

<https://helda.helsinki.fi>

ENVRI-fair-interoperable environmental fair data and services for society, innovation and research

Petzold, A.

IEEE
2019

Petzold , A , Asmi , A , Vermeulen , A , Pappalardo , G , Bailo , D , Schaap , D , Glaves , H M , Bundke , U & Zhao , Z 2019 , ENVRI-fair-interoperable environmental fair data and services for society, innovation and research . in 2019 15th International Conference on eScience (eScience) . IEEE , pp. 277-280 , International Conference on eScience , San Diego , California , United States , 24/07/2019 . <https://doi.org/10.1109/eScience.2019.00038>

<http://hdl.handle.net/10138/325616>
<https://doi.org/10.1109/eScience.2019.00038>

cc_by
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

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

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

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

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

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

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

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

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

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

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

Abstract — ENVRI-FAIR is a recently launched project of the European Union's Horizon 2020 program (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 research infrastructures (RIs) will provide a set of interoperable FAIR data services that enhance the efficiency and productivity of researchers, support innovation, enable data- and knowledge-based decisions and connect 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 RIs sustainable, transparent and auditable data services for each stage of the data life cycle, following the FAIR principles; (3) implementing 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 of the ENVRI cluster to the EOSC catalogue of services.

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

I. INTRODUCTION

Environmental changes are the result of a complex interaction of natural and anthropogenic processes on all scales from local to global. Understanding and quantifying these changes is a necessary requirement for the development of mitigation and adaptation strategies. 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 analyzing the physical and biological 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.

The environmental RIs have developed comprehensive expertise in their respective fields of research, but their integration across the boundaries of applied subdomain science is still not fully developed. However, this integration is critical for improving our current understanding of the major challenges to our planet such as climate change and its impacts on the Earth system, our ability to respond to and predict natural hazards, and our understanding and prevention of ecosystem loss. To study these interactions requires unimpeded use of multidisciplinary data from different Earth system sub-domains, with FAIRness [1] of the digital metadata and data as an absolute prerequisite.

II. ENVRI-FAIR CONCEPT AND GOALS

A. The Consortium

ENVRI-FAIR¹ builds on the RIs of the ENVRI community² which are currently listed on the roadmap of the European Strategy Forum on Research Infrastructures (ESFRI³). These RIs form the ENVRI cluster that includes the principal producers and providers of environmental research data and research services in Europe from the four subdomains of the Earth system - Atmosphere, Marine, Solid Earth, and Biodiversity/Terrestrial Ecosystems. The ENVRI cluster represents the core component of the European environmental RI landscape.

¹ www.envri-fair.eu

² www.envri.eu

³ www.esfri.eu

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

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 | Biodiversity/ Ecosystems | www.anaee.com |
| DANUBIUS-RI | Distributed | Biodiversity/ Ecosystems | www.danubius-ri.eu |
| DiSSCo | Distributed | Biodiversity/ Ecosystems | www.dissco.eu |
| EISCAT 3D | Single, with multiple sites | Atmosphere | www.eiscat.se |
| EMSO ERIC | Distributed | Marine, Solid Earth | www.emso.eu |
| EPOS ERIC | Distributed | Solid Earth | www.epos-ip.org |
| Euro-Argo ERIC | Distributed | Marine | www.euro-argo.eu |
| IAGOS AISBL | Distributed | Atmosphere | www.iagos.org |
| ICOS ERIC | Distributed | Atmosphere, Marine, Bio/Eco | www.icos-ri.eu |
| LifeWatch ERIC | Virtual, distributed | Biodiversity/ Ecosystems, Marine | www.lifewatch.eu |
| eLTER | Distributed | Biodiversity/ Ecosystems | 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 |

Table 1 summarizes the consortium. Besides the RIs on the ESFRI roadmap, the consortium includes two technical partners that are required for the implementation of common tools and standards already agreed in the recently completed EU project ENVRIplus⁴ [2].

B. Vision and Objectives

The overarching goal of ENVRI-FAIR is to advance the findability, accessibility, interoperability and reusability (FAIRness) of digital assets, in particular *research data* and *research services*, provided by the ENVRI cluster, and to connect them to the emerging service ecosystem of the EOSC⁵. Common policies, open standards, interoperability solutions, operational services, and stewardship of data on the basis of the FAIR principles [1] at the cluster level will essentially reduce development costs for an individual RI to solve those common problems. ENVRI-FAIR finally aims to provide an open access hub for interdisciplinary environmental research data 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 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 wider 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 RI, thereby adopting an open approach for sharing data and software;
4. improve the skills of RI personnel to develop and maintain the FAIR infrastructures resulting from Objectives 1 and 2 through extensive training;

5. 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 enhancing the uptake of RI services by private partners;
6. 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;
7. expose thematic data services and tools from the RI catalogues to the EOSC catalogue of services, COPERNICUS, GEO and other end-users.

