environment are being addressed in the various H2020 projects currently funded, as well as: eInfraCentral, EOSCPilot, EOSC Hub, successful INFRA-EOSC-2018 proposals, FREYA, OPENAire-Advance, RDA Europe 4, EOSC-Hub[30], which integrates and consolidates services, software and data from key e-infrastructures with common access mechanisms for the scientific community. The EOSC Portal will provide a single entry point enabling consolidated access to federated e-infrastructures for data-driven science. In addition, the pilots[31] will demonstrate trials crossing borders and scientific domains.

The pilots provide a first step towards growth and additional integration of services, resources and data. A current requisite is a set of rules for participation in the EOSC eco-system through establishment of a compliance framework. This needs to take national science infrastructure roadmaps into account to reach out into a collaborative eco-system instead of fragmented, single, national solutions.

The compliance framework shall outline the rules for participation, how to meet FAIR data principles, how services can be integrated, address security and privacy and include governance practises and risk controls. Its approach and commitments given by strategy and objectives will be overseen by the governance board of the EOSC (as part of the governance model). The policy and process set the rules for tools and processes

30 Horizon 2020 Grant no 777536
31 https/eoscpilot.eu/

which can be embedded into the EOSC eco-system. The organisational structure, human resource mechanisms, change management, performance measurement, training, communication and education cover the usage for the resources and access for people.

There are several challenges which need to be addressed before a compliance framework can be defined:

Challenge-1 enable integration of and access to resources that will be federated in the EOSC

As opposed to developing yet another interoperability framework, it is important to follow an open interface approach, which permits a seamless integration/adaptation to existing Hubs. For resources and services, it is fundamental to build services based on open standards APIs and protocols.

Rules shall be set to build services based on open standard APIs and protocols for resource and service portability reasons.

Access to services shall follow legal security and privacy requirements. This implies that the EOSC provider needs to implement the risk controls for GDPR and eIDAS compliance. **Rules** shall define the compliance statement.

Challenge-2 Cross-border and cross-discipline research collaboration.

It is not only necessary to access and find services or data sets for collaborative work, but also to enable collaboration between ad-hoc communities in the future. These communities might work in a protected space and therefore necessitate individual settings. It is important that the EOSC eco-system provides a collaboration mechanism for ad-hoc research groups with the respective settings for security and privacy. Depending on the needed resources for computing, the system also needs to provide ad-hoc resources for intensive computations.

Rules for collaboration shall include the potential of building dynamic ad-hoc groups with respective security settings, if necessary. The rules shall also include publication of the results as generated in the collaboration project.

The portal shall include social media to allow the researcher to build cross-country and cross-border collaboration.

Scientific data repositories have to follow FAIR data principles. This also includes a FAIR data certification. Here it is important to note that the European Commission is working on a European Cloud Certification[32] scheme as part of the Free Flow of Data regulation draft. As several certification schemes already exist, which can be found on the ENISA webpage[33], it is important to work together and not define yet another scheme. Furthermore, certification for services, infrastructure and resource providers generally involves costs. Therefore, it has to be determined who is pays for the certification, or whether self-certification is sufficient. Resources offered by smaller research entities also have to be taken into consideration, for example, does this have to be the same certification process required by the EOSC marketplace (via the EOSC Hub)?

Rules need to be set for FAIR data compliance but need to take into consideration risk controls already set by resource or service providers.

A new role may have come into play for the federated EOSC eco-system, or rather a kind of national or scientific communities Hub, which will be responsible for compliance and clearing of new services and data sets.

6.2 Endorsements & Commitments From The Coalition Of Doers

The first EOSC summit was held on 12th June 2017. It brought together 110 players, including scientists from a wide spectrum of scientific fields, representatives from scientific infrastructures, research funders and officials of Member States and Associate Countries acknowledging strong support for the implementation of the EOSC. Participants committed to working in areas of data culture and FAIR data. In addition, great heed was payed to the establishment of EOSC services and EOSC architecture. Standards (technical, semantic, legal

32 Digital Single Market Stakeholder Meeting December 2017

33 https://resilience.enisa.europa.eu/cloud-computing-certification

and organisational) were identified as playing an important role in the EOSC. Those present, key actors for the implementation of the EOSC, confirmed the relevant role played by the "coalition of doers" and encouraged momentum to be maintained; this led to the submission of over 70 endorsements/commitment letters as a result of the EOSC Declaration, see the figure below.

Figure 3 – The 70+ signatories of the EOSC declaration

Participation was highly representative, including scientific fields, national scientific infrastructures, research funders and ministries of Member States and Associated Countries. The EOSC HLEG will stimulate the "coalition of doers" to move to the next level and describe their vision with practical examples of an EOSC situation, practice or implementation. This could be an existing initiative, or a vision for future implementation within a timeframe of three to five years. As mentioned above, a first, selected sample of these "EOSC in Practice stories" is included in the annex to the present document.

6.4 Long Term Challenges

A number of long-term challenges have been identified by the HLEG, among which:

- Private sector involvement in the EOSC
- Human capacity development for open science
- Further evolution of procurement models for EOSC delivery
- Addressing Green IT requirements
- Emerging technological areas and priority needs (e.g., Blockchain, ethical AI, cybersecurity)
- FAIR-data-related challenges:
  - What are the criteria for "EOSC-compliant" services, particularly if they are offered from third parties?
  - Are FAIR services needed to deal with FAIR data and if so, what are the criteria for FAIR services?
  - Could criteria for FAIR services define "EOSC-compliance" of services?

More challenges will presumably be added and some of them will be addressed in the final release of the present report, trying to address them vis-à-vis the EOSC, especially trying to identify synergies among currently running pan-European initiatives and best practices.

7. RECOMMENDATIONS

In this section a set of recommendations from the 2nd HLEG are reported, clustered along the three fundamental dimensions of Implementation, Engagement and Steering, as deemed appropriate considering the current phase of the EOSC journey.

7.1 Implementation Recommendations

The focus here is to support practical implementation of the EOSC, with a set of "priority specifications" and other "technical stimuli".

I1 – The EOSC should implement "whatever works" and do "whatever it takes" to **increase the availability and volume of quality & user-friendly scientific information on-line**.

I2 – Define **EOSC interoperability standards** so that services can be interconnected and federated to be as effective as possible and be based on existing open standards.

I3 – Define an **EOSC Quality of Service (QoS) standards**, separate for all elements of the ecosystem (data, data access services, software, etc.), to develop a trustable ecosystem

I4 – Introduce, as part of EOSC's mission, that a **state of the art analysis is carried out on a national level within the Member States** for assessing statistics and key assets around the composition and relevant clustering of the community of users, with the respective eInfrastructures & research infrastructures & scientific communities.

I5 – The universal entry point to the EOSC should provide access to a **marketplace of efficient and effective services**, with lightweight integration (authentication and authorization infrastructure, order management, etc.) where service providers can find service users and vice versa. Nothing is wrong with multiple entry points which should be seen as a plus rather than a negative fragmentation.

I6 – Adopt EOSC standards – from international standards - for information encoding and for protocols for data sharing and publication to **implement FAIR principles over data and services** in a complete & efficient way. And vice versa, offer open EOSC standards to international standardisation bodies.

I7 – Promote the development of services as **independent, interoperable and exchangeable building blocks** to foster the future accreditation of innovative and/or efficient alternatives.

I8 – Promote **open software development**, where possible, for all elements of the EOSC.

I9 – Introduce a **regular assessment of EOSC against other alternatives, including commercial providers**. This could be made to either enhance an EOSC Service, or to support new Services;

I10 – Simplify early (beta) participation in the EOSC by potential data providers, service providers, and underlying infrastructure providers, by relaxing initial constraints but **without relaxing quality standards for data and services**.

I11 – **Resource allocation**, particularly for early stage pilots, needs to be dynamic and amenable to change even at short notice;

I12 – Build a workforce able to execute the vision of the EOSC by **ensuring data stewards, data and infrastructure technologists and scientific data experts who are trained and supported adequately**.

7.2 Engagement Recommendations

One central element for the success of the EOSC, part of the "participation" issue, will be engagement of all the Actors, starting with the researchers. The "Engagement Recommendations" highlight some practical ways of supporting an "EOSC Engagement Strategy"

E1 – Create **career-enhancing incentives** for researchers who open the science that they produce as also indicated by the OSPP, e.g., who lodge high quality, curated data in trusted repositories, share data services to their peers, or develop open software and services, and make the EOSC a portal to those incentives.

E2 – Develop, both at Member State and also EU level, appropriate **engagement schemes** whereby **publicly-funded research infrastructure providers and research communities take part in the EOSC**.

E3 – EOSC should take **national and international developments into account** and should be connectable to national and international frameworks (e.g. the National Research Data Infrastructure in Germany or the AOSP ICT Infrastructure Framework34)

E4 – To **stimulate the "supply side"** of the EOSC, ensure creation of economic incentives for research infrastructure providers to use and co-develop shared facilities and data repositories through the EOSC. This would be supported by the Commission through the European Data Infrastructure (EDI).

E5 – To **stimulate the "demand side"** of the EOSC, ensure establishment of dedicated funding for demonstrations of the EOSC at EU level that would include researchers and their infrastructure providers (e.g., EOSC in practice stories, cross-disciplinary success of EOSC).

7.3 Steering Recommendations

EOSC will need to select and yet receive strategic advice and steering input from a number of sources and the following recommendations highlight a few valuable paths.

S1 – Ensure that the WGs and the other advisory structures well cover for the Executive Board **the latest scientific and organisational trends and novel ideas** for the necessary decisions in those areas.

S2 – Harness **inputs and pledges of the coalition of doers** and support their activities in the context of EOSC implementation.

S3 – Ensure, both at pan-European and international level, that research communities pursue **advanced partnerships**, also supporting with grants and other incentives, so that EOSC's progress is not dependent on the slowest but on the fastest movers.

S4 – Guidelines and rules should be clearly separated into (i) domains for which stability and trustability are important and (ii) domains for which progress must take place rapidly. The former should have rules and **instructions** that remain stable in time, whereas the latter should be run by **living documents**, facilitating innovation and change. All guidelines and rules should be accounted for in the **establishment and progress of the WGs**.

34 http://africanopenscience.org.za/?p=230

8. EOSC IN PRACTICE STORIES: SELECTED TESTIMONIALS OF EXCELLENCE FROM EUROPE

"Musical harmony is based on physical principles, while in cooking, ingredients must be weighed out with precision. At the same time, you have to be able to invent because if one follows the same recipe all the time, you never create anything new. "

Fabiola Gianotti

A sample of some practical "EOSC In Practice" Stories have been collected here and they are reported in ANNEX I – EOSC IN PRACTICE STORIES. These success stories have, where possible, been woven into the Policy, Governance and Implementation recommendations. Highlights from some of these real-case stories will also be showcased at the EOSC Summit 2018 in Brussels on the 11th June 2018.

Throughout the discussions of the HLEG EOSC group to date, **8 features** were consistently raised as prominent aspects thought for a successful EOSC. On the basis of those features, the 2nd HLEG created a template to allow these inspiring stories of EOSC in Practice. The aforementioned features are:

» Research Fields covered;

» Recognition;

» Multiple member state collaboration;

» Trusted services offer;

» Certification;

» GDPR compliance;

» Sustainability model;

» User Experience.

A short description of the facilitating entity is also included in ANNEX I – EOSC IN PRACTICE STORIES, where the compliance with FAIR principles and where possible responding to the why/how have been addressed.

The success stories provide evidence and sustain some of the key recommendations described in Section 7. A multi-dimensional comparison of the "EOSC in practice stories" will be provided in the final release of the present document, also with mapping onto the envisaged EOSC MVE.

9. CONCLUSIONS

"It is the long history of humankind those who learned to collaborate and improvise most effectively have prevailed."

