

ENVRI-FAIR - Interoperable environmental FAIR data and services for society, innovation and research

Andreas Petzold
Forschungszentrum Juelich GmbH
Juelich, Germany
a.petzold@fz-juelich.de

Gelsomina Pappalardo
*CNR Institute of Methodologies
for Environmental Analysis*
Potenza, Italy
gelsomina.pappalardo@imaa.cnr.it

Helen M. Glaves
British Geological Survey
Keyworth, United Kingdom
hmg@bgs.ac.uk

Ari Asmi
University of Helsinki
Helsinki, Finland
ari.asmi@helsinki.fi

Daniele Bailo
*Istituto Nazionale di Geofisica e
Vulcanologia*
Rome, Italy
daniele.bailo@ingv.it

Ulrich Bundke
Forschungszentrum Juelich GmbH
Juelich, Germany
u.bundke@fz-juelich.de

Alex Vermeulen
Lund University
Lund, Sweden
alex.vermeulen@icos-ri.eu

Dick Schaap
MARIS B.V.
Voorburg, The Netherlands
dick@maris.nl

Zhiming Zhao
University of Amsterdam
Amsterdam, The Netherlands
z.zhao@uva.nl

Abstract — ENVRI-FAIR is a project of the European Union’s Horizon 2020 programme (EU H2020), connecting the cluster of European Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). The overarching goal of ENVRI-FAIR is that all participating RIs will provide a set of interoperable FAIR data services that enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI cluster to the EOSC. This goal will be reached by: (1) defining community policies and standards across all stages of the data life cycle, aligned with the wider European policies and with international developments; (2) creating for all participating RI sustainable, transparent and auditable data services for each stage of the data life cycle, compliant with the FAIR principles; (3) implementation of prototypes for testing pre-production services at each RI, leading to a catalogue of prepared services; (4) exposing the complete set of thematic data services and tools provided by the ENVRI cluster to the EOSC catalogue of services.

Keywords — *data services, environmental science, research infrastructures*

I. INTRODUCTION

Environmental changes are influenced by natural and anthropogenic processes on all scales from local to global. Understanding and quantification of these changes is a necessary requirement for the development of mitigation and adaptation strategies based on facts. Reliable predictions of environmental change must be based on trustworthy, well-documented observations that capture the entire complexity of the Earth system and the manifold interactions between the atmosphere, the land, and the ocean, including the impacts on all forms of life. Environmental data provides the scientific basis for analysing the physical, biological, and economic processes in the Earth system that affect all sectors of society as well as wildlife and biodiversity. Easy and fast access to reliable, high quality environmental data is fundamental for research, and for the development of

environmental prediction and assessment services. It is also vital for assessing past and defining future policies, as well as for the development of environment-friendly innovations. The demand for Earth system data is rapidly increasing, but the tools to manage, document, find, access, and use such data are still under-developed owing to the combination of data complexity and data volumes.

Environmental research infrastructures (ERI) as defined by the European Strategy Forum on Research Infrastructures (ESFRI¹) are the large producers and providers of environmental research data in Europe collected from in-situ and space-based observing systems. They cover the four main subdomains of the complex Earth system (Atmosphere, Marine, Solid Earth, and Biodiversity/Terrestrial Ecosystems), and emerge as crucial pillars for environmental scientists in their quest for understanding and interpreting the complex Earth system. Those RIs are often the result of many years of development at regional and institutional levels, typically through different projects. They face common challenges in developing effective services for managing data gathered from large collections of regional stations, and making them usable for domain scientists. It is thus strategically important to cluster these European Environmental and Earth System Research Infrastructures together to jointly tackle their common challenges.

This paper will give an overview on the ENVRI cluster, and discuss the recent activities in the context of a newly funded cluster project EU H2020 ENVRI-FAIR².

II. ENVRI-FAIR CONCEPT AND GOALS

ENVRI-FAIR brings together the current ESFRI roadmap RIs and Landmarks from the cluster of ENVRI that form part of the wider ENVRI³ community. ENVRI-FAIR will

¹ www.esfri.eu

² www.envri-fair.eu

³ www.envri.eu

ENVRI-FAIR receives funding from the EC Horizon 2020 research and innovation programme under grant agreement No 824068.

further develop and implement the tools and methods created during the preceding EU-funded ENVRI and ENVRIplus⁴ projects, and thus advance the continuing evolution of the cluster of European Environmental RIs to fully unlock their potential to deliver reliable services that are important for science and society.

