D5.2
Implementation plan for common development goals

Work Package	WP5
Lead partner	ICOS ERIC
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	2020-06-30
Submission date	2020-09-30

Deliverable abstract

During the first year of the project the cluster worked on the data service requirement tracking and a technology landscape and gap analysis on the RI FAIR data/services level, to provide an up-to-date investigation of the most common gaps the ENVRIs need to bridge with respect to the FAIR principles and the EOSC requirements. The FAIRness analysis (D5.1) established that the participating RIs display a wide range of states of readiness. The present document uses this analysis to explain the actions that have been taken by the ENVRI-FAIR partners during the second year of the project, and to define the common development goals and prepare an implementation plan at cluster level.

During the WP5 Review Workshop in M10 of the project, a decision was made to form (six) cross-cutting Task Forces, which involve representatives from all subdomains, bringing together technical, scientific and managerial staff in order to coordinate the work on a common platform. The development of the ENVRI catalogue of services and other relevant technical issues, e.g. the authentication and authorisation of the cluster end-users, the role of persistent identifiers, triple stores and the best practices, the benefits of certifying data repositories and the available options, the technical approach of machine-readable licenses and their use for citation and usage tracking, and the design of interoperable demonstrators and cross-domain service prototypes, are introduced and discussed in the text.

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

DELIVERY SLIP

	Name	Partner Organization	Date
Main Authors	Angeliki Adamaki Alex Vermeulen	ULUND/ICOS ERIC ULUND/ICOS ERIC	
Contributing Authors	WP5 Task Force leaders WP5 Task Force members WP Representatives	ENVRI-FAIR ENVRI-FAIR ENVRI-FAIR	
Reviewer(s)	Zhiming Zhao Ari Asmi	UvA UHEL	
Approver	Andreas Petzold	FZJ	2020-09-30

DELIVERY LOG

Issue	Date	Comment	Author
V 0.1	2020-08-17	First draft of the document sent for review	Angeliki Adamaki Alex Vermeulen
	2020-08-24	First draft of the document distributed to WP5 participants for feedback	Angeliki Adamaki Alex Vermeulen
	2020-08-25	Comments from reviewer 1 received	Zhiming Zhao
	2020-08-25	WP7 invited to comment on the document	Zhiming Zhao
	2020-09-03	Reminder sent to WP5 participants	Angeliki Adamaki
	2020-09-07	Reminder sent to reviewers	Angeliki Adamaki
	2020-09-16	Comments and suggestions	Margareta Hellström
V 1.1	2020-09-21	Second draft of the document distributed to WP5	Angeliki Adamaki Alex Vermeulen
	2020-09-25	Reminder sent to reviewers	Angeliki Adamaki
	2020-09-25	Comments from reviewer 2 received	Ari Asmi
V 2.1	2020-09-28	Final version sent to Coordinators	Angeliki Adamaki Alex Vermeulen

DOCUMENT AMENDMENT PROCEDURE
Amendments, comments and suggestions should be sent to the Project Manager at manager@envri-fair.eu.

GLOSSARY
A relevant project glossary is included in Appendix A. The latest version of the master list of the glossary is available at http://doi.org/10.5281/zenodo.3465753

PROJECT SUMMARY
ENVRI-FAIR is the connection of the ESFRI Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). Participating research infrastructures (RI) of the environmental domain cover the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems and thus the Earth system in its full complexity.

The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which 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 is reached by: (1) well defined community policies and standards on all steps of the data life cycle, aligned with the wider European policies, as well as with international developments; (2) each participating RI will have sustainable, transparent and auditable data services, for each step of data life cycle, compliant to the FAIR principles. (3) the focus of the proposed work is put on the implementation of prototypes for testing pre-production services at each RI; the catalogue of prepared services is defined for each RI independently, depending on the maturity of the involved RIs; (4) the complete set of thematic data services and tools provided by the ENVRI cluster is exposed under the EOSC catalogue of services.

TABLE OF CONTENTS

D5.2 - IMPLEMENTATION PLAN FOR COMMON DEVELOPMENT GOALS

1 Summary

The main objectives of the ENVRI-FAIR Work Package 5 (WP5) are to:

- identify the (ENVRI cluster) common development targets for (meta)data services, based on the gap analysis performed at RI and subdomain level,
- guide (together with WP7) the harmonisation at cluster level of the required actions the RIs need to take to improve their FAIRness level,
- prepare the ENVRI Catalogue of EOSC services,
- design and provide the guidelines for the validation of these services, and work with EOSC to formulate a strategic roadmap for future development.