Charles Darwin

The 2nd European Open Science Cloud (EOSC) High Level Group (HLEG) has built upon the work carried out by the 1st EOSC HLEG, leveraging on and supplementing the Staff Working Document (SWD) prepared by the EC. The first part of the work carried out in the period June 2017 through May 2018 has allowed the authors to touch upon a number of crucial elements of the EOSC, from definition of the **Minimum Viable Ecosystem**, to definition of the **main Rules of Participation**, and covering aspects as crucial as Governance and possible Business Models.

The preliminary conclusions reached are well captured by the set of practical **recommendations** for implementation of the EOSC that this interim report bears. The recommendations are clustered along the three fundamental dimensions of Implementation, Engagement and Steering, as deemed appropriate considering the current phase of the EOSC journey.

The definition of the agenda priorities for the **EOSC Summit 2018** has been successfully carried out and the proposal for an **open consultation** to be launched at the Summit has been put forward, also to allow the **Coalition of Doers** to convey their efforts into contributing to the definition of the rules of participation

The remaining months of 2018 will allow continuation of work along the lines of the past effort of this 2nd HLEG, namely with a strong attention to providing support to "prompting an EOSC in practice", as needed and expected by all the EOSC stakeholders. The **final report** "Prompting an EOSC in practice" will be produced in Q4/2018.

ANNEX I – EOSC IN PRACTICE STORIES

EOSC in practice Story #1 toHNSciCloud

Title	Sharing Open Science Services
Facilitating Entity	HNSciCloud

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	http://www.hnscicloud.eu/hnscicloud-user-groups » Euro-BioImaging: European Research Infrastructure for Imaging Technologies in Biological and Biomedical Sciences » CTA: Cherenkov Telescope Array » BBMRI: Biobanking and BioMolecular resources Research Infrastructure » ELIXIR: A distributed infrastructure for life-science information » ESRF Upgrades: Extremely Brilliant Source » European-XFEL: European X-Ray Free-Electron Laser Facility » HL-LHC: High-Luminosity Large Hadron Collider	Large number of end-users working with user cases from 7 ESFRI research Infrastructures
2. Recognition	The role of joint procurement of commercial cloud services in EOSC highlighted EOSC declaration endorsed by more than 70 research stakeholders.	» Helix Nebula services highlighted in e-IRG publication 'Guide e-Infrastructure Requirements for European Research Infrastructures' » Referenced by NIST/IEEE Joint Cloud Federation Working Group as good example of multi-cloud environment
3. Multiple Member State collaboration	Procurement lead by CERN, shared across EU-28 and associated countries	» 10 research performing organisations (including 3 intergovernmental organisations) hosted in 7 countries

Dimension	Description	Outstanding features
4. Trusted services offer	Innovative services addressing the needs of research communities » **Compute and Storage** – support a range of virtual machines and container configurations working with datasets in the petabyte range; » **Network Connectivity and Federated Identity Management** – provide high-end network capacity for the whole platform with common identity and access management; » **Service Payment Models** – supporting the separation between free-at-the-point-of-use access for end-users and their sponsorship by funding organisations.	» Creates a common hybrid science cloud platform for the European research community » Payment model that offers free-at-the-point-use access to the services for end-users sponsored by procuring organisations
5. Certification	Cloud services compliant with international standards. Enabling an ecosystem of trusted and recognised EU suppliers that offer services which adhere to the standards and have passed rigorous testing by the research community	» The ISO/IEC 19086 standard for Service Level Agreements » The ISO/IEC 27000 series of standards, notably ISO/IEC 27017 and ISO/IEC 27018, for information security of the resulting services » The IEEE 1016-2009 system design document template for the design of the hybrid cloud architecture » Services critically assessed for functionality, scalability, performance and availability using a testsuite collectively produced by the research community
6. GDPR compliance	Compliant since October 2017	Commercial cloud service providers have deployed GDPR-ready services as part of the hybrid platform
7. Sustainability model	Procuring organisations contribute financially to the consumption of commercial services for their sponsored end-users.	» Aggregated procurement across multiple research communities leads advantageous pricing and terms and conditions » Contractual relationship between funding agencies and service providers » Early Adopter group for simple on-boarding of new procuring organisations » In-depth Total Cost of Ownership study being performed jointly by research communities and commercial service providers during 2018

Dimension	Description	Outstanding features
8. User experience	End-users from all diverse research disciplines are deploying their applications o this hybrid cloud platform.	» Published lessons learned from the design and prototype phases – available via website (HNSciCloud.eu) » Series of webcast event highlighting domain specific use-cases scheduled during 2018

A summary note from the Facilitating Entity

Since its creation in 2011, Helix Nebula has grown to become a leading public-private partnership between public research actors and cloud service providers. The initiative has undertaken its first joint Pre-Commercial Procurement (PCP) tender called Helix Nebula Science Cloud (HNSciCloud) to create a common hybrid science cloud platform for the European research community.

HNSciCloud hybrid cloud platform links together commercial cloud service providers and research organisations' in-house IT resources via the GEANT network. The platform offers data management and compute services accessible via eduGAIN and ELIXIR federated identity and access management systems with support services, account management facilities, documentation and training.

The capacity procured by the 10 research organisations from the commercial cloud service providers to support these use-cases during 2018 exceeds 20,000 cores and 2 petabytes of storage with a network bandwidth of 40Gbs.

All the services are based on open source implementations that do not require licenses in order to be deployed on the in-house IT resources of research organisations connected to the hybrid platform.

The HNSciCloud platform can be associated with H2020 projects via the early adopter scheme to provide production quality and commercially supported cloud services that are tailored to the needs of the research communities. In particular, the HNSciCloud platform can support the connection of the research infrastructures identified in the ESFRI Roadmap to the European Open Science Cloud (EOSC).

The data management services support the data stewardship needs of research communities and are integrated with the compute services to analyse the data and run simulations.

The commercial cloud services are hosted at data centres in multiple European countries and made available according to a commonly agreed cloud service agreement that respects key national and European legislation to reduce contractual burden as well as geographical and discipline fragmentation.

EOSC in practice Story #2 – Social Data Cloud

Title	Social Data Cloud
Facilitating Entity	CESSDA ERIC

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	Our Social Data Cloud will offer data, tools and training to scholars who follow a data-driven scientific approach. It is by nature cross-disciplinary and of interest not only to researchers in the **social sciences and humanities** but also to those working on **societal challenges**, such as the UN Sustainable Development Goals.	» 100 000+ studies » 50 000+ registered users at CESSDA Service Providers » 500+ people working in CESSDA Service Providers
2. Recognition	A depositor is provided with a Digital Object Identifier (DOI) for the dataset that is deposited – a prerequisite to make the data findable, citeable, etc. CESSDA's Persistent Identifier Policy (2017) serves as a common basis and contains requirements regarding the use of Persistent Identifiers to which the CESSDA Service Providers comply.	» A DOI is automatically generated via the self-archiving tools provided. » Six principles of the PID Policy relating to identifying, locating, resolving, referencing and citation, visibility and flexibility of each CESSDA Service Provider's data holdings. » Best Practice Guidelines provide additional information and guidance on the use and implementation of PID.
3. Multiple Member State collaboration	**16 countries** are members of CESSDA and one is an observer. Further cooperation with service providers from ca. 10 other countries as a direct impact of the CESSDA Strengthening and Widening H2020 project.	» The following entities may become Members or Observers: (a) Member States of the Union; (b) associated countries; (c) third countries other than associated countries; (d) intergovernmental organizations. » CESSDA ERIC is the hub of a distributed research infrastructure. » Each member assigns a national Service Provider, or a consortium of providers, to provide CESSDA ERIC services in their country and across Europe.

Dimension	Description	Outstanding features
		» CESSDA ERIC strives for full European coverage and promotes wider participation in its research infrastructure.
4. Trusted services offer	We provide and develop tools & services for the **deposit, curation and re-use of data following the FAIR principles and elaborate on safe & secure data access and re-use.** We are renowned for our **training** covering the whole research life cycle and aimed at both **data owners and data users**. We also train and share expertise within CESSDA Service Providers & develop new expertise. We offer a joint **CESSDA Products and Services Catalogue** based on standardised metadata to make data Findable.	» Data are Findable and Accessible via the CESSDA Data Catalogue. » Standardisation of metadata helps ensure that datasets are "findable" » Common access procedures will be in place, including safe & secure environments for sensitive data. » Together with research communities, set up and maintain Data Clusters to ensure increased Interoperability and to transition away from a landscape of disciplinary silos. » Training to increase re-use and to contribute to increased professionalism in data management. » Tools & Services for multi-lingual questionnaires, common vocabularies, etc.
5. Certification	Comply with the **CoreTrustSeal** certification, which was launched in 2017 by the Data Seal of Approval & ICSU World Data System. The CESSDA Trust Support Group supports and advises members and aspirant members in all trust-related issues.	» Ensure that Quality Assurance Procedures are in place for data processing. » Reliability of the CESSDA Service Providers towards researchers and as trusted repositories for governments and research funders.
6. GDPR compliance	Guidance and tools whether data deposits meet GDPR requirements. CESSDA, together with other SSH ERICs, investigates consequences for "new types" of data such as social media data and bots.	» CESSDA Expert Seminar 2017 was on "Legal and ethical framework for the use, reuse, and archiving of new types of data". » Establishment of a Legal & Ethical Expert Group

Dimension	Description	Outstanding features
7. Sustainability model	CESSDA has an ERIC structure whereby members are obliged to contribute to the CESSDA ERIC budget, to **designate a Service Provider**, and provide national funding to allow the Service Provider to meet the necessary requirements. On membership - we have a preliminary phase at which we help the country / Service Provider to prepare membership. Membership fees are based on GDP.	» Annual Research & Innovation Budget. » Widening activities to potential new members. » Economies of scale on common infrastructures. » On membership - we have a preliminary phase at which we help the country / Service Provider to prepare membership.
8. User experience	We want to realise an ecosystem where **data, tools, and training** are available for users of the social science and humanities data. The focus is on **users and user feedback** will be gathered for the tools & services and as a general best practice through CESSDA ERIC to ensure transparency and representation.	» CESSDA Training provides direct contact with data depositors and data users. » Internal training and expert seminars for staff from Service Providers. » Communication with other Research Infrastructures in Social Sciences and Humanities and with other ERIC's.

Summary note from Facilitating Entity

CESSDA's overall goal is to bring together the expertise of Service Providers and realise an infrastructure that enables researchers to perform high quality research and that facilitates teaching and learning in the social sciences.

CESSDA ERIC has its Main Office in Bergen and **each Member has assigned a Service Provider as a national social data archive and data services organisation**. Via CESSDA these Service Providers join resources and expertise to realise this top data infrastructure for social science data, and the national Members design and discuss the overarching CESSDA strategy.

CESSDA has joined the 'coalition of the doers' of the EOSC and thereby commits to actively supporting the implementation of FAIR data. This presents a key challenge for CESSSDA not least due to the sensitivity of a number of social science data – requiring safe and secure access. At CESSDA we follow a stepping stone strategy in realising the social science data cloud. We are using the FAIR principles to direct our activities and have already realised the "F" (Findable). The **CESSDA Data Catalogue currently contains over 100 000 studies** and will become available in Spring 2018. We have pathfinder projects on the other principles and on secure access.

Next on our journey to providing FAIR data, we will focus on the legal aspects of Accessibility, on the development of Data Clusters to improve Interoperability, and on tools to increase the Reusability of data – including of sensitive data. We also appreciate the value and the importance of skills and we are therefore broadening the reach of our training activities. We currently provide training to data owners and data users at all stages in the research data life cycle. Recently, CESSDA launched its online expert tour guide on Data Management to support social science researchers and contribute to increased professionalism in data management.

To realise the **Social Data Cloud**, where data, tools and training are made available to data users, CESSDA will join forces with the other social sciences and humanities ERICs (DARIAH, CLARIN, ESS, SHARE) as well as other international organisations. Such a cooperation is vital to ensure the best possible result for users. It also ensures that the expertise across the field is shared and made available. Moreover, it stimulates multidisciplinary collaboration and has the potential for greater social impact.

