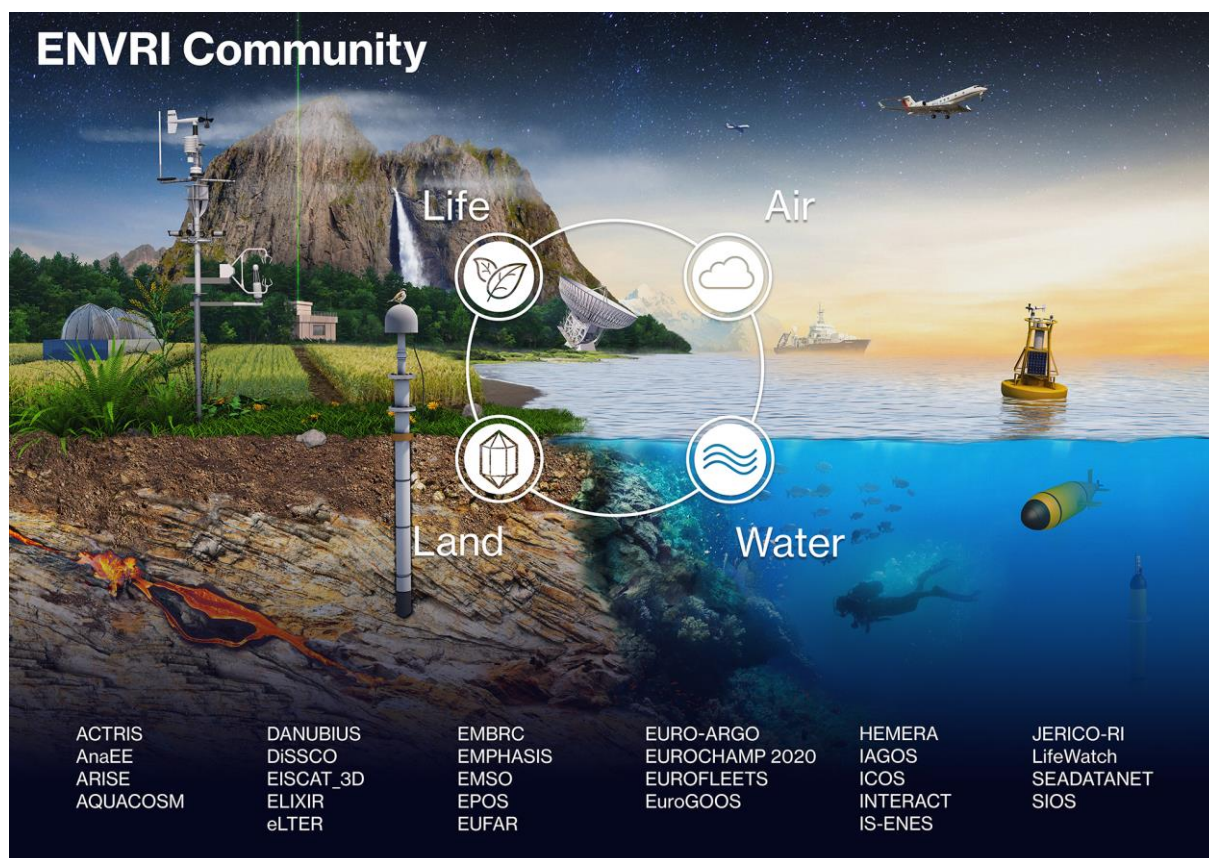# ENVRI-FAIR Position Paper on the EOSC

prepared by Andreas Petzold (a.petzold@fz-juelich.de) and Ari Asmi (ari.asmi@helsinki.fi)
endorsed by the ENVRI-FAIR Executive Board on 03 December 2019
revised on 13 February 2020 (updated illustrations added)