During the first year of the project the cluster worked on the data service requirement tracking and a technology landscape and gap analysis on the RI FAIR data/services level. This to provide an up-to-date investigation of the most common gaps the ENVRIs need to bridge and the improvements they should make with respect to the FAIR principles, while reviewing the EOSC requirements and their impact. The previous WP5 Deliverable (*D5.1 Requirement analysis, technology review and gap analysis of environmental RIs*) presented the first evaluation of the status of the ENVRIs and the needed steps to improve their FAIRness (see section 2 for a summary and relevant references). The present document uses this analysis to explain the actions that have been taken by the ENVRI-FAIR partners during the second year of the project, and to define the common development goals and prepare an implementation plan at cluster level. The FAIRness analysis laid out in D5.1 established that the participating RIs display a wide range of states of readiness. A summary of the identified main actions and the priorities to improve the general level of readiness in each subdomain is given in section 3. There are many differences between the subdomains, but many common characteristics can also be identified. The ENVRI cluster is working on some common developments, with suggestions from WP5 and WP7, using the experience acquired with solutions developed by the different RIs and the training opportunities organised by WP6.

During the WP5 Review Workshop in M10 of the project, a decision was made to form (six) cross-cutting Task Forces (see section 4), which involve representatives from all subdomains, bringing together technical, scientific and managerial staff in order to coordinate the work on a common platform on the following specified topics:

- Design of the ENVRI Catalogue of services
- AAI implementation
- PIDs, identification types and registries
- Triple stores and data storage certification
- Licences, citation and usage tracking
- User oriented cross-domain demonstration cases

The development of the ENVRI catalogue of services is orchestrated by a dedicated working group (see section 4.1) and is of course already a separate task in WP5. The ENVRI catalogue will include information regarding resources provided by each RI in the environmental cluster. These can be datasets, web services, Graphical User Interfaces (GUIs) or portals, software, computational services, equipment, workflows and others. Different definitions of “services” (as well as metadata for data and/or services) seem to exist, thus convergence on a common definition is one of the first steps for this task force. The current suggestion is to use the term “service” not only with a technical meaning, but also to describe “everything that implements a functionality that serves a purpose”. It is generally agreed that the ENVRI catalogue should map machine readable assets. The heterogeneity between the participating ENVRIs (e.g. in terms of accessing datasets, data formats and semantics) brings along major challenges when a common approach of an ENVRI catalogue is attempted. To achieve the required interoperability within the cluster, a rich metadata schema is fundamental.

Other relevant technical issues, e.g. the authentication and authorisation of the cluster end-users, the role of persistent identifiers, triple stores and the best practices, the benefits of certifying data repositories and the available options, the technical approach of machine-readable licences and their use for citation and usage tracking, and the design of interoperable demonstrators and cross-domain service prototypes, are discussed in sections 4.2-4.6. To formulate guidelines for validating the overall quality of software development of the ENVRI-FAIR services, an active dialogue with EOSC concerning relevant EOSC integration criteria and EOSC developments is a key component. Most of the WP5 participants represent the ENVRI community in several EOSC groups and related initiatives, delivering valuable feedback to the ENVRI-FAIR partners. Relevant references and links can be found as footnotes throughout the document. The next steps to be followed by the ENVRI community are described in section 5.

2 Introduction

Within the ENVRI-FAIR project the overarching goal of the participating RIs is to build FAIR-based data services which will be provided to the ENVRI community end-users, which can be policy makers, research communities and data users or other service providers, as e.g. the European Open Science Cloud (EOSC) and Copernicus. To start with, some of the FAIR principles[1] have been translated into requirements, to which the repositories and their data, products and services should adhere. Such requirements define guidelines for further development. The work of the ENVRI community towards more FAIR data and services takes place at different levels, starting with the individual RIs that are grouped into the 4 environmental subdomains (Atmosphere, Marine, Solid Earth and Biodiversity/Ecosystems) to form the ENVRI cluster. The ENVRIs have documented the steps they need to follow to improve their FAIRness and successfully complete the implementation actions as those are agreed at subdomain level.

Table 1. Summary of the Rules of Participation (RoP) for the EOSC, as mentioned in the version 0.2 draft[2] which has been distributed for discussion by the RoP WG[3] of the EOSC since January 2020. The aim of the document is to provide the conceptual framework for the participation policy and will be finalised taking also into account the recommendations that are currently being developed by other EOSC WGs. The rules will apply to all digital resources made accessible via EOSC.