A. Vision and Objectives

The overarching goal of ENVRI-FAIR is to promote the findability, accessibility, interoperability and reusability (FAIRness) of digital assets, in particular *data* and *services*, provided by the ENVRI cluster and connect them to the emerging service eco-system of the EOSC⁵. Common policies, open standards, inter-operability solutions, operational services, and stewardship of data on the basis of the FAIR principles [1] at the cluster level essentially reduce development costs for an individual RI to solve those common problems. The final goal is to provide an open access hub for interdisciplinary environmental research data in the European Research Area utilizing the EOSC.

Specifically, ENVRI-FAIR will

1. further develop common standards and policies for the data life cycle, including cataloguing, curation, provenance and service provision within the cluster of ENVRI with specific consideration of the FAIR principles and interoperability within the context of EOSC; these policies and standards are also aligned with wider European policies (e.g. INSPIRE) as well as relevant international developments. Each participating RI will implement sustainable, transparent and auditable data services that are compliant with the FAIR principles;
2. develop and implement the necessary tools for reaching Objective 1 in each RI, thereby adopting an open approach to sharing data and software;
3. improve the skills of RI personnel to develop and maintain the FAIR infrastructures resulting from Objectives 1 and 2 through extensive training;
4. increase the potential for innovation of each RI 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 for enhancing uptake of RI services by private partners;
5. establish cohesion with the global RI landscape, including RI clusters and regional/international initiatives in the environmental sector; maintain ENVRI community knowledge with particular consideration of developing integrated activities;
6. expose thematic data services and tools from the RI catalogues to the EOSC catalogue of services, COPERNICUS, GEO and other end-users.

B. Participating Research Infrastructures

The RIs participating in ENVRI-FAIR represent the core component of the European environmental RI landscape. The main participants are the 11 RIs in the ENVRI cluster identified as ESFRI Projects (Preparatory or Implementation

phase) or Landmarks. ENVRI-FAIR also includes selected new and upcoming environmental RIs in order to ensure their cohesive development as part of the ENVRI cluster throughout future phases of the ESFRI Roadmap, and guarantees that further evolution of these RIs is aligned with the FAIR principles.

The ENVRI-FAIR consortium also includes several technical partners that are either required for developing solutions in one specific subdomain (e.g. SeaDataNet), or for securing the implementation of common tools and standards already agreed in the previous ENVRIplus project. Table 1 summarizes the consortium of participating RIs and other partners.

Other environmental RIs and related communities will be encouraged to engage with the project and utilize the ENVRI-FAIR developments through dedicated project activities. ENVRI-FAIR will implement, further develop and test tools, methods, and services at the RI level with strong synergies across the subdomains and at the ENVRI cluster level. Recognizing the subsidiarity principle, the ENVRI-FAIR implementation strategy will favour the most efficient common solutions at the ENVRI and subdomain levels. Ensuring knowledge transfer from the ENVRIplus project [2] to support the implementation of tools and services in ENVRI-FAIR offers two fundamental benefits: 1) it promotes coherent development of the leading RIs fostering synergies, complementarity and interoperability, and 2) ensures upcoming RIs are aligned with the principles of the ENVRI cluster.

C. Ambition

The environmental RIs contributing to ENVRI-FAIR have developed comprehensive expertise in their respective fields of environmental research, but integration of research activities across the boundaries of applied domain science are still not fully developed. However, the largest uncertainties in our current understanding of climate change, and the related impact on the Earth system, are associated with the complex feedback mechanisms in the climate system and the corresponding responses to these changes at the ecosystem and domain level. To study these complex interactions requires unimpeded use of multidisciplinary data from different domains, with FAIRness of the digital metadata and data an absolute prerequisite.

