ERIC Forum Implementation Project

Technical and Innovation Report: Report on integration of ERICs into pan-European infrastructures. The example of EOSC.

Work Package 6 – Deliverable 6.11

Deliverable no	D.6.11
Deliverable Title	2st Technical and Innovation Report: Report on integration of ERICs into pan-European infrastructures. The example of EOSC
Contractual delivery month	24
Responsible Partner	LifeWatch ERIC
Author(s)	Juan Miguel González-Aranda, Cristina Huertas Olivares, Christos Arvanitidis

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report has been produced in the context of the **ERIC FORUM H2020 Project**. It forms part of the **Work Package 6: "The Role of ERICs in European science policy and research strategy"**. It presents a review on the integration of ERIC development into pan-European infrastructures, particularly in the new EOSC.

The first chapter, which describes Research Infrastructures (RI) and RI clusters, introduces the reader on the rationale for their cooperation and collaboration in using the resources of the pan-European infrastructure EOSC. Through this process, a degree of integration of their basic assets (e.g. data and information) is achieved.

In the following chapter, the efforts that have been made to establish EOSC in the past are reviewed. Key players are described, along with the current prospects. After that, conclusions are listed. At the end, some recommendations for the enhancement of this integration process are provided.

1. INTRODUCTION

The European Research infrastructures (RI) are facilities that provide resources and services for the research communities to conduct research and foster innovation in their fields. Europe has at its disposal some of the best **science facilities** in the world. These **RIs** include as assets: major equipment or sets of instruments, knowledge-related facilities such as collections, archives or scientific data infrastructures, computing systems and communication networks. RIs have also the potential to foster excellence in science. One of their fundamental features is their compliance to the FAIR[1] principles.

The urgent **societal challenges**[2] have made the European Commission to spend additional effort to reinforce the capacity of European research to develop evidence-based knowledge and nature-based solutions that when co-designed with all the interested parties can effectively **lead** to economic, social and environmental **transitions**. It defines, evaluates and implements strategies and tools to provide Europe with world-class sustainable **RIs**. The strategic focus of this effort has been on the creation of a specific legal form called **ERIC** (European Research Infrastructure Consortium) **to facilitate the establishment and operation of RIs** which have been identified as key in the science and research strategy from Europe.

Several ERICs are e-Infrastructures that facilitate research by offering access to open access data and reproducible analytics, with practically unlimited power. This way, they help research communities to mobilize their members and achieve their research objectives, produce world-class knowledge and assist the society to address its major challenges. However, very often, they work isolated in a few disciplines and within a the same research domain. Decisions to tackle the urgent societal challenges require evidence based on interdisciplinary and cross-domain science, which means collaboration between the RIs and ERICs. Scientists need to shift their attitude from working in isolation as focused experts with only self-generated data towards a new paradigm based on their discoverability, collaborative analyses and comparison of the results in terms of patterns and processes as derived by multiple disciplines and domains. This requires a sustainable set of core of e-infra services. This above have brought the collaboration of the ERICs, first in the form of Science Clusters (e.g. ENVRI-FAIR, EOSC Life, etc.).

As an **example,** Environmental RIs and ERICs are collaborating within their Science Cluster ENVRI-FAIR. Each environmental ERIC shows a high level of specialization on their main

---

[1]Find the data and metadata in our Catalogues of Services; Freely access, use and share large datasets of different types and sources; Work with interoperable data, thanks to our standards, thesauri and ontologies; Reuse and combine data for different research questions, generating new services and meet community standards; Cluster data repositories are minting identifiers for the community. They provide open data and continuous, trusted working environments and networking opportunities to the research community. Long-term open data archives with high performance storage and computing services, to enable sustainable use of data. Open data and a virtual research space for open science where scientists can create content and collaborate. They collaborate: Major challenges Scientific communities collaborating on common topics Þ grand challenges. Alignment of practices and tools. Common data standards, annotation, ontologies and catalogues. Vertical and horizontal composability of services. Facilitation of data mobilisation.