Rules **o**f **P**articipation (**RoP**) for the **E**uropean **O**pen **S**cience **C**loud (**EOSC**)	
Ground Rules **G1.** EOSC is open to all **G2.** EOSC resources are registered in an EOSC recognised catalogue	
Data	Services
D1. Data resources exposed through EOSC are free of charge at the point of access	**S1.** Services exposed through EOSC are free of charge at the point of access
D2. Data products adhere to principles of proper research conduct	**S2.** Service providers adhere to principles of proper research conduct
D3. Data providers determine the terms of use of data resources	**S3.** Service providers determine and publish the conditions of use of their services
D4. Data providers will respect principles of FAIR data	**S4.** Services align with EOSC service architecture
D5. Data users adhere to the terms of use of data resources	**S5.** Service users adhere to the terms of use of the services they consume
D6. Data users reference the source	**S6.** Service users reference the source

[1] GO FAIR – FAIR Principles explained https://www.go-fair.org/fair-principles/
[2] Draft RoP version 2, last visited August 2020, https://repository.eoscsecretariat.eu/index.php/s/QWd7tZ7xSWJsesn#pdfviewer
[3] EOSC Rules of Participation WG https://www.eoscsecretariat.eu/working-groups/rules-participation-working-group

The challenge of ENVIR-FAIR is to work together with a diverse set of Research Infrastructures that have in common that they work on environmental observations, but also differ with regards to the level of operationality (some are just starting and designing, others have mature operational systems in place that need to continue undisturbed), most are single-domain but some are multi-domain, all have as reason of existence their own dedicated users and bread-and-butter services for these users and most already have connections or even are an integral part or leading global research infrastructures in their areas. Most of them are distributed which contributes to the complexity of maintaining and building systems in different locations and cultures. And most of them have problems with the sustainability of the operational distributed networks because of the often national funding situation. And now next to the integration of national networks into global RIs comes an additional challenge of interfacing the data and metadata with a moving target called EOSC. On the other hand, the ENVRI community has already been building strong foundations in the last 10 years to overcome these challenges, and this also forms a perfect basis to also successfully integrate with the EOSC. The participation in EOSC dictates just yet another sets of actions[4], defining requirements for service providers and other roles within EOSC, as well as clarifying definitions and setting standards for the type of services the participants will contribute with. Such a dynamic system introduces more levels of complexity to an ongoing project of a scientific community which already includes high heterogeneity. The ENVRI community however is motivated to overcome these challenges and constructively bridge the differences between the participating RIs while interacting within the project to determine a common strategy to achieve their goals.

2.1 First FAIRness assessment

During the first year of the project, the ENVRIs performed an analysis to estimate the current level of FAIRness of their repositories, with help from technical experts and in collaboration with other international initiatives in data management. 34 repositories participated in the survey, coming from the 4 environmental subdomains. Summarising the actions taken, the WP5 partners:

- designed the FAIRness assessment approach using questionnaires
- revised the approach to obtain more structured and unambiguous machine-readable answers
- developed a YAML[5] template and reference lists for the answers from the questionnaires
- organized subdomain workshops on FAIRness assessments and guided the RIs
- extracted the key information from RI answers
- analyzed the FAIRness evaluation results and presented them explicitly in relevant documents.

The responses from 11 ENVRIs (with a total of 30 repositories) where interpreted and described in detail in the deliverable D5.1 “Requirement analysis, technology review and gap analysis of environmental RIs”[6]. Specifically, the analysis refers to 12 Atmospheric RIs, 6 from the Marine subdomain, 8 from the Solid Earth and 8 from the Ecosystem subdomain (another 4 repositories were at a planning stage at the time of the first evaluation). Every step of the process and a run-down of the results at subdomain and cluster level were presented in the D5.1 document, together with a description of the background work the project partners did while preparing the appropriate methodology to adopt (for the cluster and the project), considering that there are several tools to use in order to evaluate the FAIRness level of a repository (with quantitative or qualitative approaches). Detailed information of the FAIRness status of each RI was made available internally through the project communication channels. This first round of FAIRness assessment will serve as a reference point for future analyses which are planned within the project, to better record the progress made by the project partners in terms of improving their FAIRness level. A summary of these results is given below (**Table 2**). The main remarks on the gap analysis are given in the following section.