The high-impact ambition of ENVRI-FAIR is to prepare the foundation for the successful implementation of a virtual, federated machine-to-machine interface to access environmental data and services provided by the contributing ENVRI. This interface, the ENVRI-hub, will be realized as the services across RIs and even between environmental subdomains become progressively more integrated.

Each RI is optimized for a number of selected parameters related to its specific competences, for example, essential climate variables in the atmospheric subdomain [3]. For users that require a broader or full spectrum of environmental parameters, the ENVRI-hub will offer a platform that reflects the complexity and diversity of the ENVRI landscape, while preserving the structure of the individual RI and continuing to fulfil the requirements of their designated communities.

⁴ www.envriplus.eu

⁵ www.eosc-portal.eu

TABLE 1. RESEARCH INFRASTRUCTURES AND TECHNICAL PARTNERS PARTICIPATING IN ENVRI-FAIR

	RI type	Earth System Subdomains	Link
<i>Research Infrastructure</i>			
ACTRIS	Distributed	Atmosphere	www.actris.eu
AnaEE	Distributed	Bio/Eco	www.anaee.com
DANUBIUS-RI	Distributed	Bio/Eco	www.danubius-ri.eu
DiSSCo	Distributed	Bio/Eco	www.dissco.eu
EISCAT_3D	single, multiple sites	Atmosphere	www.eiscat.se
EMSO	Distributed	Marine, Solid Earth	www.emso.eu
EPOS	Distributed	Solid Earth	www.epos-ip.org
Euro-Argo	Distributed	Marine	www.euro-argo.eu
IAGOS	Distributed	Atmosphere	www.iagos.org
ICOS	Distributed	Atmosphere, Marine, Bio/Eco	www.icos-ri.eu
LifeWatch	virtual, distributed	Bio/Eco, Marine	www.lifewatch.eu
eLTER	Distributed	Bio/Eco	www.lter-europe.net
SIOS	regional, distributed	all subdomains	www.sios-svalbard.org
<i>Technical Partner</i>			
SeaDataNet	Federated	Marine	www.seadatanet.org
Univ. of Amsterdam	Systems and Networking lab	IT Expertise	ivi.fnwi.uva.nl/sne

The concept of the ENVRI-hub represents a highly ambitious goal. The ENVRI-hub will be a federated system of harmonized subdomain (Marine, Atmosphere, etc.) or RI-specific systems that include data management and accessplatforms, virtual research environments, and be underpinned by common data policies. The system will be completely open source, modular and scalable. It will leverage the experience available in the consortium and already operational systems. ENVRI-FAIR will provide the means to further develop the ENVRI architecture by building on the concepts from the RDA Data Fabric interest group [4]. Several RIs and institutes involved in ENVRI-FAIR have developed systems that contain elements of this vision including the ICOS Carbon Portal [5] and the TOAR ozone surface observation portal [6].

The ENVRI-hub community metadata and data store will be based on semantic web technology, ontologies and open linked data, allowing integration of the vocabularies and metadata standards developed as part of the implementation work packages in ENVRI-FAIR. The cross-subdomain development will have significant benefits in terms of efficiency and robustness for the ENVRI-hub system. It will enable true interoperability of access to metadata and data objects across the RIs and subdomains and thus facilitate the development of (joint) higher-level services.

By following a strict modular design the developments can be (re-)used in the different RIs. All components should contain couplers to the EOSC services such as Authentication and Authorisation Infrastructure (AAI), data storage solutions and catalogue of services, in order to enable the seamless integration of ENVRI data and services into the EOSC.

III. METHODOLOGY AND WORK FLOW

A. Methodology

The main focus of ENVRI-FAIR is the implementation and further development of RI-level services while ensuring the highest possible level of standardization at the domain level. According to the terminology used here, a service (or collection of services in a workflow) allows a researcher or

other end-user to achieve their objectives. ‘Prototype services’ indicate services being tested by the end-users, i.e. after internal alpha-testing and commonly named beta-testing, which indicates that it works but there may be bugs and performance issues left to be solved. ‘Production services’ indicate services that are released to end users, implying the software has an appropriate technology readiness level (TRL) and there is associated information concerning sustainability. Fig. 1 illustrates schematically the ENVRI-FAIR data and services flow.