[2] Such societal challenges are the impacts from climate change, from non-indigenous invasive species, from pollution and eutrophication, fragmentation of habitats, coastal development, etc., which result in loss of the ecosystem services offered to mankind with implications to the income of the populations, shifts in professions, immigration, etc.

tasks, e-Infras they develop and user communities. All together, they include a wide range of disciplines, with their own methodologies and research cultures. They are in charge of complex phenomena needing highly specialized and varying observation and analysis platforms. Within this framework of collaboration, they organize joint activities to foster **interdisciplinarity, such as on data annotation, cataloguing and ontologies used.**

One-step further down from forming a Science Cluster of ERICs, is their **integration into pan-European Infrastructures**. This integration is a key in maximizing the impacts on the European Research Area (ERA). In **ERA**, the EU policy of Open Science is a priority, with the objective to change the way European science works. **Such a pan-European infrastructure is EOSC (**European Open Science Cloud), which was created to support the EU policy of Open Science.

EOSC is a research environment for accessing and processing data to support EU science policy. This initiative offers researchers a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines by **federating existing data infrastructures**. The Cloud is meant to provide Europe with a global edge in reaping the full benefits of data-driven science.

The aim of **EOSC** is to give the EU a global lead in **research data management** by providing European science, industry and public authorities with: (a) world-class data infrastructure to store and manage data, (b) high-speed connectivity to transport data and (c) unlimited high performance computer systems to process data. It provides universal access channel to **EOSC** services and resources. Through the portal, researchers and professionals can access open and seamless services, data and other resources from a wide range of national, regional and institutional public research infrastructures across Europe. The portal facilitates interoperability of datasets and tools from different providers, and enables researchers to perform their work more quickly, in a collaborative way and disseminate their research results more widely. It already covers a wide range of disciplines (e.g. medical and health, natural sciences, physics, earth sciences, arts, humanities, agriculture, and engineering).

Like EOSC, there are other Pan-European Infrastructures Initiatives such as the EuroHPC (High-Performance Computing -HPC-), European Data Spaces, and the Long Tail of Science. This report, however, only focuses on **EOSC.**

2. PAST EFFORTS TOWARDS INTEGRATION OF ERIC INFRASTRUCTURE INTO EOSC

The main milestones of EOSC until 2020 are:

Date	Development Stage
May 2015	The creation of a **European Open Science Cloud** (**EOSC**) is proposed to the Competitiveness Council.
April 2016	The EOSC was proposed as part of the European Commission Communication *"European Cloud Initiative - Building a competitive data and knowledge economy in Europe"* with the aim: ● to federate existing research data infrastructures in Europe ● to implement a web of FAIR data and related services for science, making research data interoperable and machine actionable following the FAIR guiding principles.
March 2018	Investment of 320M€ to start prototyping the EOSC through project calls in H2020 under the EC research and innovation funding programme. The EC published the **EOSC Implementation Roadmap** detailing the main action lines towards the first **EOSC** implementation phase until 2020. Then, a multi-layered, interim governance structure was established from November 2018 to steer and oversee the implementation of the **EOSC** from 2019-2020.
November 2018	The **EOSC portal** was <u>launched</u> with its progressive integration and consolidation of e-infrastructure projects. It started as a collective effort from OpenAIRE, EOSC-hub, eInfraCentral and EOSCpilot projects building on the experience and technology of major pan-European e-infrastructures, universities and research infrastructures.
2018 - 2020	The **EOSC portal** was <u>enhanced</u> by the addition of further high-level functionalities and a wide range of scientific data and data analysis services. **EOSC** has been co-created in a series of funded projects and initiatives from Member States and Associated Countries.
July 2020	The EOSC Association was set up to provide a single voice for advocacy and represent the broader EOSC stakeholder community. This association aims to become operational by early 2021 and rapidly expand its membership. To achieve the ambitious objectives of EOSC, a governance structure was put in place, bringing together public funders and the EOSC stakeholders (researchers, service providers, industry, public sector and other users) (see Key players)

EC (2019). European Open Science Cloud (EOSC) strategic implementation plan.	EC Communication (2016) "European Cloud Initiative - Building a competitive data and knowledge economy in Europe"