ENVRI-FAIR Cluster Project Mission Statement and Objectives

European Environmental Research Infrastructures on the ESFRI level provide data, research products and services from key areas of the Earth system. These research infrastructures (RI) form the ENVRI Cluster and include the principal producers and providers of environmental research data and research services in Europe from the four segments of the Earth system - Atmosphere, Marine, Solid Earth, and Biodiversity/Terrestrial Ecosystems. The data, products and services provided by the ENVRI Cluster are crucial European contributions to the integrated global observation system monitoring the state of the Earth system and climate. They are vital for assessing past and defining future policies, as well as for the development of environment-friendly innovations and adaptation as well as mitigation strategies. The ENVRI Cluster represents the core component of the European environmental research infrastructure landscape with the ENVRI community as their common forum for collaboration and co-creation.
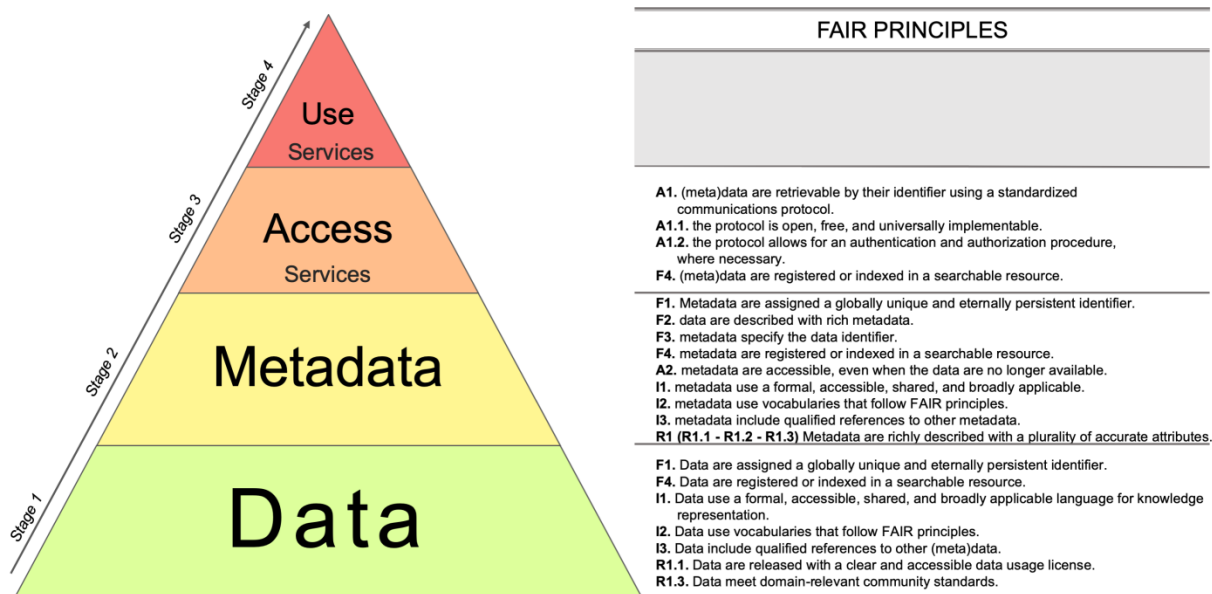
ENVRI-FAIR is the cluster project aiming towards the provision of services according to FAIR principles by the ENVRI community. It develops tools and resources for easy and seamless access to environmental data and services provided by ENVRI research infrastructures, with the high-impact ambition to prepare the foundations for the successful implementation of a federated machine-to-machine interface – the ENVRI-hub – to access environmental data and services provided by the contributing ENVRIs. The highest priority is on the provision of high-quality data using open licenses, standard mechanisms and protocols. The overall architecture is designed as strong integration at sub-domain level with a layer of integrated services on top which may serve as the nucleus for the ENVRI-hub. The hub forms the interface to the EOSC and will be realized as the services across ENVRIs and even between environmental subdomains become progressively more integrated [1].