EOSC in practice Story #3 – Federation of research data infrastructures to fostering cross-disciplinary research data management

Title	Federation of research data infrastructures to fostering cross-disciplinary research data management
Facilitating Entity	Generic Research Data Infrastructure (GeRDI)

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	http://www.gerdi-project.de/ 2016-2019: life sciences, marine sciences, and economics (cross-disciplinary) 2020+: rolling out of the technology to other disciplines	Close cooperation between research communities and infrastructure providers is guaranteed by GeRDI community managers, who serve as an interface between the two groups and enable a continuous and effective dialogue and ensure multidisciplinarity
2. Recognition	GeRDI is highly recognised in German Science Policy and among the German community of research infrastructure providers	
3. Multiple Member State collaboration	GeRDI serves as a model for a future German Research Data Infrastructure and is considered to be a building block of the German contribution to the emerging federated research data infrastructure environment in Europe	» Lessons learned in GeRDI will be provided for the implementation of the EOSC
4. Trusted services offer	GeRDI explicitly seeks to support researchers throughout the successive stages of the research data lifecycle. GeRDI thus provides functionalities associated successively with searching for and finding research data, processing it, analysing it and deriving new data from the analysis	GeRDI's software is based on a micro-service architecture which allows a highly flexible application of single, several or all services supporting different phases of a data life cycle. GeRDI will offer a single point of access from which researchers can define the characteristics of the data that they are looking for, irrespective of » where it is located » from which discipline they are, » any information about their location
5. Certification		.
6. GDPR compliance	The connected research data centres that are federated through GeRDI make sure that research data is compliant with GDPR	

Dimension	Description	Outstanding features
7. Sustainability model	GeRDI has an entire work package which will develop a sustainable business model.	It is planned to base the sustainability of GeRDI on the following three pillars: » In-kind contributions from the research data centers which will be using GeRDI software and services (e.g. staff costs) » Continuous financial contributions for the maintenance of the GeRDI software and services (e.g. through membership fees from the data centers and/or communities) » Public funding (e.g. seed funding for initial investments for storage and computing capacities in the research data centers)
8. User experience	GeRDI's project work is organized along use cases which have been defined with the involved communities. The implementation will continuously aligned with these uses cases. The Community managers ensure a continuous dialog between the research communities and the infrastructure providers..	» ·Published *lessons learned* from the design and prototype phases

A summary note from the Facilitating Entity

The GeRDI project deals with the development of a Generic Research Data Infrastructure (GeRDI). It will provide generic, sustainable and open software connecting research data repositories to enable multidisciplinary and FAIR research data management. This software will be based on common standards and be developed in close collaboration with various research communities to ensure the best match to the requirements of different disciplines. GeRDI will promote a wide usage of its software and thus contribute to establishing an active GeRDI community which will continue to flourish beyond the lifespan of the project. All project results, in particular software, training support, and business model, will form a German contribution to the European Open Science Cloud.

The aim is to enable all scientists in Germany, especially those who hold only small amounts of data, to store, share and re-use research data across disciplines. In a first phase, three pilot data centers supporting the management of research data will be linked up with each other so that research data can be used across disciplinary boundaries, enabling new opportunities for multi-disciplinary research. In a second phase, the developed solution can be rolled-out in Germany and – if appropriate funding is available – serve as a model for a future German Research Data Infrastructure. In particular, GeRDI will be able to support universities and research institutes in providing research data, in linking up their existing data stores and in establishing new research data stores. GeRDI pursues the idea of the European Open Science Cloud, supplementing current efforts to implement infrastructures for research data management and taking them a step further.

The project started in November 2016 and will run three years. It is funded by the German Research Foundation (DFG) with about 3 Mio euros.

EOSC in practice Story #4 – Cost Savings and operational success through joint cloud IaaS procurement in GÉANT; a use case from Irish institute QQI

Title	Cost Savings and operational success through joint cloud IaaS procurement in GÉANT; a use case from Irish institute QQI
Facilitating Entity	GÉANT IaaS Framework

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	GÉANT organised a collective tender, across 36 European countries, for procuring Infrastructure as a Service (IaaS) solutions, enabling 10.000 European Research and Education institutions serving 50 million users, to consume the selected IaaS offerings safely and securely. GÉANT and its National Research and Education Network (NREN) partners awarded framework agreements to Amazon (AWS) and Microsoft (Azure) through a number of resellers, Cloudsigma, Dimension Data, Interoute, ITsoft, KPN, Lattelecom, Telecom Italia Sparkle, T-Systems and Vancis. With this operational cloud portfolio, GÉANT has established a Digital Single Market for cloud services and contributes to the European Open Science Cloud. https://clouds.geant.org/geant-cloud-catalogue/geant-cloud-catalogue-iaas/ A good example of an institution using the GÉANT IaaS tender outcome is Quality and Qualifications Ireland. QQI is an independent State agency founded in 2012 as a result of the Qualifications and Quality Assurance (Education and Training) Act 2012.	» Buying IaaS through the GÉANT framework agreements save institutions money and time. » There are discounted prices available, which apply to all institutions, large and small. Institutions already using cloud services have achieved cost savings of up to 30%, after moving to the framework contracts. » Billing and purchasing models were designed to match financial structures in research and education institutions. Institutions can manage usage and control spending. » Network traffic costs are significantly lower thanks to connecting the suppliers to the GÉANT and NRENs high-performance data network. » End-users can login using their institutional account, through single sign-on with SAML2 support. » Many of the selected vendors allow existing educational licensing arrangements to be transferred (Bring Your Own License).

Dimension	Description	Outstanding features
	QQI is the single authority for a wide range of certifications and quality assurances in many areas. From business to child care, and of course higher education, QQI certifies over 170,000 individuals and approves providers' quality assurance, keeping Ireland on its toes when it comes to local and international quality standards. While many research and education institutions were depending on costly data centre-based solutions and underperforming applications, QQI embarked on a gradual migration of the application to the cloud with Microsoft Azure, which is	
2. Recognition	QQI achieved a 33% cost reduction, after Irish NREN HEAnet helped them transfer their Azure usage to the GÉANT IaaS Framework in July 2017.	
3. Multiple Member State collaboration	HEANet provides connectivity and consultative services for QQI. HEANet is a member of GÉANT and can therefore access, and offer, the IaaS framework, which is available to GÉANT's pan-European membership.	
4. Trusted services offer	Ready-to-use agreements, which comply with EU data protection law.	
5. Certification	» Governing law and jurisdiction local to the customer. » Ensuring GDPR compliance including ISO (and other) certifications and standards » SAML2 single sign-on support	
6. GDPR compliance	QQI was concerned about data protection and security. QQI are pleased with Microsoft's efforts to ensure they pass all relevant security audits. They are confident that QQI could not do better than that on their own.	» The GÉANT framework agreement ensures that there is vendor compliance with EU data protection law.

Dimension	Description	Outstanding features
		» The ready-to-use agreements comply with EU data protection law and were established using the EC 2014/24/EU procurement directive, which allows institutions to use the services directly, without running their own tender / call-for-competition.
7. Sustainability model	QQI's use of the GÉANT IaaS framework, with usage discounts, reduction of data traffic costs and transfer of existing on-premise server licenses, brought the actual total costs down by 33%. Since then, QQI has deployed some other services on Azure, so the cost savings are probably even higher today.	» Services such as this prove that cloud service delivery can ease the work conducted by end-user researchers whilst driving down significant costs. The specialised targeting of such services to researchers who are in need of them will ensure the long-term sustainability of EOSC if delivered in a focussed manner.
8. User experience	QQI quote; "After moving 98% of our infrastructure in Azure, we don't have to worry about the physical hardware and administrative and contract costs are much lower," "Taking all this into consideration, the value of HEAnet facilitating our migration to the cloud is really immeasurable."	» Researchers across Europe can use the services in the GÉANT cloud portfolio as a tool for research. This cloud framework is one of many GÉANT services that can bring added value to the EOSC as a whole.

A summary note from the Facilitating Entity

GÉANT is a fundamental element of Europe's e-infrastructure, delivering the pan-European GÉANT network for scientific excellence, research, education and innovation. Through its integrated catalogue of connectivity, collaboration and identity services, GÉANT provides users with highly reliable, unconstrained access to computing, analysis, storage, applications and other resources, to ensure that Europe remains at the forefront of research.

Through interconnections with its 38 national research and education network (NREN) partners, the GÉANT network is the largest and most advanced R&E network in the world, connecting over 50 million users at 10,000 institutions across Europe and supporting all scientific disciplines. The backbone network operates at speeds of up to 500Gbps and reaches over 100 national networks worldwide.

The IaaS Framework Agreements are the result of a joint procurement 'umbrella agreement' between GÉANT and IaaS Suppliers. These IaaS suppliers were awarded contracts with GÉANT, following a pan-European call-for-competition (tender). The framework agreements are made available through the NRENs and enable their end-user organisations to acquire and use IaaS cloud services via individual contracts, throughout the four-year duration of the agreement. The end-user organisations do not need run a tender of their own. The terms address price, quality, quantity, the procedure for ordering services, and the obligations of the suppliers. Individual contracts (call-offs) can be established via direct award or a mini-competition.

By delivering a digital single market across the European NREN community, institutions are ensured the best-value services. Individual institutions will see that they can obtain the most discounted rates that meet the specific technical requirements of educational institutions. For NRENs, a digital single market ensures an equal playing field within the community, and saves time and resources negotiating the labyrinth of legal terms and conditions. In addition, a wide and diverse offering of such vital services serves to strengthen the relationship between NRENs and constituent institutions.

EOSC in practice Story #5 – Shaping the development of the global Internet of FAIR Data & Services (IFDS)

Title	Shaping the development of the global Internet of FAIR Data & Services (IFDS)
Facilitating Entity	GO FAIR

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	Research fields are represented by GO FAIR implementation networks. As of Jan 25th, 2018 involved research fields are » Chemistry » Geophysical sciences » Marine data » Biodiversity » Rare diseases » Metabolomics » Routing services and core infra » Personal Health Train » FAIR data service (OPEDAS) » Distributed Learning and Training	GO FAIR will enable an early phase in which we work by 'federating the gems'. Crucial activities in any research field can commence with motivated early movers already organised in Implementation Networks without any delay, while adaptations can be made as seen fit by the coordinating networks. At any time, new Implementation Networks can be added and as such, the GO FAIR consortium is entirely open to any research field, inclusive and stakeholder driven.
2. Recognition	GO FAIR is recognised in European Science Policy and among the Dutch, French and German communities of research infrastructure providers	GO FAIR support offices are directly funded by the responsible national ministries in Germany, the Netherlands and France.
3. Multiple Member State collaboration	The Netherlands, in close partnership with Germany, France, and a growing number of additional countries is to foster the coherent development of the global Internet of FAIR Data & Services (IFDS), with the main focus on early developments in the European Open Science Cloud (EOSC). A key instrument to achieve this is to establish GO FAIR Implementation Networks (INs) in the European Member States and beyond. So far, six GO FAIR INs have been established or are under development, coordinated in four European Member States.	GO FAIR follows a bottom-up open implementation strategy for the technical governance and funding needed to establish the first phase of the European Open Science Cloud (EOSC) as part of a broader global Internet of FAIR Data & Services. The approach is largely based on the EOSC communication and the recommendations of the High Level Expert Group.

Dimension	Description	Outstanding features
4. Trusted services offer	GO FAIR provides input and impetus for the development and implementation of cross-disciplinary scientific standards, supports the alignment with FAIR Principles and promotes synergies to develop best practices with the aim of turning them into widespread applications. In this respect, GO FAIR organises workshops and sets up working groups so that discipline-specific knowledge and interdisciplinary approaches about the FAIR use of research data and services can be exchanged. These platforms will ensure interconnectivity regarding the establishment and operation of existing and future research data infrastructures.	GO FAIR fosters public private partnership and wants to ensure that no monopolies develop
5. Certification	GO FAIR may organise the certification of services, tools, organisations, and people (including data stewards). This would help research funders and other stakeholders to promote open science, for instance by enabling researchers to incorporate a certified service in their data stewardship plans. GO FAIR's potential role in certification is still the subject of debate within the GO FAIR community.	The stakeholder community will define the 'certification scheme' for a certain service or person. The GO FAIR International Support & Coordination Office (GFISCO) will support this process and act as the 'certification scheme owner' at least during the first phase of GO FAIR. In each participating country, one or more certification bodies can be engaged or started, to certify for instance data stewards (including exams), repositories, algorithms, core resources, registries, ontologies, etc. These certification bodies should apply for accreditation from their national accreditation body to ensure their independence and competence, while performing certification of specific elements of the Internet of FAIR Data & Services. They will have a contract with the Certification Scheme-owner (initially the GFISCO) and are allowed to charge the requesters (companies, service providers, people). All this can be regulated by country and by funder. It is conceivable that the GFISCO will hand over the ownership of the certification schemes to another party as soon as the scheme is mature enough.