Figure 1. Main reference documents for the establishment of EOSC until 2020.
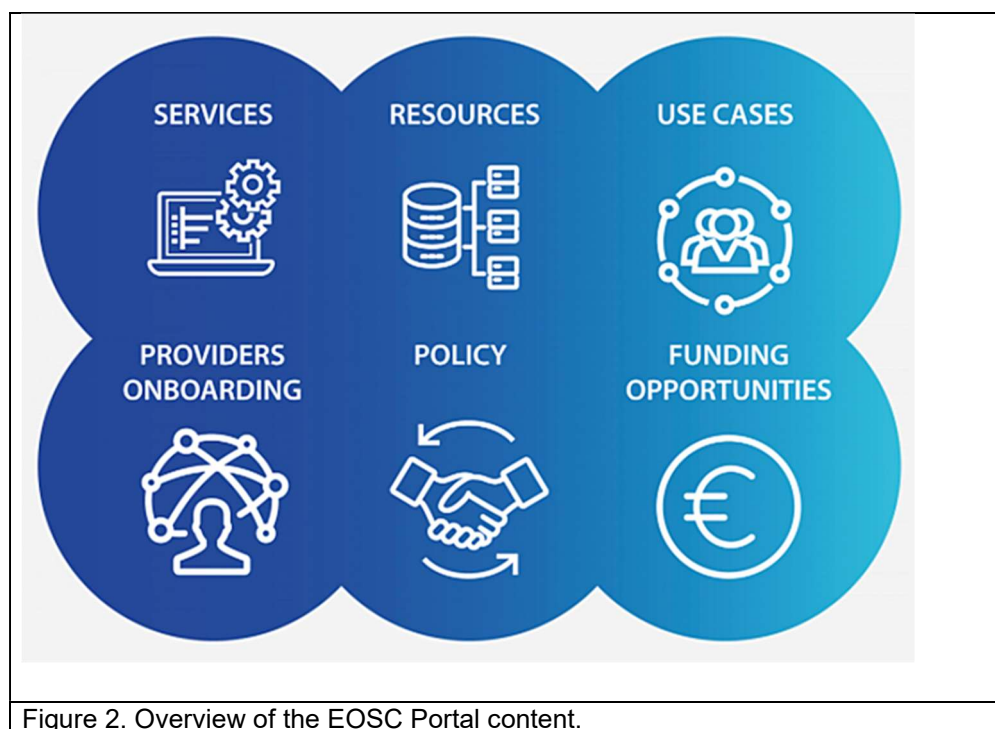

Figure 2. Overview of the EOSC Portal content.

ERICs have already contributed to the development of EOSC through their direct involvement in: (a) the various consultation rounds over these years (2015-2019), (b) the projects such as EOSC-hub and EOSC Enhance, (c) the formation and operation of the EOSC Association as legal entities, (d) the development and implementation of demonstration projects on the use of EOSC platform, such as ENVRI-FAIR, EOSC Life, etc and (e) the drafting and implementation of EOSC Future project.

The ERIC Forum considers that data quality and reproducibility are two critical factors for the success of EOSC. It also recognizes that research infrastructures are, to some extent, responsible for delivering data and services that meet the quality requirements set by the EOSC Rules of Participation. To support the efficient use of EOSC, the ERIC Forum suggests to develop a robust policy on data and service (software) quality assessment. It is anrecognised issue that several research domains face structural difficulties related to delivering FAIR data. Such difficulties do not only pertain to the later phases of data management, such as validation and publication. The biggest challenge in the open data culture is that they require completely new workflows and significant infrastructural support throughout and eventually even beyond the life-cycle of a single research project. Research communities therefore need support, new skills and appropriate tools and services related to each phase of the data management process. The broadening and opening scope of access to data through EOSC will make it even more important and urgent to address this issue. The ERIC Forum advocates for a concrete data policy framework and looks forward to working closely with EOSC on making such policy. To further support the development of EOSC, members of the ERIC Forum are involved in five different cluster projects of the EOSC panorama, i.e. ENVRI-FAIR (environmental research), EOSC-LIFE (life sciences), ESCAPE (astronomy and particle physics), PaNOSC (multidisciplinary scientific analysis), and SSHOC (social sciences and humanities). In addition, the ERIC Forum Implementation Project has a task (Task 6.2: ERICs technical and innovation multi-scale integration) dedicated to the integration of ERIC-related developments into pan-European e-infrastructures, and particularly into EOSC.