The organizational structure must reflect the different levels of interaction between the single RIs, subdomains, and the whole ENVRI cluster with the EOSC and other services using environmental data. RIs offer services directly to users of environmental research data, they are connected within their subdomains and interact at this level with the user community. A key achievement of ENVRI-FAIR, will be RIs connecting to the EOSC and other users of environmental research data through services at the ENVRI cluster level.

B. Work Flow

ENVRI-FAIR begins with recognition of the challenge of connecting RI clusters to the EOSC and environmental research data users. From this starting point the work flow is designed to address three levels of implementation and development; see Fig. 2 for details of the work package structure and Fig. 3 for the schematic work flow.

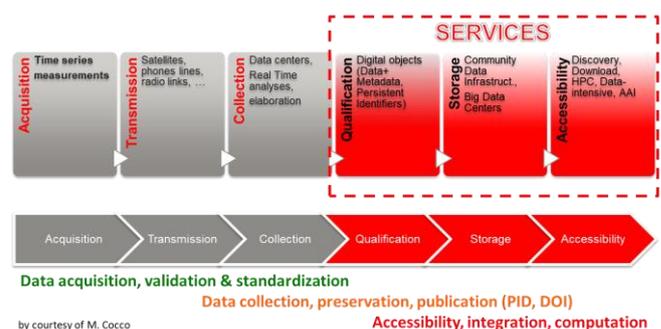


Fig. 1. Schematic ENVRI-FAIR data and services flow.

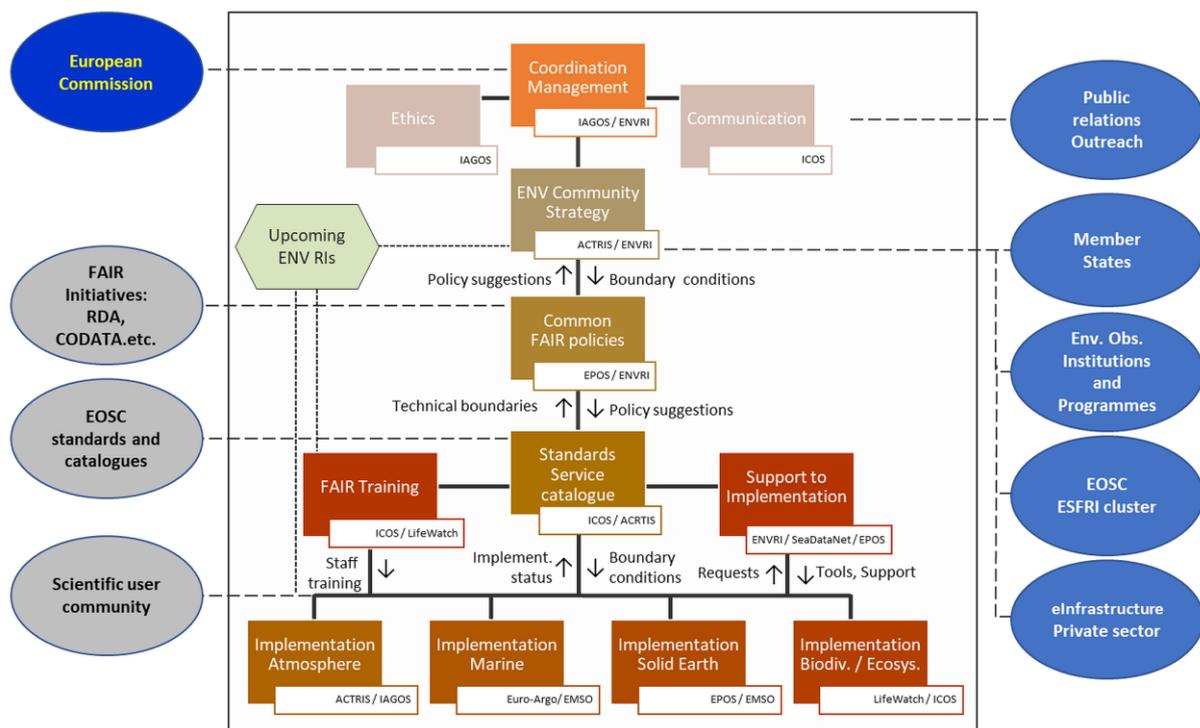


Fig. 2. Schematic structure of the ENVRI-FAIR work flow.