[4] EOSC Executive Board, Work Plan 2019-2020 https://op.europa.eu/s/ob9w
[5] YAML files with FAIR results for ENVRIs https://github.com/envri-fair/fairness-assessment/descriptions
[6] ENVRI-FAIR Deliverable D5.1 DOI: 10.5281/zenodo.3884998

Table 2. Summary of the main results (per FAIR Principle-group) from the first FAIRness assessment analysis in the ENVRI cluster. The level of compliance (when possible to estimate) is displayed. Detailed results at subdomain and cluster level are given in the Deliverable D5.1. The results at RI level have been made available to the ENVRI-FAIR partners through supplementary material.

Findability	**A**ccessibility	**I**nteroperability	**R**eusability
23 Identifier Systems	22 HTTP access protocols, 1 FTP	All (but 1) RIs use machine-readable exchange formats	4 data usage licenses are reported
48% of the repositories use unique PIDs	46.7% of the repositories make statements on access policy in metadata	Diverse metadata schemas are used - most of them FAIR compliant	14 ENVRIs have not provided license information
33.3% of the cluster provide working IRIs to machine-readable metadata documents	No longevity plans	Diverse vocabularies – most of them FAIR compliant	1 RI provides FAIR compliant machine-readable provenance information about data
50% of the cluster include PIDs in their metadata description		40% of the repositories have categories in their schemas which are defined in open registries	
63.3% of the cluster provide search on data			
14 registries for the repositories			

2.2 Gap analysis

From the FAIR principles to the FAIR practices, gaps might be identified when there is need for further development to reach a different improved level, with practices that are compliant with the FAIR principles. In the case of the ENVRI-FAIR project, the gaps can be identified at different levels (and as a result addressed and handled from different perspectives) within the FAIR principles framework.

The gap analysis for the ENVRI cluster has been derived from the FAIR assessment activity that involved all participating ENVRIs. The deliverable D5.1 presented the results and a synthesis of the common technical requirements at the cluster level (see **Table 3**). These form the basis for the facilitation by WP7 of the implementation activities planned to be undertaken by the RIs in the upcoming years, to achieve FAIRer data and services. During the lifetime of the ENVRI-FAIR project, the FAIRness level of the ENVRIs will be measured again, presumably in the middle period and before the end of the project. The findings of the gap analysis of the middle period will also be considered in the cluster implementation plan.

Each RI used the observations made by WP5 to discuss its identified gaps and strengths, to consolidate common implementation strategies at subdomain level and to compile its reports. Some of the most important points in the subdomain implementation plans are briefly discussed in the next paragraphs.

Table 3. Summary of the main gaps (per FAIR Principle-group) as observed during the first FAIRness assessment analysis in the ENVRI cluster. Specific recommendations per RI and FAIR sub-principle are given in Deliverable D5.1.

Findability	**A**ccessibility	**I**nteroperability	**R**eusability
PIDs not fully implemented	Standardized solutions for access protocols	Harmonisation of metadata standards	Machine-readable usage license needs to be part of a common strategy
PIDs should be included in metadata	Access policy needs to be mentioned in the metadata	Need for mapping and wider use of vocabularies	Machine-readable provenance information needs to be implemented
Harmonisation of metadata and metadata discovery is required	Authentication and Authorisation	Register schemas in common registries	Metadata should be included in provenance
Need for machine-readable metadata	Need for machine-readable metadata longevity plans	Categories in metadata marked up with vocabularies	Compliance validation service
Need for registered metadata or indexed in searchable resources			

3 ENVRI subdomains: Implementation plans

In this section a short statement of the focus points of the subdomain implementation plans is given, based on the reports[7] of the 4 environmental subdomains delivered during the first 18 months of the ENVRI-FAIR project. The subdomains have used the first year FAIRness assessment results at RI level to define their common priorities, taking into account the importance of each function for their end-users, the different maturity levels of their participants as well as the objectives of each subdomain WP to achieve the harmonisation of their service and meet the interoperability requirements as those are defined within the ENVRI cluster.