ENVRI-FAIR Objectives are to

1. further develop common standards, protocols and policies for the data life cycle, including cataloguing, curation, provenance and service provision within the ENVRI Cluster, with specific consideration of the FAIR principles including interoperability, and of the tools and methods created during the preceding EU-projects ENVRI and ENVRIplus;

2. align these policies and standards with more extensive European policies (e.g. ISO 19115 INSPIRE) as well as with relevant international developments;

3. develop and implement the necessary tools for reaching Objective 1 in each research infrastructure, thereby adopting an open approach for sharing data and software;

4. improve the skills of research infrastructure personnel to develop and sustain knowledge on Research Data Management and FAIRness, including both cross-cutting and subdomain-specific knowledge, and on the FAIR infrastructures resulting from Objectives 1 and 2 through an extensive training;

5. increase the potential for innovation of each research infrastructure by establishing a specific ENVRI-FAIR section in the EOSC service catalogue, with the aim of stimulating common pre-commercial procurement processes and dissemination of outcomes and thus enhancing the uptake of research infrastructure services by private partners;

6. establish cohesion with the global research infrastructure landscape, including research infrastructure clusters and regional/international initiatives in the environmental sector; maintain ENVRI community knowledge with particular consideration of developing integrated activities;

7. expose thematic data services and tools from the research infrastructure catalogues to the EOSC catalogue of services, COPERNICUS, GEO, and other end-users.

The implementation strategy of ENVRI-FAIR follows the FAIRness maturity pyramid developed by EPOS [2]. The FAIR principles discussed by Wilkinson et al., 2016 [3] are analysed with respect to FAIRness of data, metadata and services, and are broken down into ascendant stages towards increasing maturity. These stages are reflected in the FAIRness implementation plans of the participating research infrastructures, taking into account the maturity of each research infrastructure. The maturity of the participating research infrastructures is assessed periodically in an approach inspired by the FAIRification process adopted by GO FAIR [4] and is applied in accordance with specific needs of the ENVRI cluster and those of the other ESFRI domains.

FAIR PRINCIPLES

A1. (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1. the protocol is open, free, and universally implementable.
A1.2. the protocol allows for an authentication and authorization procedure, where necessary.
F4. (meta)data are registered or indexed in a searchable resource.

F1. Metadata are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. metadata specify the data identifier.
F4. metadata are registered or indexed in a searchable resource.
A2. metadata are accessible, even when the data are no longer available.
I1. metadata use a formal, accessible, shared, and broadly applicable.
I2. metadata use vocabularies that follow FAIR principles.
I3. metadata include qualified references to other metadata.
R1 (R1.1 - R1.2 - R1.3) Metadata are richly described with a plurality of accurate attributes.

F1. Data are assigned a globally unique and eternally persistent identifier.
F4. Data are registered or indexed in a searchable resource.
I1. Data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. Data use vocabularies that follow FAIR principles.
I3. Data include qualified references to other (meta)data.
R1.1. Data are released with a clear and accessible data usage license.
R1.3. Data meet domain-relevant community standards.

The FAIRness Maturity Pyramid [2].

ENVRI-FAIR Cluster Project and its relation to the EOSC

What is the ENVRI community expecting from the EOSC?