EOSC ecosystem, fragmentation of disciplinary and national silos

After this first period of scoping, design and implementation, it has been realized by all key-players, including the EU, that EOSC development was attempted primarily through multiple approaches which were attached to different disciplines with a little degree of cross-collaboration. Both the previously-mentioned Science Clusters projects and the EOSC Future project have helped tremendously the multidisciplinary and cross-domain approach. Within the EOSC Future, Science Clusters attempt to break current cross-disciplinary barriers by building collectively the development of Science Projects (SPs). These SPs have helped the Science Clusters to start identifying their collaborative interfaces, their "*trading zones*", in which they will then have to invest in common to produce new, synthetic knowledge, technology and innovation, along with the continuous support and collaboration of the e-Infrastructures. Trading zones are the most fruitful areas for such investments. In addition, these areas will provide the EOSC Infrastructure with new challenges to face if this is meant for as wide use as possible. If successful, this exercise will enforce Science Clusters, e-Infrastructures and EOSC Future to facilitate the big societal challenges to be addressed with a degree of success. They have also the potential to support ERICs to become part of the Consortia which will help the EU to implement its policies (e.g. New Biodiversity Policy, European Green Deal, Missions, etc.).

3. KEY PLAYERS

- **ESFRI Science Clusters**. They act as the key interfaces between the scientific communities, their infrastructures, and the EOSC: (a) provide vital links to the scientific community and EOSC governance; (b) build and maintain key community initiatives; (c) provide an integrative, rapid and efficient platform for connecting to RI cluster services, even those from across the clusters, for innovative research in support of the attempt to address the important societal challenges; (d) provide platforms for scientific interoperability in EOSC.

- **e-Infrastructures** (e.g. GÉANT, EGI, OpenAir, etc.). They act as thrusters of IT applications by providing massive Infrastructure in terms of data storage, computing power, applications (software) including web services. All of them agree a common reference of operation through the so-called e-Infrastructures Reflection Group https://e-irg.eu, a strategic body to facilitate integration in the area of European e-Infrastructures and connected services, within and between member states, at the European level and globally. In fact, e-IRG mission is to support both coherent, innovative and strategic European e-Infrastructure policymaking and the development of convergent and sustainable e-Infrastructure services. It is important to remark that e-IRG cooperates with ESFRI and ERIC processes, as well as with the long tail of science in the effect possible.

- **EOSC Association**: This is the legal entity set in 2020 to represent European stakeholders of EOSC in a collective manner by its four founding members: GÉANT, CESAER, CSIC and GARR. Prior to the formation of EOSC Association, there was an interim EOSC governance scheme, tasked with the preparatory work,* whose the job came to an end by 31 December 2020. The EOSC Association is the legal entity that has enter into a contractual arrangement with the EC under the European partnership (An MoU was signed between the EOSC Association and the EC to officially launch the EOSC partnership at the end of 2020). https://www.eosc.eu/ (* The previous governance structure brought together public funders (from Member States, Associated States and the European Commission) and the EOSC stakeholders (researchers, service providers, industry, public sector and other users). The 3-tier structure included: 1) a Governance Board, consisting of representatives from the Member States and the European Commission who jointly decide the future of EOSC; 2) an Executive Board, consisting of experts from key European research and e-infrastructures communities who advise on the EOSC's implementation and the way forward; 3) a Stakeholders Forum, giving voice to the wider community of users, service providers, industry and the public sector).

- **Copernicus**: Copernicus is the European Union's Earth observation Programme, looking at our planet and its environment to benefit all European citizens. It offers information services that collect data from satellite earth observations and from in-situ (non-space) ones. The European Commission manages the Programme. It is implemented in partnership with the Member States, the European Space Agency (ESA), the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), the European Centre for Medium-Range Weather Forecasts (ECMWF), EU Agencies and Mercator Océan. Vast amounts of global data from satellites and ground-based, airborne, and seaborne measurement systems provide information to help service providers, public authorities, and other international organisations improve European citizens' quality of life and beyond. The information services provided are free and openly accessible to users.

- **EuroHPC:** The European High Performance Computing Joint Undertaking (EuroHPC JU) is a joint initiative between the EU, European countries and private partners to develop a World Class Supercomputing Ecosystem in Europe (https://eurohpc-ju.europa.eu/).

4. CURRENT PROSPECTS. FROM A PROJECT-BASED TO A STAKEHOLDER-FOCUSED APPROACH

The EOSC under Horizon Europe 2021-2027: EU countries and countries associated with Horizon 2020 agreed to run the EOSC as a co-programmed **European Partnership under Horizon Europe** https://ec.europa.eu/info/horizon-europe/european-partnerships-horizon-europe/candidates-across-themes_en from 2021. This has been seen as the best instrument to provide a framework for collaboration and the pooling of resources at European, national, regional and institutional levels.