Dimension	Description	Outstanding features
6. GDPR compliance	The connected research data centres that are federated through GO FAIR make sure that research data is compliant with GDPR	The PHT IN specifically addresses this in a generic way (also beyond personal health data) by enabling self-control of how personal data can re reused.
7. Sustainability model	The offices are funded by the governments, but implementation networks are not funded through GO FAIR but bring their own resources and hopefully will be more succesful in acquisition by speaking with one, authoritative voice for their component	GO FAIR is a movement to get the Internet of FAIR data going and also contributes to the EOSC, but GO FAIR as such needs no sustainability plan as the best that can happen to it would be that 'the lighter' can be stopped when the fire is blazing (read: the IFDS is self sustaining like the current internet)
8. User experience	GO FAIR is a member state-up movement which brings together various different disciplines to jointly collaborate in implementation networks. Through this approach, user experience is by the nature of GO FAIR part of in every GO FAIR activity.	

A summary note from the Facilitating Entity

The GO FAIR Initiative is a movement for putting the European Open Science Cloud (EOSC) into practice in the broader context of the Global Internet of FAIR data and services. It aims to make research data findable, accessible, interoperable and reusable across national and disciplinary boundaries. In Europe (EOSC), The United States (The Commons), Australia (AORC), and Africa (ADIRC, EAOSCH), efforts to collaborate are already underway to prevent renewed silo formation and to ensure international interoperability. GO FAIR has a very open, bottom-up approach that allows to include all research fields and member states, thus enabling common use and reuse of research data in the future. Since most research fields today are interdisciplinary, this adds significant value to the advance of knowledge in data-based research.

Germany, France and the Netherlands perceive the GO FAIR Initiative as a suitable pathfinder for the European Open Science Cloud and an Internet of FAIR Data and Services. In a joint position paper, Germany and The Netherlands announced their intent to each establish a support and coordination office, operating internationally to kick-start the GO FAIR Initiative successfully. Frencj joined at an early stage. The three offices are tasked with building a worldwide network of academic communities and promote the acceptance and implementation of the FAIR principles. The core of GO FAIR is a federation of existing topical networks of excellence that collectively commit to the FAIR approach and capitalise on their critical mass to make choices in the implementation of the FAIR principles regarding standards, protocols, and best practices. The GO FAIR implementation approach is based on three interactive processes: building the technical infrastructure (GO BUILD) is complemented by a change programme involving relevant stakeholders (GO CHANGE) and training the data stewards capable of providing FAIR data services (GO TRAIN).

Close collaboration and continuous exchange with the numerous initiatives and working groups (e.g. RDA, EOSC Pilot, OSPP) at national and international level in the area of research data management is vital to ensure mutual interconnectability and to help build an EOSC accepted by the member states and an internet of FAIR Data and Services. The expected impact and benefits are substantial. Users will be able to search and analyse linked FAIR data sources much more efficiently, in turn supporting and enabling more effective research, the potential to discover unexpected associations and developments, and ultimately to expedite innovation.

EOSC in practice Story #6 – OPENCoastS

Title	OPENCoastS
Facilitating Entity	LNEC, LIP, Universidad de Cantabria, CNRS/Univ. LaRochelle

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	The OPENCoastS service builds on-demand circulation forecast systems for user-selected sections of the North Atlantic coast and maintains them running operationally for the timeframe defined by the user. This daily service generates forecasts of water levels, 2D velocities and wave parameters over the spatial region of interest for periods of 72 hours, based on numerical simulations of all relevant physical processes. Forcing conditions at the boundaries and over the domain are defined by the user from global forecast databases (including global Copernicus-based CMEMS circulation forcings) or by user-uploaded files. Automatic comparison with real-time in-situ sensor data can be provided for a number of user specified locations.	» The first EOSC example in coastal engineering in Europe » Wide range of applications (research, teaching, consultancy and coastal zones management) and users: scientific and academic communities, SMEs and stakeholders operating in the coastal zones. » Innovative service providing on-demand circulation forecast systems as-a-service.
2. Recognition	The service is supported by INCD the Portuguese National Distributed Computing Infrastructure funded by the Portuguese Science Foundation. The partners are internationally recognized research groups and organizations.	
3. Multiple Member State collaboration	Yes, in the scope of EOSC-Hub. Integration in EOSC resources and services is being done by Portuguese, Spanish and French partners. The research user community that will explore the service is international.	» Available for use in European Atlantic coasts
4. Trusted services offer	Innovative service addressing the needs of multiple communities » **Researchers** –OPENCoastS can contribute directly to the development of new research methodologies and workflows regarding water quality, biological, biochemical and coastal erosion studies.	» Create a standard for circulation forecasting in European Atlantic waters that can be tested and evaluated by all

Dimension	Description	Outstanding features
	» **SMES consultancy** – they will benefit from these operational systems to feed their own higher-level service portfolios to respond to other societal needs, without the need to invest time and resources to deploy forecast systems from scratch.; » **Port and coastal authorities** –This platform will provide coastal managers all the information required to fulfil their responsibilities (examples of uses include facilitating navigation, reducing port operation costs, reducing emergency planning and response of coastal hazards, and better exploring recreational uses of the coast)	
5. Certification	Compliance with international standards	» Will be FitSM compliant
6. GDPR compliance	The service is GDPR compliant.	
7. Sustainability model	Community support. Existing national users have been supporting this service for the past 10 years. Support letters at proposal stages anticipate the interest at global level. Potential for future pay-per-use for commercial engineering applications.	
8. User experience	Users have been supporting the development and using the service for the past 10 years	

A summary note from the Facilitating Entity

Coastal forecast systems have start to be deployed by LNEC in 2008, following a long term collaboration with CMOP (USA), integrating IT and coastal modelling competences to address a recognized demand by the coastal managers and port authorities.

Currently, the service is part of LNEC's catalogue of site-specific research and advanced consultancy services (http://www.lnec.pt/hidraulica-ambiente/en/core/information-technology-in-water-and-environment-research-group/activity-2/ , http://ariel.lnec.pt). In this scope it has been deployed at several sites (Atlantic ocean for waves and water levels, Aveiro lagoon, Tagus estuary and Lisbon drainage system and Faro lagoon). These services continue to be used by research partners after the end of these projects for their educational and research activities, as well as by the involved end-users for their management tasks (National and Local civil protection agencies, Port authorities). The site-specific service is already part of the set of services supported by the Portuguese NGI - Portuguese National Distributed Computing Infrastructure (INCD), integrated in their catalogue of services. LNEC's wave forecast service is also integrated in the National Authority for Maritime Resources (DGRM) ocean data portal.

In spite of the importance of coastal forecasts for both coastal management and coastal research, generic, site-independent tools to build these services are not available yet at European level. Current efforts are concentrated in site-specific deployments that are often of restricted access, promoting duplication of similar results in several providers and the need to build vast teams with little production gains.

The proposed OPENCoastS - On-demand Operational Coastal Circulation Forecast Services, integrated within European resources, will be of great importance for coastal managers, public institutions, research

groups and private sector with responsibilities in emergency and monitoring purposes across Europe. It takes advantage of LNEC's team long-term work on coastal modelling as model developers, integrated in several open source modelling communities (SCHISM, ELCIRC, SELFE, ADCIRC) as well as coastal science researchers tightly coupled to advanced competences in developing forecast frameworks and operating forecasts deployments.

EOSC in practice Story #7 – Lifewatch

Title	LIFEWATCH AlgaeBloom: from monitoring to prediction
Facilitating Entity	LifeWatch-JRU-ES plus ECOHYDROS (SME)

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	Water management, Biodiversity (algaebloom is the most serious problem in lakes and water reservoirs), Climate Change. AlgaeBloom is an example of a complete case study combining: » data collection from on-line complex instruments (meteo station, CTDs, radiometers, etc.) placed in-situ (water reservoir) » curation and processing of data » publication of data (for water management authorities) » re-use of data as input to large model for simulating the hydrological and biological conditions leading to algae-bloom in a water reservoir » requires HPC resources (large simulation, for large grid at a realistic scale, 50x50 m cells covering several km2 surface, and 1m step in depth, up to 40m.) and also large storage » integrates external data (meteo, radiation, bathimetry)	Combines the use of HPC resources under a cloud model, for simulation and prediction, with the access to databases and datasets in the cloud. The databases connect directly to the instrumentation, and implement several stages of the data lifecycle: collect, curate, process, up to ingestion. High added value: the platform has been reused for other similar projects, integrating instrumentation, data curation, and data publication that is then used as initial input to large HPC-cloud simulations. Full support from the consortium.
2. Recognition	Integrated in LifeWatch Thematic Services for EOSC-Hub Developed in hand-hand collaboration with an SME (Ecohydros) Open source Validated with real data Presented at DELFT-days In production since 2015 Paper accepted and presented at EGU (European Geophysical Union meeting)	Use Case will be further developed with the integration of satellite (SENTINEL) data in EINFRA funded project. Connection to OpenAire via the use of data in DIGITAL.CSIC: historical data to understand the evolution of the clima at given sites, to assess future impact.
	Poster at DI4R 2016 (Krakow) In use in LifeWatch community Used by LIFE+ project ROEM+ Extended in several master and graduate projects. Integrated into the official DataScience Master UIMP-Universidad de Cantabria	

Dimension	Description	Outstanding features
3. Multiple Member State collaboration	IFCA (Santander, Spain) SME: Ecohydros LifeWatch ESFRI: Spain, Italy, Netherlands, Portugal, Belgium, Greece, Slovakia	» First line research institutions in Europe, including an SME (Ecohydros) » Well connected to EOSC-Hub » Contributing to eINFRA
4. Trusted services offer	Access to online databases for stakeholders (including water management authorities) DELFT-3D software (open) applied to water reservoir hydrological and biological modelling	Access to HPC resources and online databases
5. Certification	Quality Assurance using standard EOSC-Hub procedures, adapted from INDIGO-DataCloud.	In-house quality assurance expertise
6. GDPR compliance	Cloud services at public research institutions and compliant with GDPR.	
7. Sustainability model	The sustainability model is supported by the SME Ecohydros, that has included the ALGAE-BLOOM in their product portfolio. The customers of Ecohydros are the management authorities of river basins. These public bodies issue public tenders to support monitorization of ecological parameters. The SME applies to those public tenders, with a portfolio of services, in which Algae-Bloom is a successful product. It is currently part of the services included to support monitorization of lakes and dams belonging to the basin of the Ebro, Tajo and Duero rivers in Spain.	Here follows a description of the current customer portfolio of ECOHYDROS using this application: 1. Lake of Sanabria (Leon-Spain): A lot of information of all kinds (and models) is being generated. There is no continuous data, it is all manual. Data from satellites could be very interesting for the vegetation maps of the basin and especially for the growth of macro-algae and vegetation on the shores, which could be a very relevant issue. 2. Dam of Beche (Galicia-Spain): Here there are blooms of toxic cyanids, that we are monitoring. 3. El Val Reservoir (Aragon-Spain): The Ebro River Administration wants us to progress towards a hydrodynamic and reservoir quality model that covers at least the period from May to October. Here there are blooms of phytoplankton.
		4. Alarcón Reservoir (Cuenca - Spain): An emblematic case, which is at the heart of the controversial Tajo-Segura transfer. There are no blooms of algae but there are problems with the drought and the associated risks of fish mortality. 5. Laguna Negra (Soria - Spain): It has phytoplankton growth above normal.