ENVRI-FAIR supports the view of EOSC and its services as a public good. We also consider that long-term and sustained funding is required to ensure the EOSC continues to exist and serve its users. This funding should also reflect the resources requested by the supported communities, with necessary periodic updates and related development initiatives.
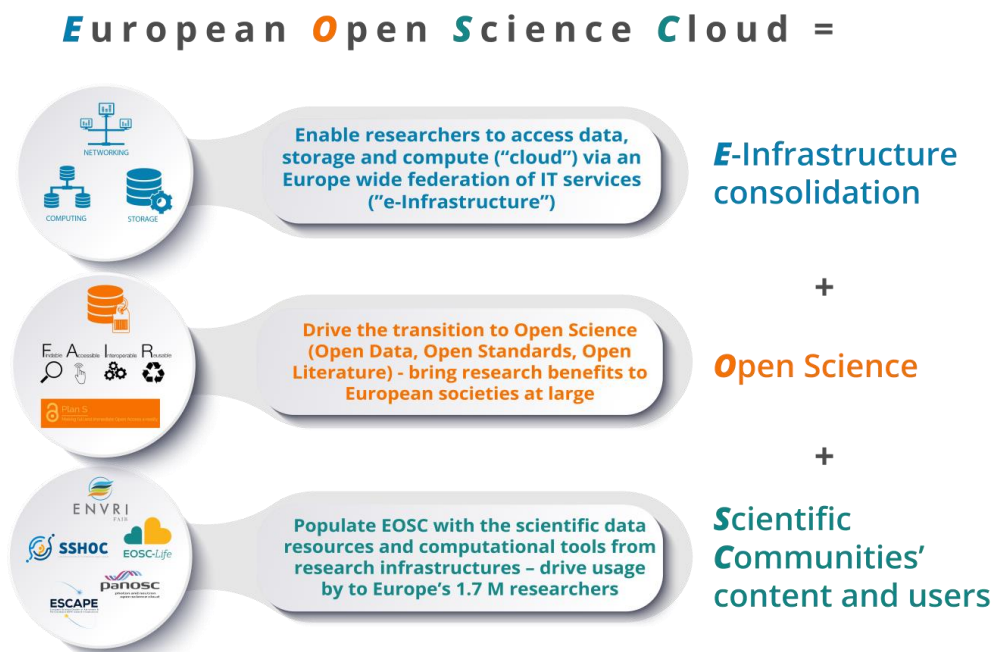
To our understanding, the Federating Core with its components Compliance Framework, Hub Portfolio, and Shared Resources, will serve as the power house of the EOSC, whereas the majority of curation of research data remains the responsibility of the contributing research infrastructures.

We also consider that the Federating Core and Shared Resources can form an essential part of the ENVRI Research Infrastructures operation models, if the sustainability and access issues can be solved.

The ENVRIs play an important and multi-facetted role for EOSC as both providers of data and services of all kinds (e.g., data services, research products services) and as first users of services provided by EOSC. The relationship between RI funding and EOSC-derived funding for the services and resources provided for the RIs need to be detailed, together with the precise specification of services and resources provided for each facility.

In the EOSC ecosystem, the ESFRI research infrastructures organized in five thematic clusters have evolved into key players when it comes to the provision of data and services to their research communities and integrating this to the use of the emerging EOSC infrastructure. The five ESFRI

research infrastructures cluster projects[1] span a broad range of scientific fields: life sciences, environmental sciences, social sciences and humanities, photon and neutron sciences, and astronomy and particle physics. The expectations, needs and offers from the clusters point of view towards the EOSC are very similar and developed into a common view on the EOSC. For the ESFRI cluster projects EOSC is built from three levels: **E**-Infrastructure consolidation *plus* **O**pen science *plus* **S**cientific **C**ommunities' content and users.

E u r o p e a n   O pen   S c i e n c e   C l o u d =

Enable researchers to access data, storage and compute ("cloud") via an Europe wide federation of IT services ("e-Infrastructure")

E-Infrastructure consolidation

+

Drive the transition to Open Science (Open Data, Open Standards, Open Literature) - bring research benefits to European societies at large

Open Science

+

Populate EOSC with the scientific data resources and computational tools from research infrastructures – drive usage by to Europe's 1.7 M researchers

Scientific **C**ommunities' content and users

ESFRI thematic cluster view on EOSC [5]

What are the added values of EOSC to our community?

The primary goal of ENVRI-FAIR is the implementation and further development of data and research services at RI and subdomain (atmosphere, marine, etc.) levels while ensuring the highest possible level of standardisation at the whole environmental domain (cluster) level. ENVRI-FAIR implies motivation to share solutions and strategies, moving towards a shared approach to EOSC. Key added value to the RIs can be the wider findability and accessibility of RI data and research services for the greater public and thus beyond the ENVRI's "traditional" scientific communities. Additionally, the EOSC can offer the basis for the provision of platforms for scientist-developed virtual research environments, more extensive use of shared workflows (including their publication), and access of less resource-rich country researchers to these facilities.