C. Ambition

The high-impact ambition of ENVRI-FAIR is to establish the foundation of a successful implementation of a virtual, federated machine-to-machine interface to access environmental data and services from the contributing RIs. 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 single platform that hides 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.

After its full implementation, the ENVRI-hub will be a federated system of harmonized subdomain or RI-specific systems that include data policies and management as well as access platforms through virtual research environments, and will be underpinned by common data policies; see Fig. 1 for the schematic ENVRI-FAIR architecture.

⁴ www.envriplus.eu

⁵ www.eosc-portal.eu

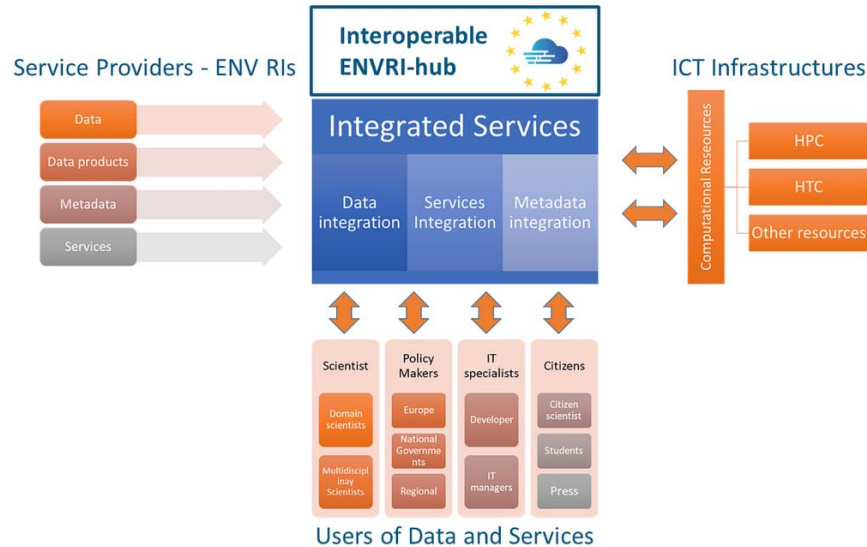


Fig. 1. ENVRI-hub architecture.

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 cluster by building on the concepts from the RDA Data Fabric interest group [4]. Several RIs involved in ENVRI-FAIR have developed systems that contain elements of this vision including, e.g., the ICOS Carbon Portal [5], the TOAR ozone surface observation portal [6], or the EPOS Thematic Core Services [7].

III. METHODOLOGY AND WORKFLOW

A. Methodology

The main focus of ENVRI-FAIR is the implementation and further development of RI- and subdomain-level services while ensuring the highest possible level of standardization at the cluster 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.

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 development across subdomains 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 on cluster level.

By following a modular design the developments can be (re-)used in the different RIs. All components will contain

couplers to EOSC services such as Authentication and Authorization Infrastructure (AAI), data storage solutions and catalogue of services, to enable the seamless integration of ENVRI data and services into the EOSC.

B. Workflow

The ENVRI-FAIR workflow is basically organized as a bottom up process with few but central coordinating and harmonizing activities (see Fig. 2), and addresses the following three levels of implementation and development:

- i. *RI level*: main level of service implementation, taking into account that participating RIs are of different maturity and have individual constraints when implementing FAIR compliant services;
- ii. *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;
- iii. *domain (or cluster) level*: most effective for addressing standardization with EOSC and e-infrastructures (dominantly EOSC) or with international organizations (GEO, WMO, RDA etc.).

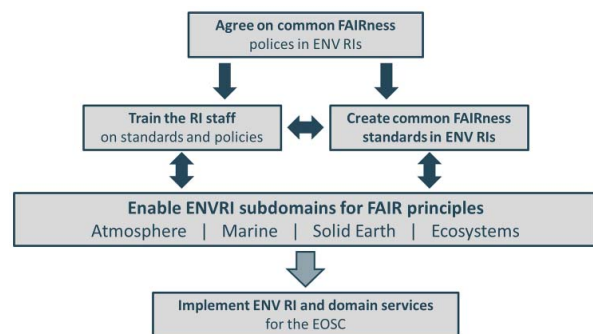


Fig. 2. Schematic structure of the ENVRI-FAIR workflow; arrows indicate directions of interaction.