Dimension	Description	Outstanding features
8. User experience	In use by SME biologists.	

A summary note from the Facilitating Entity

LGAE-BLOOM for LifeWatch is an outstanding example of how the "data cloud" can address a very relevant problem: it integrates data collection from a remote monitoring platform, with its use for predicting several weeks in advance the possibility of algae-bloom episodes, understanding the main reasons (like for example low water level or excess of nutrients due increased population) allowing to explore different alternatives to solve the problem in advance.

This is an outstanding example of how EOSC can help to provide SME researchers, not experts in computing, to powerful simulation models that require performant computing resources.

Moreover, the service can be easily integrated into new initiatives, and in the Data Science education curricula.

These performant techniques are very much in demand by the industry

Fully open project.

EOSC in practice Story #8 – OpenAIRE

Title	Science. Set Free. Open Science: participatory and co-designed by default
Facilitating Entity	OpenAIRE

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	OpenAIRE is a scholarly communication e-Infrastructure and as such, **it is by nature cross-disciplinary**. At the same time, **being a socio-technical infrastructure**, with the former having the **institutional setting at its core**, OpenAIRE provides an effective channel of research support mechanisms. OpenAIRE reaches out to researchers helping them to open up and share their research from the early stages, where data is being generated and processed. It sets in place a **decentralized and interoperable infrastructure of services** to help researchers, content providers, research administrators and policy makers **to implement open science tailored to their needs.** Spearheading open access to publications, OpenAIRE supports all open science activities: making data FAIR and open, linking and sharing publications and data to software, protocols and methods, opening up research to the public and to industry.	An active network of **34 organizations/experts in Europe (National Open Access Desks)** facilitating a culture change to open and FAIR science through a **multiplier effect:** in 2017 alone OpenAIRE NOADs contacted over 250 content providers and were present in 125 events across Europe, organized 46 national workshops, and coordinated 12 multi-stakeholder workshops on open science with an attendance of 2,000+, making sure open science is in all agendas and is effectively implemented. OpenAIRE connects 1,100+ repositories (literature + data), national aggregators, OA Journals and produces a **curated scholarly communication graph** with 24 mi OA publications, 600K data objects, 3K software, linked to grants from 20 funders, thus becoming **the authoritative OA dataset in Europe**.
2. Recognition	OpenAIRE has strongly influenced the landscape of scholarly communication in Europe and the world and has been instrumental in putting in practice EC's Open Access policies. On the global scene, OpenAIRE is Europe's voice in the "**Aligning repositories worldwide**" initiative by the Confederation of Open Access Repositories, COAR, with representatives from similar networks in China, Japan, Australia, US, Canada, Latin America and Africa.	OpenAIRE is a major European e-Infrastructures, responsible for shaping 34 open access policies and coordinating their implementation at national and institutional level. The OpenAIRE guidelines for metadata exchange have been adopted by many stakeholders around the globe, while key national or regional initiatives base their aggregation services on OpenAIRE's software or backend services, some operational, others in their design phases in Latin America, Poland, Spain, Argentina, Chile, Turkey, Greece, Canada.

Dimension	Description	Outstanding features
	OpenAIRE data is increasingly recognized as the authoritative database for EC's and Europe's OA publications and it is retrieved by public and commercial organizations. It is extensively used in policy making, more notably in FP7 and H2020 evaluation.	OpenAIRE is viewed as a successful model of decentralized operations, participatory and co-design, and is consulted in setting up similar initiatives around the world.
3. Multiple Member State collaboration	OpenAIRE's power lies in its **34 member network**, the National Open Access Desks (NOADs). They are experts from key national organizations, familiar with national and local developments, cultural, governance and financing systems in their specific jurisdiction, thus providing **support and solutions (infrastructure)** for open science policy implementations, and wider open science issues, such as aligning and transferring **practices and innovative ideas** for research publishing practices.	OpenAIRE's dual capacity as a social and technical infrastructure readily **provides the link** to and follow up on national policies and initiatives to streamline with European/EOSC policies, hence contributing to a more seamless European research environment.
4. Trusted services offer	OpenAIRE services support, *accelerate and monitor* the implementation of Open Science. A rich service portfolio (15+ services) targets all players involved in the research process and implements services at several e-Infrastructure layers: **LOCAL: A distributed Open Science Helpdesk** implementing interoperability in policy, technical, legal, organizational aspects, for OA publications, OA/FAIR research data and software. The Helpdesk is accompanied by an extensive, hands-on **training** (*train the trainers*) programme building Europe's human data and open science skills capacities. **MIDDLE: Open Science accelerator (brokering)** services providing connectivity in research, i.e., they effectively allow information exchange and distribution across repositories and databases building a linked open science environment. **UPPER: Value added services** for integrated research monitoring, reporting and research analysis, bringing Europe's research publications and data to 3rd party service providers.	**Making sense out of science** OpenAIRE's **discovery portal** (EXPLORE) serves 1 mi unique users every year, while its data is indexed by commercial library services and the global science gateway worldwidescience.org; **Zenodo** is a trusted repository with a global outreach hosting 350K objects, the biggest software DOI minter; **ScholExporer** has 18 mi bi-directional links, from 1000 publishers and 10 data centers, and is ready to serve millions of daily requests; **Amnesia**, an anonymization tool promoting a safe & secure environment for sensitive data, is already requested by health data related infrastructures; the **Validator** has been operated in Latin America increasing literature metadata interoperability in the continent; OpenAIRE's **APIs** are **daily** used by 1000 machines (with a number of requests ranging anywhere from 100K to 1mi) whose aim is to re-use the curated scholarly communication graph; OpenAIRE's backend engine (**D-NET**) has been operating national platforms in 5 countries.

Dimension	Description	Outstanding features
		Open Science as a Service – OpenAIRE's **Research Community Dashboard**, a suite of services for contextualized research view and monitoring, allows research communities to build their own thematic data portals addressing all research artefacts: publications, data, software, protocols, methods. It is used by 5 pilot communities, with 3 more to follow
5. Certification	The OpenAIRE Guidelines are critical to the interoperability, findability and accessibility of the wide range of European research. They are used for metadata exchange of: » Literature repositories & OA journals » Data repositories » CRIS systems » National/Thematic aggregators » Software repositories (pending)	Apart from European repositories, OpenAIRE **Guidelines** for literature repositories have already been adopted in Latin America and in Japan. OpenAIRE operates a **Validator** service which checks data sources for metadata and protocol compliance to OpenAIRE guidelines or other metadata schemata.
6. GDPR compliance	The role of OpenAIRE is to bring the GDPR process of compliance to services in Universities and Research Performing Organizations. With a targeted activity that has just started in 2018, OpenAIRE has a **dedicated task force on legal issues to recommend best-practices for legal compliance**.	**Amnesia** is one service used to anonymize data. Plans are to make it GDPR compliant.
7. Sustainability model	OpenAIRE has been running as a project since 2009, but the imminent establishment of the OpenAIRE Legal Entity will ensure a **permanent structure for a Europe-wide national policy and infrastructure implementation**. Following a hybrid model of member organization and member state representation, the OpenAIRE legal entity will be the foundation for national coordination on open access and open science.	OpenAIRE members are already establishing links at the national level for open science coordination synergies, participating or leading working groups which pop up organically and are supported by ministries.

Dimension	Description	Outstanding features
8. User experience	OpenAIRE has been a human-centric infrastructure from its onset, embedding concepts like participatory and co-design in the DNA of our approach and our services. Having institutions and research communities at the forefront, OpenAIRE reaches out to researchers and all players with a stake in the research process at the local level - where it matters - effectively being a broker of user requirements and surfacing them to the European and global level.	OpenAIRE services target a variety of users (researchers, libraries, funders, publishers, repository managers, policy makers, schools), and its services are used by thousands on a daily basis. The pending, large scale and continuous "customer experience" feedback programme, will centralize the user experience. In addition, OpenAIRE's *train-the-trainer* approach allows the NOADs to be in contact with users in the field, which essentially brings the tailored, not one-size-fits-all open science implementation to the center of developments.

Summary note from Facilitating Entity

Open Science needs pragmatic, participatory infrastructures to make it work. The diverse research communities and cultural variety of Europe will accept no "one-size-fits-all" solutions. While some see such variety and diversity as an insurmountable barrier, for OpenAIRE it is the foundation upon which we build.

In order to be effective, policies must go hand-in-hand with infrastructures that implement and support them. OpenAIRE bases open policy implementation on Europe's **repository infrastructure, leveraging** the **existing investments** of institutional, thematic repositories and OA journals (both literature and data). All key stakeholders (researchers, libraries, administrators, funders, ministries) are involved in this process, showing that **de facto member states already share responsibilities, coordination and costs**.

Big science operates out of a few excellence centers around Europe, but most science takes place in universities and RPOs. The following **key points** state this and OpenAIRE's relation to EOSC:

» To reach researchers as the main users of EOSC, a **language of scholarly communication** should be spoken. Excellent researchers do not have time to design infrastructure or temper technology but are eventually targeted towards communicating with peers through publishing any forms of research results, which contribute to their career development.

» Making **existing, funded localized services** the cornerstones of EOSC architecture, which can efficiently mitigate risks of EOSC becoming an artificial *de novo* construction. Making efficient use of the investments that governments already provide for institutional services in Universities and RPOs (libraries, IT Services, data centres) thus implementing the principles of "standing on the shoulders of giants" and solving the well-known communication problem of the "last mile". Fundamental IT and repository services are successfully serving science today and the technological, financial and governing interface between those and EOSC need clear specification.

» **An EOSC "user" is contextual**: the primary end-user may be defined as the "researcher" – there are, however, local infrastructure services (e.g. libraries and IT Services) who are 'users' of EOSC, in that they provide or support EOSC-related services locally in the institutional context of Universities and RPO's with technology adaptation, training or helpdesk activities. **Infrastructures are governed and funded locally.**

EOSC in practice Story #9 – ELIXIR

Title	ELIXIR Compute Platform
Facilitating Entity	EMBL-EBI, CSC, CESNET

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	Life sciences	
2. Recognition	ELIXIR is a landmark on the ESFRI 2016 roadmap	
3. Multiple Member State collaboration	The compute platform work is coordinated by EMBL-EBI (UK), CSC (FI), CESNET (CZ) with members contributing from several other ELIXIR national nodes.	
4. Trusted services offer	1. Federated Cloud for life sciences: an open federation of cloud providers, offering IaaS cloud resources with pre-deployed applications and datasets for life science researchers. 2. ELIXIR Authentication & Authorisation (ELIXIR-AAI) services to establish a common federated ELIXIR identity and provide managed resource permission authorisation services to the ELIXIR community 3. ELIXIR Data storage and replication services provide seamless access to reference datasets across ELIXIR cloud sites.	» ELIXIR Authentication & Authorisation (ELIXIR-AAI) services to establish a common federated ELIXIR identity and provide managed resource permission authorisation services to the ELIXIR community
5. Certification	Cloud providers integrated into the ELIXIR Compute platforms are self-certified and compliant with the needs of the communities that they serve using industry standards for IT-Security where needed. ELIXIR AAI supports a majority of academic IdPs through the eduGAIN network and industry standard SAML and OpenID authentication mechanisms for authentication. Additionally, ELIXIR is also working with GÉANT and with the AARC2 project to build a sustainable AAI infrastructure for ELIXIR and other BMS RIs.	
6. GDPR compliance	ELIXIR complies with the GDPR through the compliance of individual ELIXIR nodes, their member institutes and any associated service providers.	

Dimension	Description	Outstanding features
7. Sustainability model	The service offers are operated from local funding by ELIXIR node member institutes or supported by the ELIXIR-Hub as commissioned services. Further development of the offerings is funded from EC projects or supported by the ELIXIR Hub through exploratory implementation studies.	
8. User experience	ELIXIR Compute Platform supports seven communities from diverse domains in life-sciences, that help drive the technical developments within the platform.	