The proposal for a candidate EOSC partnership was published in June **2020** following a process of co-creating its vision and developing its strategic and operational objectives to be achieved by 2027. **It aims to deploy and consolidate by 2030 an open, trusted, virtual, federated environment in Europe to store, share and re-use research data across borders and scientific disciplines and provide access to rich array of related services. The Partnership will bring together institutional, national and European initiatives and engage all relevant stakeholders to jointly design and deploy EOSC, based on science commons where data are Findable, Accessible, Interoperable, Reusable (FAIR) and where research-enabling and other services are made available throughout Europe. This European contribution to a "Web of FAIR Data and Related Services for Science" will enhance the possibilities for researchers to find, share and reuse publications, data, and software and develop and access a service ecosystem, boosting the research capacity and innovation potential of Europe.**

The FAIR principles provide high-level ways to address at least some of the challenges in managing research data, such as:

- **Heterogeneity**: Data are defined by the scientific discipline that produces, manages or analyzes them. They are coming from observations performed by specific equipment, by surveys or any other data production /collection process (e.g. modelling). Data are also the results of processing steps developed after the raw data collection. As a consequence of this multifaceted landscape, a very wide variety of data frameworks have already been designed by research communities.

- **Granularity**: While raw data may be produced in large volumes, users may not need access to raw data but only to the results of post-processing steps, such as processed data, patterns or derived processes stemming out of their analyses. Deciding at what stage data should be preserved and made available to users has been also become something which is specific to the scientific domain.

- **Versioning**: Data are collected as time goes by. The sampling rate, the importance of preserving knowledge on the sequence of events also depends on the domain.

- **Disambiguation**: When it comes to representing data in digital form, the choice of identifiers is a problem in itself. It may occur that one identifier points to multiple objects introducing one more level of complexity for the user.
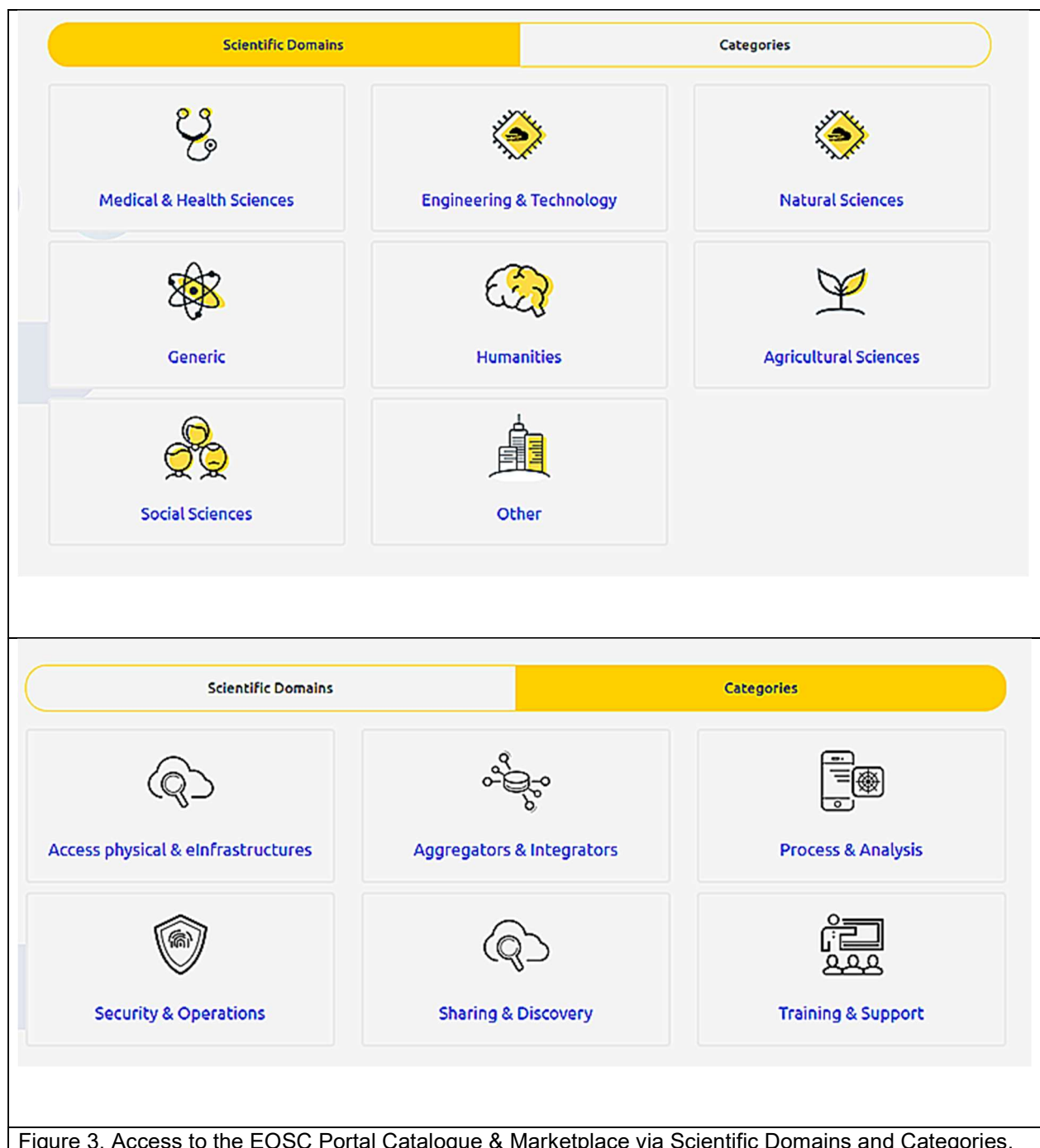
- **Diffusion prevention**