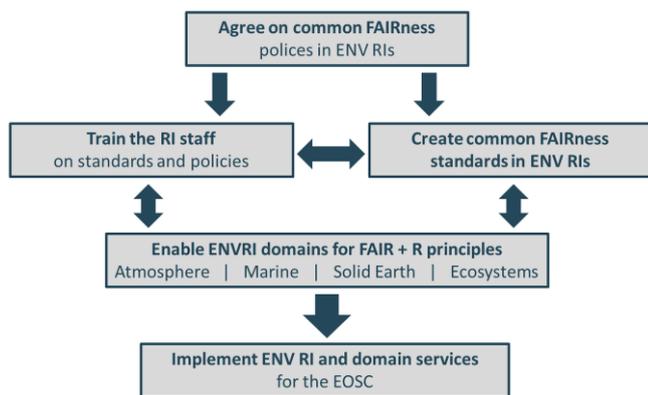


Fig. 3. Schematic structure of the ENVRI-FAIR work flow.

The three levels of the ENVRI-FAIR workflow are:

- i. the RI level: taking into account the fact that participating RIs are at different maturity levels and have individual constraints when implementing FAIR compliant services,
- ii. the subdomain level: (Atmosphere, Marine, Solid Earth, and Biodiversity/Terrestrial Ecosystems) currently recognized as the most-effective science-driven user level including for example institutional users such as COPERNICUS, and
- iii. the domain (or cluster) level which is most effective for addressing standardization with EOSC, e-infrastructure (dominantly EOSC) or with international organizations (GEO, WMO, RDA etc.).

IV. THE TEAM

ENVRI-FAIR is coordinated by a team of data and research specialists from all participating RIs and technical experts across all subdomains of the complex Earth system:

A. Petzold (IAGOS, coordinator), A. Asmi (ENVRI, co-coordinator), U. Bundke (IAGOS), M. Brus (ICOS), P. Laj (ACTRIS), S. Sorvari Sundet (ACTRIS), H.M. Glaves (EPOS), A. Vermeulen (ICOS), G. Pappalardo (ACTRIS), M. Hellström (ICOS), N. Fiore (LifeWatch), J. Konijn (LifeWatch), Z. Zhao (LifeWatch/UvA), D. Schaap (SeaDataNet), D. Bailo (EPOS), C. Lund Myhre (ACTRIS), D. Boulanger (IAGOS), T. Carval (Euro-Argo), J.J. Daño-beitia (EMSO), S. Pouliquen (Euro-Argo), G. R. Baker (EPOS), M. Cocco (EPOS), A. Basset (LifeWatch), D. Papale (ICOS ECO).

REFERENCES

- [1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al., "The FAIR guiding principles for scientific data management and stewardship", *Sci. Data*, 3, 160018, doi: 10.1038/sdata.2016.18, 2016.
- [2] Zhao, Z., Martin, P., Grosso, P., Los, W., de Laat, C., Jeffrey, K., et al., "Reference model guided system design and implementation for interoperable environmental research infrastructures, 1st international workshop on interoperable infrastructures for interdisciplinary big data research (IT4RIs)" In the proceedings of IEEE eScience Munich Germany, p583-588, doi: 10.1109/eScience.2015.41, 2015.
- [3] Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M., "The concept of essential climate variables in support of climate research, applications and policy", *Bull. Am. Met. Soc.*, 95, 1431-1443, doi: 10.1175/bams-d-13-00047.1, 2014.
- [4] <https://rd-alliance.org/group/data-fabric-ig/wiki/df-configuration-pid-centric-data-management-and-access.html>
- [5] www.icos-cp.eu
- [6] Schultz, M. G., Schröder, S., Lyapina, O., Cooper, O. R., Galbally, I., Petropavlovskikh, I., et al., "Tropospheric Ozone Assessment Report: Database and metrics data of global surface ozone observations", *Elementa-Sci. Anthropol.*, 5, 26, doi: 10.1525/elementa.244, 2017.