A summary note from the Facilitating Entity

Data analysis on life sciences, which is the focus area of the ELIXIR Compute Platform, is a fast moving field.

During the first half of 2017 ELIXIR identified a set of 'Core Data Resources' that are of great significance to the European life-science community from the available data resources in the community. Having these ELIXIR Core Data Resources available for researchers locally on cloud platforms across Europe would allow computing pipelines to access these data resources with lower latency thereby increasing the throughput of their analysis activities. The local processing will also support analysis of datasets that for legal or similar reasons cannot be moved elsewhere.

Within the EGI-Engage and ELIXIR-Excelerate projects ELIXIR members established a federated cloud infrastructure for bioinformatics for users possessing ELIXIR IDs. The infrastructure allows the developers of bioinformatics analysis software to deploy applications in any of the clouds that participates in the infrastructure, to make those applications available for researchers. This infrastructure is further developed within the EOSC-Hub project, enabling Core Data Resources to be pre-deployed within the partner clouds via the Reference Data Set Distribution Service, simplifying user analysis applications to work with life science reference data.

The integrated, scalable compute-data setup will support life scientists in achieving FAIR research outputs by:

» Offering discovery services of the federated clouds, the replicated datasets and pre-staged applications > Findable

» Building the federation on community-maintained, open interfaces and technologies > Accessible

» Working towards standardised description of applications (VMs, containers, appliances) datasets and cloud providers > Interoperable

» Supporting workflow tools to facilitate machine-driven reuse of pre-staged applications and datasets in analysis > Re-usable

EOSC in practice Story #10 – CLARIN

Title	CLARIN - European Research Infrastructure for Language Resources and Technology
Facilitating Entity	CLARIN ERIC

Dimension	Description
1. Research fields covered (cross-disciplinarity)	Humanities, Social Sciences and beyond
2. Recognition	CLARIN ERIC is officially recognized by ESFRI as a Landmark
3. Multiple Member State collaboration	Austria, Bulgaria, Czech Republic, Denmark, Dutch Language Union, Estonia, Finland, Germany, Greece Hungary, Italy, Latvia, Lithuania, the Netherlands, Norway, Poland, Portugal Slovenia, Sweden (members) France, United Kingdom (observers)
4. Trusted services offer	Most language resources and tools are directly provided via the CLARIN centres. On top of that, CLARIN ERIC offers several services through EOSC-hub: The **Virtual Language Observatory**, a facet browser for fast navigation and searching in huge amounts of metadata. The **Virtual Collection Registry**, a service that allows researchers to create their own citable digital bookmarks. The **Language Resource Switchboard**, a web application that suggests language analysis tools for specific data sets.
5. Certification	CLARIN-compliant repositories have gone through a strict internal certification procedure and are certified externally by the CoreTrustSeal (formerly Data Seal of Approval)
6. GDPR compliance	CLARIN ERIC complies to the GDPR.
7. Sustainability model	The majority of operations, services and centres of the CLARIN infrastructure is provided and funded by CLARIN ERIC members (and observers). They set up a national consortium, typically consisting of universities, research institutions, libraries and public archives, of which at least one has the status of CLARIN centre. The contribution expected from members and observers is to create and provide access to digital language data collections, and digital tools and expertise for researchers to work with them.
8. User experience	Researchers are the driving force behind the CLARIN infrastructure. This connection is even strengthened through the user involvement platform and activities.

A summary note from the Facilitating Entity

CLARIN is a European Research Infrastructure providing access to language resources and tools. It focuses on the widely acknowledged role of language as social and cultural data and the increased potential for comparative research of cultural and societal phenomena across the boundaries of languages. Access to language data is crucial for scholars in the social sciences and the (digital)

humanities. Language is a carrier of cultural content and information, both synchronically and diachronically. Language also plays a role as the reflection of scientific and societal knowledge, as an instrument for human communication, as one of the central components of the identity of individuals, groups, cultures or nations, as an instrument for human cognition and expression, and as an object of study or preservation. These insights underline the vision of CLARIN: all digital language resources and tools from all over Europe and beyond are accessible through a single sign-on on-line environment for the support of researchers in the humanities and social sciences.

To achieve this, CLARIN is creating and maintaining an infrastructure to support the sharing, use and sustainability of language data and tools for research. This infrastructure is designed with the FAIR-principles in mind.

Findable

CLARIN's rich CMDI metadata ecosystem provides a sophisticated framework to describe language data and tools. Together with clearly defined protocols in terms of metadata exchange and a powerful search engine, this leads to an optimal findability.

Accessible

All metadata provided via CLARIN is openly accessible. For those language resources and tools that cannot be provided via open access (because of copyright or privacy concerns), easy single-sign on is provided. Persistent identifiers help to ensure the sustainability of the access.

Interoperable

Standardized vocabularies and well-described mechanisms for establishing links between metadata and Digital Objects, also relying on the results from CLARIN's close collaboration with the Research Data Alliance and the C2CAMP initiative, result in an enhanced level of interoperability. Tools like the Language Resource Switchboard – provided via EOSC-hub – illustrate this in practice.

Re-usable

Data sets and tools should come with clear licenses and provenance information. In combination with community-proven standards, this paves the way for high levels of re-usability.

EOSC in practice Story #11 – ARDC, the Australian Research Data Cloud

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	Research fields are supported through precursor activities of the Australian National Data Service (ANDS)	Building upon the approach being taken by EOSC, Australia has a roadmap of research infrastructure that includes the recommendation of establishing an Australian Research Data Cloud (ARDC). This has commenced by integrating three existing initiatives – a data focused (ANDS), a data services focused (RDS) and a collaboration focused (Nectar) being combined to provide an investment that is focused on building four strengths for Australian research: data, collaboration, innovation, and translation.
2. Recognition	This initiative is recommended in a 10 year roadmap, and the first steps in 2017-18, with a formal commencement in 2018-19	This initiative is supported by the Australian Government through its National Collaborative Research Infrastructure Strategy
3. Multiple Member State collaboration	The intent of establishing the ARDC is to form infrastructure of a similar form to EOSC so that Australian researchers can more easily collaborate internationally, but particularly with EU member states who also have similar infrastructure making it much easier to collaborate	International engagement in marine science is a very good demonstration of how research is strengthened by international collaboration, delivering much research of greater impact as ocean systems can be studied in their entirety.
4. Trusted services offer	A very important component of the ARDC will be a network of trusted research data repository services that are compliant with international standards. The RDA developed standard for trusted repositories – the CoreTrustSeal Trustworthy Data Repository certification. A number of these services have already been established – both for domain repositories and institutional repositories.	Using an internationally agreed seal of approval, it makes it much easier to validate an approach being undertaken for any organization. Access to many of the services offered are free to all participants: data publication support, training, policy, etc. Those services that have to be allocated – in particular cloud services, data storage, software development, are allocated on the basis of need, and ability to deliver transformational outcomes.
5. Certification	The CoreTrustSeal Trustworthy Data Repository certification is used.	This approach has strong European, and international backing, so makes it more useful.
6. GDPR compliance	This is not currently relevant to Australia	

Dimension	Description	Outstanding features
7. Sustainability model	The ARDC is intended to be funded for 10 years under the NCRIS scheme funded by the Australian Government, as the value of the investment to the country as a whole is important to the whole of the country.	An Australian investment in a research data cloud is made much more sustainable by doing it collaboratively with international partners, so that investments are shared and made more robust to changing environments.
8. User experience	The intent of the investment is to enable data intensive research to be conducted collaboratively by all Australian research and their partners – both within research, and with partners beyond research.	By using common approaches it should be much easier for European and Australian researchers to collaborate.

A summary note from the Facilitating Entity

The Australian 2016 National Research Infrastructure Roadmap says:

Australia has the opportunity to consolidate the gains of the past decade and create a more integrated, coherent and reliable system to deal with the various needs of data-intensive, cross-disciplinary and global collaborative research. An Australian Research Data Cloud would build on existing eResearch infrastructure to create a cohesive, seamless experience for researchers that provides a fully integrated system.

The Australian Research Data Cloud should broadly align with the European Open Science Cloud and other global initiatives. It should support research data management from creation and discovery, through description and provenance, integration and storage, manipulation and analysis, and preservation. This improves the quality, reliability, durability, and accessibility of data, ensuring the outputs of research are more transparent. It should provide digital platforms that meet specific research requirements and integrate other data rich research infrastructure. It should support the sharing of informatics and software techniques to enable the deployment and wide use by researchers.

The underpinning Australian eResearch infrastructure should include cloud computing, HPC, networks, access, authentication and trusted data repositories. Data, collaboration and software services, skills and knowledge provided by the Australian Research Data Cloud will be an essential part of the new system.

EOSC in practice Story #12 – EUDAT CDI Data Shared Across Borders and Boundaries

Title	EUDAT - Data Shared Across Borders and Boundaries
Facilitating Entity	EUDAT Collaborative Data Infrastructure

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	EUDAT is a Europe-wide service network of major research infrastructures, data repositories and high-performance computing and data centres. The EUDAT Collaborative Data Infrastructure (CDI) has been set up through ongoing collaboration between service providers and research communities working as part of a common framework for developing and operating an **interoperable layer of data services**, the EUDAT service suite. The scientific fields cover diverse domains including social sciences and humanities, Earth and atmospheric science, climate science, biodiversity, life sciences, and physics.	The EUDAT CDI services are developed according to a **service-centric strategy** based on service co-design, involving users in all phases of the service definition, implementation, configuration and operation lifecycle. Current services have been developed in close collaboration with over 50 research communities spanning across many different scientific disciplines and involved at all stage of the design process making the EUDAT services cross-disciplinary by definition. Among others, these communities include CLARIN, Elixir, ENES, EPOS, Europeana, ICOS, LifeWatch, West-Life, etc. Cross-disciplinarity is taken into account also in the composition of the CDI: 6 thematic service providers have joined the EUDAT CDI as full members
2. Recognition	In 2011, the Collaborative Data Infrastructure emerged as a new concept and was presented as “the vision for the future”. 6 years later, it has become a concrete entity bringing together service providers and research communities and providing a range of data services and solutions for sharing, preserving, accessing and performing computations with research data.	Being a member of the EUDAT CDI means to be recognized as a European leader in data management and data infrastructure services. Driving the strategic directions of the EOSC-hub project, the EUDAT CDI positions itself as a key pillar in the ecosystem of the European Open Science Cloud

Dimension	Description	Outstanding features
3. Multiple Member State collaboration	With a network of more than **20 European research organisations, data and computing centres in 14 countries**, the EUDAT Collaborative Data Infrastructure (CDI) is one of the largest infrastructures of integrated data services and resources supporting research in Europe.	EUDAT is not a member state based organisation. Its members are research organisations (involving one intergovernmental organisation) hosted in 14 countries. Research organisations can join as individual members regardless of their geographical provenance (Multiple organisations from the same country may join the CDI) already promoting one of EOSC's principal governance recommendations of the Declaration in a direct dialogue with Member States
4. Trusted services offer	EUDAT is supports researchers with their day-to-day needs of **storage and sharing**, and helping them meet their long-term obligations under research data management plans. Covering both access and deposit, from informal data sharing to long-term archiving, and addressing identification, discoverability and analysis of both long-tail and big data, EUDAT services address the full lifecycle of research data and offers a set of integrated solutions. EUDAT delivers comprehensive research data management solutions for European researchers and research infrastructures through its partner network, following a model of tailoring solutions in close collaboration with end-users.	The EUDAT CDI suite of data management services supports researchers and research communities to ensure their data is FAIR compliant and promote open access to data. EUDAT promotes the release of data under standard licences, with the long- term aspiration that all data in the CDI should become fully open access. Access to data in the CDI is free at the point of use. **User Training** is also a core component of the EUDAT offer. EUDAT services are at the core of the EOSC-hub marketplace.The EUDAT suite of services include the following building blocks: **B2SAFE**: a robust, safe and highly available service which allows community and departmental repositories to **implement data management policies** on their research data across multiple administrative domains in a trustworthy manner. **B2DROP** is an easy-to-use, user-friendly and trustworthy **storage environment**. **B2SHARE** is a user-friendly, reliable and trustworthy way for researchers, scientific communities and citizen scientists to **store and share small-scale research data** from diverse contexts. **B2FIND** is a **discovery service based on metadata steadily harvested from research data collections** from EUDAT data centres and other repositories. **B2ACCESS** is the easy-to-use and secure **Authentication and Authorization platform developed by EUDAT**.