Figure 3. Access to the EOSC Portal Catalogue & Marketplace via Scientific Domains and Categories.

The EOSC Future Project, a Research and Innovation Action (RIA) under the Call H2020 - INFRAEOSC -2018-2020 (INFRAEOSC-03-2020) has recently being funded. The aim of this

project is to integrate, consolidate, and connect e-infrastructures, Science Clusters along with their research communities, and initiatives in EOSC to develop the EOSC-Core and EOSC-Exchange.

EOSC Future will expand the EOSC ecosystem, integrate existing but scattered initiatives, and hand over key project outputs to the EOSC Association. EOSC Future will unlock the potential of European research through a vision of Open Science for Society. The digital transformation of science is no longer a coming trend but a reality. Beyond better connectivity, faster processing, and more reliable storage, this has enabled a wealth of value-generating services and an ever-growing wave of data and research products. As these developments become a part of our daily lives, the digital transformation also brings new opportunities for European research through the digital platforms which provide it with opportunities for significantly more reproducible research and a passage to a more data-intensive and Open Science. The European Union's leadership in Open Science contributes to the advancement of the scientific discoveries and market innovation, as well as opening science to the public and addressing societal challenges more effectively.

EOSC Future will reduce the current research fragmentation of disciplinary and national silos, as well as promote FAIR data, professionally provided services, and open research products. EOSC Future will build on the work of EOSC-hub, EOSC Enhance, OpenAIRE-Advance, and other EOSC projects and initiatives. This includes projects awarded under the EINFRA-12-2017, INFRAEOSC-05-2018, and INFRAEOSC-04-2018 calls as well as the INFRAEOSC-07-2020 projects generating services for the Portal.

Therefore, EOSC Future will:

- strengthen, simplify, and increase access to different types of EOSC resources by effectively providing the mechanisms to implement a data and service interoperability framework (utilizing standard and concrete APIs and AAI) through an ecosystem of regional and thematic portals all connected to the EOSC Portal;

- support better cooperation and synergies with other European research initiatives such as High-Performance Computing (HPC), European Data Spaces, and the Long Tail of Science,

- closely monitor and support policy priorities of EU Member States and Associated Countries (MS/AC), including for Digitalisation, Artificial Intelligence (AI), Big Data, the European Green Deal, and Europe in the World.

5. CONCLUSIONS, RECOMMENDATIONS

Conclusions:
1. There is much potential when ERICs seek a degree of collaboration and integration of at least some of their assets. The outcome of this process surpasses the outcomes of the individual ERICs and reaches far wider audiences (added values);
2. The European Science Open Cloud (EOSC) provides a good example of this process;
3. Many useful lessons may be taken from the history and current operation of EOSC and at multiple levels, such as: strategic-political, structural-developmental, operational, communicational, etc.