Fundamental infrastructure components and metadata services (AAI, PID, provenance, workflow management, etc.) need to be integrated at sub-domain and RI levels, but also across the entire

---

[1] EOSC-Life (life sciences, https://www.eosc-life.eu/), ENVRI-FAIR (environmental sciences , https://envri.eu/home-envri-fair/), SSHOC (social sciences and humanities , https://www.cessda.eu/About/Projects/Current-projects/SSHOC), PaNOSC (photon and neutron sciences , https://www.panosc.eu/) and ESCAPE (astronomy and particle physics , https://projectescape.eu/)

cluster. EOSC may provide generic solutions that can be tailored to specific RI needs and then adopted by the single RI.

The provision of resources such as repositories, high-performance computing (HPC) and high-throughput computing (HTC) resources and data management resources may foster the FAIRification process of the involved ENVRIs and particularly of those RIs at an early stage of their life cycle.

To fulfil these expectations, the provision of EOSC resources needs to be sustainable. Otherwise, the deactivation of services implemented at the RI level may pose a high risk on the RIs which have adopted EOSC services, because in this case their operational status will be threatened.

Key message for EOSC Boards - both executive and governing boards:

- ENVRIs can provide the EOSC with their amassed collective domain-specific knowledge and competencies that underlie all the data and other services provided.

- Making and sustaining data FAIR as well as sustaining maintenance of the respective infrastructures require resources and expertise which are available at the ENVRI scientific communities.

- Ensuring the coherence of methodologies and technologies for the FAIRification process across subdomains and ensuring the sustained functionality of provided services are essential; these processes can be strongly supported by EOSC resources.

- Sustainable and long-term funding of the observations and the production and provision of high-quality open and FAIR data and services are crucial.

- It is essentially needed that the financial model of the Federating Core must be positioned at a national level where governments have a political interest in encouraging open research and the means to define national policies that can support it.

ENVRI-FAIR Requirements Table [6]

What ENVRI-FAIR needs from the EOSC
Generic infrastructure services such as for AAI, PID, and provenance, for tailoring to specific Research Infrastructure needs and adoption by individual research infrastructures
Generic workflow management tools and services, for tailoring to specific Research Infrastructure needs and adoption by individual RIs
Access to shared resources such as repositories, HPC, HTC and data management tools
Standard APIs to support remote data discovery, access, and sharing
Provision of notebook-based environments which allow to access and integrate data services for the community
What ENVRI-FAIR can offer to EOSC
Collective domain-specific knowledge and competencies that underlie all the data and other services provided by the European ENVRIs
FAIR-based tools and resources for easy and seamless access to environmental data and services provided by the European ENVRIs
ENVRI-hub – a federated machine-to-machine interface to access environmental data and services provided by the contributing ENVRIs

Relevant Links and References

[1] Petzold, A., Asmi, A., Vermeulen, A., Pappalardo, G., Bailo, D., Schaap, D., Glaves, H. M., Bundke, U., and Zhao, Z.: ENVRI-FAIR - Interoperable environmental FAIR data and services for society, innovation and research (Version Camera ready), Proc. IEEE International Conference on eScience 2019, 1-4, DOI: http://doi.org/10.1109/eScience.2019.00038, 2019.

Accessible at https://zenodo.org/record/3462816#.XfjYJHsxlaQ

[2] Bailo, D., Paciello, R., Sbarra, M., Rabissoni, R., Vinciarelli, V., and Cocco, M.: Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructures, Front. Earth Sci., 8, https://doi.org/10.3389/feart.2020.00003, 2020.

[3] Wilkinson, M. D., et al.: The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3, 160018, DOI: 10.1038/sdata.2016.18, 2016.

[4] GO FAIR https://www.go-fair.org

[5] ESFRI thematic cluster view on EOSC https://doi.org/10.5281/zenodo.3631247

[6] EOSC Federating Core Community Position Paper

https://www.eoscsecretariat.eu/eosc-liaison-platform/post/eosc-federating-core-updated-proposals-and-first-draft-community-position

http://tiny.cc/FedCorePPv1