Dimension	Description	Outstanding features
5. Certification	Trust and quality of service are both extremely important and are usually broader in scope than that of a data repository. Rather than pursue trustworthiness under an "ISO 16363" track (for trustworthy digital repositories), generic service providers in the CDI have sought, and are actively seeking, accreditation and external audit in more general standards of service quality (under ISO 9000) and information security (under ISO 27000). In addition, EUDAT also initiated a programme for all CDI sites to participate in data quality assessments following the guidelines of the Data Seal of Approval (DSA), which recently became the the CoreTrustSeal after a merger with the ICSU World Data System.	Several EUDAT service providers are certified against internationally recognised standards. These organisations, being more "generic" than a dedicated digital repository, are increasingly pursuing broader measures of trustworthiness through audited standards like ISO 9001 (service quality management) and ISO 27001 (information security management). Other EUDAT sites have been or will be certified under the Core Trust Seal.
6. GDPR compliance	EUDAT encourages the open publication and sharing of research data under permissive licence conditions (we recommend CC BY 4.0) in line with the FAIR data principles of findability, accessibility, interoperability and reusability. CDI services are being adapted to align with the new and strengthened rights of data subjects. From this baseline of openness, EUDAT recognises that certain data cannot be fully open. The personal data of European citizens, as defined and codified under the General Data Protection Regulation (GDPR) of 2016, form a major class of restricted data.	The EUDAT CDI Collaboration supports, through common guidelines and suggested best practices, its members and service providers in devising and implementing appropriate technical solutions to make their infrastructure and services compliant with the GDPR. EUDAT recently published two well-received and widely circulated reports on managing sensitive research data under the rules of the incoming General Data Protection Regulation and organized workshops on this important topic at various events.

Dimension	Description	Outstanding features
7. Sustainability model	Following two grants from the EC to develop the foundation of the EUDAT Collaborative Data Infrastructure, the collaboration evolved into a partnership network constituted through the June 2016 CDI Agreement. Through this Agreement, EUDAT partners have committed to sustain the collaboration and its services for an initial period of ten years and have established a Secretariat responsible for managing, on a daily basis, the CDI Agreement... ... and for coordinating the development and operation of the CDI infrastructure. In February 2018, a limited liability company (EUDAT Ltd) was formally established to operate on a non-profit making basis as the voice of European organisations working together as part of the EUDAT Collaborative Data Infrastructure in order to provide services related to scientific and research data storage and lifecycle management. To cover the cost of the management, marketing and core services EUDAT ltd receives revenue in the form of fees from the CDI partners.	The EUDAT CDI is a growing organisation based on a contractual agreement between its members, signed for an initial period of 10 years. It is one of the key pillars of the European Open Science Cloud. The CDI is an open enterprises and offers the possibility for generic and thematic service providers to join the network with various levels of engagement and integration. As of April 2018, 23 partners have formally acceded the CDI. The current sustainability model relies on a partnership Agreement in which the CDI members sustain the collaboration and its services with a yearly fee.
8. User experience	EUDAT has, since its inception, been working on the principle that the research communities should be in the driving seat for selecting the main services, and that they should also directly participate in the design and development of these services, as part of multi-disciplinary task forces. Coming from different scientific domains, the research communities involved in EUDAT bring with them **specific requirements and knowledge regarding the development of the CDI** and its services, and a concrete wish to pilot and take up these services within their own research infrastructure.	Clear communication links between users and providers have been defined from the start of EUDAT. EUDAT facilitates the uptake of services by these communities through specific uptake plans and the active involvement of community experts in the development and deployment of solutions within their community. EUDAT services can also be used by individual researchers and citizen scientists.

Dimension	Description	Outstanding features
	Their requirements range from **registration and replication of data objects across multiple EUDAT nodes** for greater availability, safety and preservation of the data, to the ability to move large volumes of data to computational facilities, through better discovery and sharing of the data beyond the initial community.	

Summary note from Facilitating Entity

The establishment of the EUDAT CDI is timely with the imminent realization of the European Open Science Cloud which aims to offer open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines.

EUDAT customers fall into three segments: EU research infrastructure funding programmes; European researchers (including ESFRI data managers, and both research data producers and consumers); and the partner organisations in the CDI network. We operate in the "e-infrastructure" market for research services, alongside natural partners EGI.eu, PRACE, OpenAIRE and GÉANT.

EUDAT has five core value propositions:

1. management & marketing services for the CDI partnership network;
2. federated data management services for medium-to-large research consortia;
3. research data storage and sharing for small-to-medium research consortia, individual researchers and citizen scientists;
4. research data discovery services for all researchers and the research-interested citizen;
5. data management training for all researchers.

EOSC in practice Story #13 - The Astronomical Virtual Observatory on a FAIR path toward Open Science

Title	The Astronomical Virtual Observatory on a FAIR path toward Open Science
Facilitating Entity	The astronomical Virtual Observatory (VO)

Dimension	Description	Outstanding features
1. Research fields covered (cross-disciplinarity)	Astronomy and astroparticlepysics. Although astrophisics might look like one discipline, with the multi-messenger approach the field has a wide variety of techniques and data structures, from electromagnetic spectrum, (from radio, infrared, optical, x-ray, to gamma ray wavelengths), to neutrinos and cosmic ray particles and gravitational waves.	The astronomical Virtual Observatory (VO) provides the underlying framework of standards, which has developed since its beginnings in 2001, into a now mature and operational e-infrastructure that is widely implemented in Europe and internationally. VO services provide astronomy researchers with seamless access to hundreds of global astronomy services, with innovative 'all-sky' tools for accessing a wide diversity of data including the largest billion source sky surveys.
2. Recognition	The astroparticle physics scientist will get recognition through peer reviewed publications. The data scientist work can be credited by referencing to software or data products. In the VO, credits to tools and data can be linked almost automatically, leaving a trail of credits linked in the final scientific publication	· Astronomy is at the forefront of data sharing with a well-established record of using common data formats, and of interoperability between data archives, services and tools.
3. Multiple Member State collaboration	EURO-VO ASTERICS IVOA-	Institutes in several EU member states 24 organisations in 6 EU member states truly worldwide
4. Trusted services offer	With this success story, we may see the Virtual Observatory as an excellent precursor and example for the EOSC.	With the experience and lessons learned from building the astronomical interoperability framework, the VO has much to offer to the development of the EOSC.

Dimension	Description	Outstanding features
		The EOSC in turn provides the opportunity for VO to realise its next major challenges of strengthening data stewardship activities for open science, and interfacing the VO with interoperable computing to address the big data challenge presented by the unprecedented data rates that will be generated by the astronomy ESFRI and space missions.
5. Certification	IVOA standards Many standards are defined and implemented. Meetings on setting standards every half year	The VO is FAIR by design. The architecture of the International Virtual Observatory Alliance (IVOA) enables data to be Findable via VO registries, Accessible via standard data access protocols, Interoperable with core VO standards, and Re-usable by multiple tools and services. The remarkable direct mapping of the VO concepts to the FAIR principles means that the VO is ready for the EOSC, a view that has been reinforced by participation in EOSC Pilot activities.
6. GDPR compliance	Every institute is bound by the rules of their country, so they will have to follow GDPR as well.	The VO will implement standards evolving from the GDPR.
7. Sustainability model	All archives are maintained by the individual observatories and funding will be obtained through the observatory regular contacts, normally linked to local funding per country.	· The VO provides central services as a registry to data products in the archives, a repository of tools and the language to make the tools interact with each other and standards for interoperability. These services are supported through EC projects and local funding at the various institutes and countries involved.
8. User experience	The VO users group is growing. There are VO schools for unexperienced users, on-premises training and a forum for experienced users and one for data providers. With the Gravitational Wave success of last year, a recently developed tool to look for archive data at the location in the sky where the Gravitational wave is expected to originate from, was very useful.	Succeeded in helping these large infrastructures to become not just users of the VO, but participants in its development — ensuring the relevance and applicability of the framework, and illustrating a good model for developing an e-infrastructure that meets the needs of the scientific domain it serves.

A summary note from the Facilitating Entity

The Astronomical Virtual Observatory on a FAIR path toward Open Science

Astronomy is at the forefront of data sharing with a well-established record of using common data formats, and of interoperability between data archives, services and tools. The astronomical Virtual Observatory (VO) provides the underlying framework of standards, which has developed since its beginnings in 2001, into a now mature and operational e-infrastructure that is widely implemented in Europe and internationally. VO services provide astronomy researchers with seamless access to hundreds of global astronomy services, with innovative 'all-sky' tools for accessing a wide diversity of data including the largest largest billion source sky surveys.

The Data Access, Discovery and Interoperability, 'DADI', work of the H2020 ASTERICS project builds on these achievements by engaging the astronomy and astroparticle ESFRIs, and their 'pathfinders', in a coordinated effort to implement the data from these multi-wavelength (electromagnetic spectrum) and multi-messenger (neutrinos, cosmic rays, gravitational waves) facilities in the VO framework. Importantly 'DADI' has succeeded in helping these large infrastructures to become not just users of the VO, but participants in its development — ensuring the relevance and applicability of the framework, and illustrating a good model for developing an e-infrastructure that meets the needs of the scientific domain it serves.

The VO is FAIR by design. The architecture of the International Virtual Observatory Alliance (IVOA) enables data to be Findable via VO registries, Accessible via standard data access protocols, Interoperable with core VO standards, and Re-usable by multiple tools and services. The remarkable direct mapping of the VO concepts to the FAIR principles means that the VO is ready for the EOSC, a view that has been reinforced by participation in EOSC Pilot activities.

With this success story, we may see the Virtual Observatory as an excellent precursor and example for the EOSC. With the experience and lessons learned from building the astronomical interoperability framework, the VO has much to offer to the development of the EOSC. The EOSC in turn provides the opportunity for VO to realise its next major challenges of strengthening data stewardship activities for open science, and interfacing the VO with interoperable computing to address the big data challenge presented by the unprecedented data rates that will be generated by the astronomy ESFRI and space missions.

ANNEX II – REQUIREMENTS FOR CONFORMANT PROVIDERS OF CLOUD COINS – A WORKING EXAMPLE

Definitions:

1. Digital Object: An electronic artefact, including, but not limited to data, software, metadata, and/or workflows that can be stored or manipulated in an electronic information system.
2. Digital Object Steward: The individual or organization that created and/or controls a digital object and that has formal responsibility for its security, integrity and/or availability.
3. Investigator: A user who interacts with the Commons.
4. Cloud Credits Coordinating Center: The organization (including subcontractors, where relevant) that distributes computing resources (Credits) to stewards and prospective users of digital objects for use with providers.
5. Provider: An organization that makes a conformant cloud infrastructure available to users of the Commons and accepts NIH Commons Credits
6. Reseller: an entity which provides capabilities as a result of reselling or providing access to another provider's capabilities.
7. FISMA: Federal Information Security Management Act (44 USC § 3541 et seq) enacted as Title III of the E-Government Act of 2002, defines federal agency responsibilities for Information Assurance.
8. NIST: National Institute of Standards and Technology
9. IaaS: Infrastructure as a Service, based on NIST definitions[35]
10. PaaS: Platform as a Service, based on NIST definitions
11. SaaS: Software as a Service, based on NIST definitions
12. REST: Representational State Transfer; an implementation independent protocol for exchanging information over networks.
13. SLA: Service level agreement
14. CPU: Central Processing Units
15. VM: Virtual Machines
16. FTP: File Transfer Protocol
17. SFTP: Secure (SSH) Files Transfer Protocol.