Recommendations:
1. The example of EOSC shows that once a framework is in place, such as EOSC, the ERICs should, first, run the exercise of identifying their surfaces of interaction or trading zones at which their integration may take place;
2. When there is a plan, this integration has a lot more chances to succeed;
3. In any such process of ERIC integration, it should always be clear that the fundamental properties of producing evidence-based knowledge are guaranteed through the process of integration: quality, transparency and reproducibility. Open (FAIR) access data, reproducible analytics and mobilized communities are such property examples;
4. Robust policies should be built to ensure the continuous and effective operation of the integration process.
5. Any integration should be regarded as an ongoing process and not on a project-by-project one.
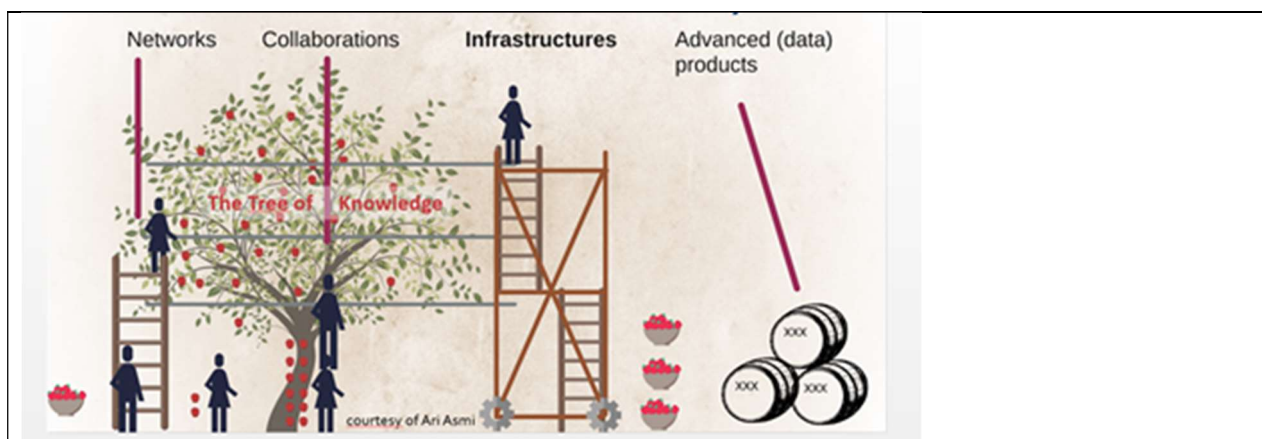

Figure 4. RIs and the Scientific Community.

6. REFERENCES

EOSC Portal, 2021. https://eosc-portal.eu/about/eosc-projects

European Commission, Our Digital Future 2021.
https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future_es

EOSC interoperability framework. Report from the EOSC Executive Board Working Groups FAIR and Architecture, 2020: https://op.europa.eu/en/publication-detail/-/publication/d787ea54-6a87-11ebaeb5-01aa75ed71a1/language-en/format-PDF/source-190578229

Landscape of EOSC-related infrastructures and initiatives. Report from the EOSC EB WG Landscape, 2020:
https://op.europa.eu/en/publication-detail/-/publication/cbb40bf3-f6fb-11ea991b01aa75ed71a1

Draft proposal for a European Partnership under Horizon Europe. EOSC, 2020: https://ec.europa.eu/info/files/europeanopen-science-cloud-eosc_en Progress on Open Science: Towards a Shared Research Knowledge System. Final Report of the Open Science Policy Platform (OSPP): https://ec.europa.eu/research/openscience/pdf/ec_rtd_osppfinal-report.pdf

EOSC strategic implementation plan, 2019: https://publications.europa.eu/s/mIli

Turning FAIR into reality. Final report and action plan from the European Commission (EC) Expert Group (EG) on FAIR data, 2018: https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source89516642

EC, Candidates for European Partnership across various themes, 2021.
https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/european-partnerships-horizon-europe/candidates-across-themes_en

EC (2019). European Open Science Cloud (EOSC) strategic implementation plan.

EC Communication (2016) "European Cloud Initiative - Building a competitive data and knowledge economy in Europe"