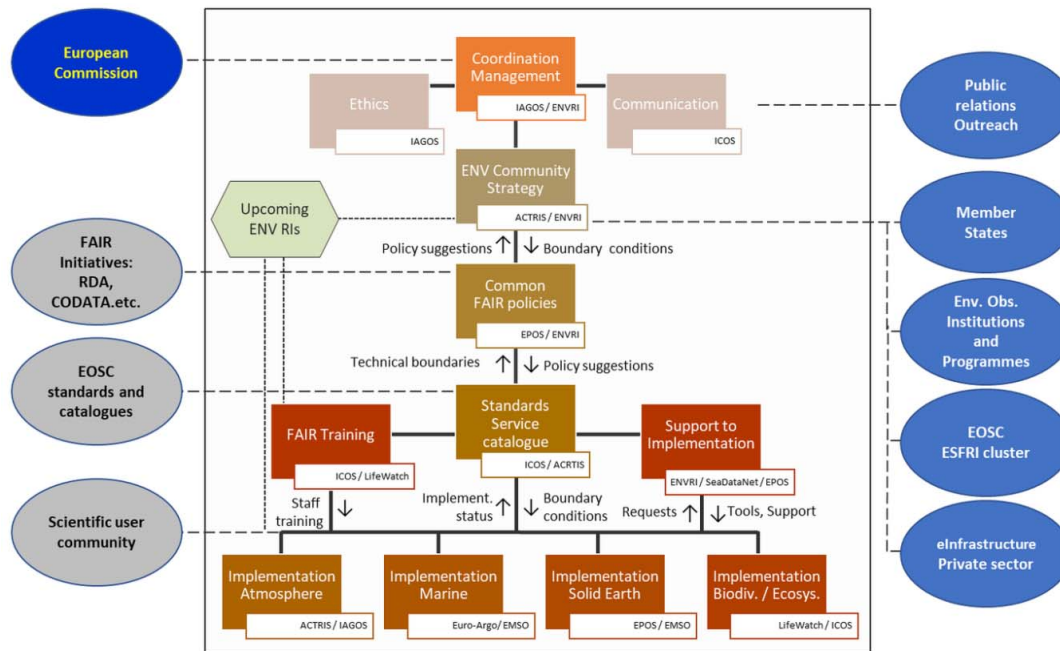


Fig. 3. ENVRI-FAIR work package structure.

The ENVRI-hub architecture as shown in Fig. 1 and the work package structure of ENVRI-FAIR illustrated in Fig. 3 mirror the three levels of implementation and development introduced above, and 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, but they are also connected within their subdomains and interact at this level with the user community. All levels of interaction with the user communities are maintained in ENVRI-FAIR.

C. Start-up Activities

During the start-up phase of ENVRI-FAIR, FAIRness strengths and weaknesses of each RI are evaluated using a standardized questionnaire. The process is inspired by the FAIRification process adopted by GO FAIR⁶ and is applied in accordance with specific needs of the ENVRI cluster and those of the other domains of the ESFRI roadmap. Solutions to address identified weaknesses are then developed to provide coordinated FAIRness roadmaps for the subdomains.

A cluster-level knowledge base is built in order to share technical practices, identify common data and service requirements and design patterns, and facilitate search and analysis of existing RI solutions for interoperability challenges that are shared among environmental RIs.

For each subdomain, reference material for both data license and data policy used by the contributing RIs is collected. Starting from the present state, recommendations for licenses on data and metadata and respective policies are developed and implemented.

⁶ <https://www.go-fair.org>

IV. THE TEAM

The ENVRI-FAIR coordination team includes data and research specialists and technical experts from all RIs: A. Petzold (IAGOS), A. Asmi (ENVRI), 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. Dañoibeitia (EMSO), S. Pouliquen (Euro-Argo), G. R. Baker (EPOS), M. Cocco (EPOS), A. Basset (LifeWatch), D. Papale (ICOS).

REFERENCES

- [1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., 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., 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., 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.
- [7] <https://www.epos-ip.org/thematic-core-service-index>