General Requirements:

1. Providers must offer one or more of the following cloud services: IaaS, PaaS or SaaS. When included in a provider's offering to reduce the effort needed for developing or running computational or visualization tools, PaaS or SaaS-only offerors must also include an available data access API (or equivalent), which can be used by recipients of credits and the general public, to facilitate access to, and egress of, key public-use data sets. The only exceptions to this requirement are interfaces specified in these conformance standards.
2. Resellers: A reseller of services can act as a conformant provider so long as the provider upon which they operate their service layer is able to meet the conformance requirements, including and the reseller is able to meet the service-layer portion of those requirements (e.g., interface).

35The NIST Definition of Cloud Computing, Peter Mell and Timothy Grance. NIST Special Publication 800-145

3. Business Relationships with the Cloud Credits Coordinating Center: The provider[36] must set up an appropriate business relationship with the NIH designated cloud credits coordinating center that will be distributing NIH resources/credits to stewards of digital objects. The exact nature of the relationship is to be determined by the provider and the cloud credits coordinating center.
4. Credit Distribution Model: The provider must accept the financial mechanism by which the Government intends to deliver payment and to provide monthly on pre-defined and mutually agreeable reporting of Commons user metrics for those utilizing their services.
5. Validation of Conformance: providers will need to document compliance with conformance standards prior to receiving approval to receive Commons credits from the cloud credits coordinating center. Provider self-certifications will be bound by the False Claims Act (31 USC §§ 3729-3733) and the general provision about making false statements to the government encapsulated in 18 USC § 1001 and is punishable by fine and/or imprisonment. During the current pilot period (October 1, 2015 to September 30, 2018), providers will not need to re-certify. Providers will need to recertify upon the transition to regular operations and thereafter recertify on a bi-annual basis.
6. Changes to NIH requirements: The NIH will post proposed changes to the requirements for conformant clouds by an appropriate mechanism (to be determined) for comment by interested parties. Approved changes will become effective six (6) months from the date of approval. All approved providers will retain their ability to participate in the Commons during the three (3) years of proposed Pilot activities. Conformant providers, once approved, will be grandfathered against changes to the conformance requirements for the remainder of the pilot. However in order to retain their accreditation, they will need to meet the current requirements that are effective at the time of any post-pilot activities.
7. General access considerations: In order to be part of the Commons, providers must make their services available to the broad research community, comprising of NIH grant holders who have applied for, or received, Commons credits. Thus, a cloud that is inaccessible outside of the provider organization will not be considered conformant, since it does not make the digital objects contained within that cloud available to the broad research community.
8. Business relationships and liability: Digital Object Stewards and other investigators that interact with the Commons will do so under a business relationship with the provider(s); the government will not be a party to these agreements. Similarly, the government and providers will not participate in a direct relationship for the purposes of the distribution of resources; rather resources will be distributed and managed by a third party (the cloud credits coordinating center) with whom the government will have a contractual relationship. The government therefore accepts no liability for the actions of investigators in the Commons. Providers are encouraged to seek appropriate counsel and take appropriate steps to understand their potential liability associated with the actions of investigators in the Commons. The government will not define the terms and conditions which providers choose to offer their license agreements, or equivalent, to investigators. Providers are free to define terms and conditions, so long as they are (a) consistently applied, (b) are part of the regular account provisioning process, (c) do not violate federal or other applicable law and regulation, (d) do not assert intellectual property or other title on data or other digital objects, and (e) do not otherwise violate the conformance requirements.
9. Providers must follow a defined protocol when attempting to raise rates for users of the Commons in this pilot. The maximum price that can be charged for a given service by a given provider) is the published price on the Commons Portal for that provider for that service. Notification of price increases for any service listed on the Commons Portal must be made through the Coordination Center, who will update the prices on the Commons Portal upon request. This may be done up to once per month for any provider. The lead time for processing by the Pilot Coordination Center will be no more than 1 week.

Interfaces:

1. Interfaces: All interface standards and specifications germane to Commons Credit Pilot operations should be published and available to the research community. Although the interface must be open source, there is no requirement for the software itself to be open source.

36In the interest of brevity, it is intended that a reference hereafter to "provider "or "reseller" be read as "provider or reseller, as applicable"

2. Data deposit interface: Providers make a series of data deposit interfaces available to investigators. These interfaces include an interactive, web-based interface or other services that the provider feels will add value to the research community. The interfaces should provide an identifier (such as a URI) that can be used to access the digital object.
3. Data download: All providers are required to support a set of simple defined data download interfaces for use by pilot participants (e.g., FTP, SFTP, REST or Web Service interfaces).
4. Management: Providers must provide a minimally-functional web-based management console available that:
   a. Enables investigators to actively manage access controls for their data, and to manage Credits provided to them by the Coordination Center, as approved by the NIH,
   b. Provides historical and current service metrics to investigators, indicative of their storage, compute and network usage.
   c. Generates configurable alerts based on threshold usage and cost
   d. Monitors the Provider's SLA
5. Computational Tools: Providers must make available a relatively simple mechanism to launch software applications against any accessible data. Prospective cloud providers should demonstrate how to launch an application in their environment against such data (e.g., launching a sequence alignment algorithm, such as BLAST, against a set of stored sequence data).

Identifiers and Metadata:

1. Identifiers: Providers must provide a mechanism to apply a resolvable identifier at the granularity requested by the investigator. This identifier may be a URI, URL or other identifier, so long as it can be resolved into an address that can be used for appropriately authorized access to the digital object for use, retrieval and indexing.

Compute and Storage:

1. Storage: Conformant providers must provide:
   a. A minimum of a total of 5 Petabytes of persistent storage pool available at published commercial rates for service zones to be used in the pilot[37].
   b. Elastic capability to provision storage based on computational need triggers.
   c. Notification service with selectable alerts for storage usage.
2. Compute: A conformant provider must have:
   a. Minimum pool size of 100,000 CPU cores (or equivalent measure, per provider's specific nomenclature) for allocation to VMs. These should be at published commercial rates[38]. The CPU core pool need not be reserved for pilot use, but represents the composite capability of the provider pooling all regions of their service zones to be used in the pilot. The provider conformance application package should provide a clear definition of the computing unit measure used.
   b. Offer a selection of VMs for various computational needs.
   c. Notification service for compute with selectable alerts.

Networking and Connectivity:

1. Commodity High-speed Internet: The provider must provide connectivity to the commodity internet that can be configured to support networking with a minimum bandwidth of 40 Gb/sec[39] for service zones used in the Pilot.

37A published commercial rate in this case is a service that is advertised on the Commons Portal and available to users of Commons Credits at those defined prices.

38 Ibid

39The government recognizes that networking in a public cloud environment is dependent on the configuration selected by the user of those services. Thus, the requirement is not to provide 40 GB/sec in all cases but that the environment can be configured

2. Network pool: The Provider must provide capabilities for configuring devices such as virtual firewalls, network switches, and load balancers for ease in implementing a scalable and redundant connectivity, or link aggregation.

Information Assurance:

1. Information Assurance is a shared responsibility of all system stakeholders to manage information-related risks. Providers must demonstrate the capability of managing their part of the responsibility stack in protecting and defending information by ensuring its confidentiality, integrity, authentication, availability, and nonrepudiation.
2. FISMA Compliance: Conformant providers must meet FISMA requirements. The minimum standard is that the offering meets FedRAMP Low baseline controls. Providers are also free to provide additional environments that meet FedRAMP Moderate and/or High baseline controls (at the same or different price points) as desired, so long as the provider has an available Low environment and the relative costs at the various levels of assurance are posted. Providers should consult relevant NIST publications for current standards. Presence on the published GSA list[40] may be used as evidence of compliance in the application process.
3. Authorization and Accreditation (A&A): Providers must carry out the standard authorization and accreditation activities on their standard IaaS, PaaS and SaaS offerings, as well as such additional PaaS capabilities are provided to support their approval as part of the Commons.
4. Providers must provide a mechanism to assure the security of data in their IaaS, PaaS and SaaS environments (e.g., data encryption at rest and in transit).

Assessment and Authorization:

1. General requirements: Providers must follow relevant NIST guidelines with regard to authentication and authorization, specifically NIST Special Publications 800-63 Rev 2, Electronic Authentication Guideline (http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63-2.pdf).
2. Authentication: The provider must allow and support authentication in a form that complies with good and standard industry practices.
3. Authorization: Providers must provide a straightforward mechanism to enable investigators to set permissions on data and other digital objects that have been stored in the Commons (see 'Interfaces' above). The minimum requirement is for a digital object to be made available either (a) only to the depositor, (b) to such individuals as defined by the investigator (who must know the relevant account names) or (c) publicly available to all users.

Desirable Features (preferred, but not required):

1. Continuity: A published Continuity of Operations and Disaster Recovery Capability (COOP/DR) plan, or a published SLA covering COOP/DR features is desired.
2. Interfaces: Published and licensed interface specifications under an approved, non-viral, Open Source license is desired. A list of approved licenses will be made available by the NIH Office of the Associate Director for Data Science, but the Apache license version 2.0 (http://www.apache.org/licenses/LICENSE-2.0) is the preferred license.
3. Interfaces: Adherence to Open Cloud Computing Interfaces (OCCI) and/or Cloud Infrastructure Management Interface (CIMI) specifications for APIs is desired.
4. Encryption service: An encryption mechanism using a FIPS 140-2 approved encryption algorithm is desired for storing sensitive data, which may include capabilities for data in motion, data at rest, and data under computation.
5. Storage: Perpetual or other long-term data maintenance available on commercial terms is desired.

---

by a user to support 40 GB/sec if needed. The government also recognizes that such capabilities would likely incur additional costs to the Provider, costs which can be passed on to investigators that require such service levels

40http://www.gsa.gov/portal/category/105279

6. Data Download: Reduced charges with separate metering for data egress in the case of large or frequently accessed data is desired.
7. Virtualization: Additional Virtualization methods where the kernel of an operating system allows for multiple isolated user-space instances is desired.
8. Messaging: Options to use an available messaging system for facilitating information exchange between systems is desired.
9. Processing: Options for supporting batch and real time processing of tasks are desired.
10. Internet2: Connectivity to internet2 (http://www.internet2.edu/) at a bandwidth compliant with the provider's published SLA is desired.
11. Authentication: It is desired that the provider allows and supports authentication based on in Common credentials (http://www.incommonfederation.org/).
12. User Support: Helpdesk and technical support for assisting users are desired.

Getting in touch with the EU

IN PERSON

All over the European Union there are hundreds of Europe Direct Information Centres.
You can find the address of the centre nearest you at: http://europa.eu/contact

ON THE PHONE OR BY E-MAIL

Europe Direct is a service that answers your questions about the European Union.
You can contact this service
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696 or
– by electronic mail via: http://europa.eu/contact

Finding information about the EU

ONLINE

Information about the European Union in all the official languages of the EU is available on the Europa website at:
http://europa.eu

EU PUBLICATIONS

You can download or order free and priced EU publications from EU Bookshop at:
http://bookshop.europa.eu. Multiple copies of free publications may be obtained
by contacting Europe Direct or your local information centre (see http://europa.eu/contact)

EU LAW AND RELATED DOCUMENTS

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

OPEN DATA FROM THE EU

The EU Open Data Portal (http://data.europa.eu/euodp/en/data) provides access to
datasets from the EU. Data can be downloaded and reused for free, both for commercial and
non-commercial purposes.

Legal Notice

Luxembourg: Publications Office of the European Union, 2018

To help drive forward and implement the EOSC, the main thread of the report is to understand how the EOSC can effectively interlink **People, Data, Services and Training, Publications, Projects and Organisations**. The ideas presented here bring together, reflect on and further expand on various policy papers and recommendations contributing to the establishment of the EOSC that have been published by ongoing Horizon 2020 projects and national initiatives, as well by the Commission FAIR Data expert group and by the Open Science Policy Platform, with whom the group have collaborated actively. The report shows how Europe, with its strong scientific base and investments made in infrastructures has the skills, knowledge and capacity to turn EOSC into a reality in less than a year from now.

Studies and reports

Publications Office