3.1 Atmosphere

Five ENVRIs collaborate in the atmospheric subdomain, namely ACTRIS, EISCAT, IAGOS, ICOS (atmosphere), SIOS (atmosphere). Following their FAIRness gap analysis during the first months of the project (see D8.1[8] for results), the atmospheric RIs prepared their individual implementation plans aiming to increase the FAIRness level of the atmospheric data and metadata. Utilizing technical solutions which are common for the subdomain, the atmospheric RIs first focus on the technical standards which are considered to be the most important and thus need to be implemented immediately. A list of the WP8 implementation priorities as presented in the D8.3[9] document includes:

- consolidation of consistent use of PIDs throughout data production workflow
- common standard interfaces for metadata and data access
- indexing of data resources
- domain vocabulary/ontology for observed parameters, discovery and use metadata
- common use of authentication schemes
- consistent documentation of provenance throughout data
- recommendations for licenses on metadata and data
- semantic search for atmospheric RI user interfaces
- improve Graphical User Interfaces (GUIs)

[7] D8.1, D9.1, D10.1, D11.1; List of ENVRI-FAIR Deliverables https://envri.eu/deliverables/
[8] ENVRI-FAIR Deliverable D8.1 DOI: 10.5281/zenodo.3885160
[9] ENVRI-FAIR Deliverable D8.3 DOI: 10.5281/zenodo.3885240

WP8 has grouped these tasks into 3 categories, depending on their importance for the subdomain common strategy, and thus prioritized their implementation also considering the specification maturity in each case. Conclusions derived from the ENVRIplus project provide the necessary basis for further developments in the subdomain, while interaction with the ENVRI-FAIR WP5 and WP7 contributes to a productive synergy within the cluster.

3.2 Marine

The marine subdomain involves Euro-Argo, EMSO, ICOS (marine), LifeWatch and SeaDataNet. Starting with an evaluation of their FAIRness status, the marine RIs delivered a roadmap[10] during the first year of the project to list their main priorities in becoming more FAIR. Based on the identified gaps in their FAIRness, the RIs will focus mostly on their back-end services and deal with issues regarding Interoperability and Reusability, from a machine-to-machine perspective, as those will consequently improve their front-end services at subdomain and cluster level. The implementation plan of the subdomain can be found in the D9.2[11] document. Each RI maps their technical conditions concerning data services, building on the mapped user requirements, and plans the necessary moves for FAIRness improvement. Those might include upgrades of already existing services, or development of new ones depending on each RI.

One of the outcomes of the ENVRI-FAIR project for the Marine subdomain will be the implementation of a subdomain demonstrator, which will be delivered at later steps of the project. The harmonisation of the actions taken from the marine RIs is a prerequisite for this. The harmonisations concern vocabularies, machine-to-machine access to data and metadata, etc. More specifically, in the WP9 Deliverable D9.3[12] each RI explicitly documents the technical specifications of their machine-to-machine services, the interfaces to be designed for accessing data and metadata for their implementation at RI level, and discusses the technical choices as well as priorities in the development process. The planned activities can be found in:

- the upgrade of existing machine-to-machine services and interfaces, by either improving or adding features,
- the development of newly defined services where necessary,
- other operations which will be focused on the upgrade of the shared marine (meta)data as those are published via the subdomain services, to improve their interoperability and re-usability.

Regarding the subdomain demonstrator, the subdomain identified some global pre-requirements to be taken into consideration, including:

- the use of harmonised parameters/vocabularies (with at least the defined essential variables for the marine RIs) enhancing the NERC Vocabulary Server,
- the accessibility of the RIs data and metadata which will allow for machine-to-machine interaction through the appropriate services (e.g. APIs),
- the redirection of the user requests to access the individual RIs when necessary, and to include crucial provenance information into the data files.

3.3 Solid Earth

The subdomain involves the European Plate Observation System (EPOS), which represents RIs, data centres and repositories operating in the field of solid earth, as well as EMSO and LifeWatch, which participate in more than one environmental subdomain.

EPOS in particular is organised into thematic cores. The architecture is based on a micro-services approach. Internally, the components communicate using standard APIs. The subdomain catalogue is maintained within a relational Database Management System (based on Postgres). The Common

[10] ENVRI-FAIR Deliverable D9.1 DOI: 10.5281/zenodo.3885296
[11] ENVRI-FAIR Deliverable D9.2 DOI: 10.5281/zenodo.3885327
[12] ENVRI-FAIR D9.3 DOI: 10.5281/zenodo.3885330

European Research Infrastructure Format (CERIF) in the EPOS catalogue facilitates the conversion to different metadata formats and using a Resource Description Framework (RDF) allows for interoperation with systems which choose a linked data structure for their metadata representation.

The FAIR roadmap for the Solid Earth aims to match technical solutions to the subdomain goal for implementing the FAIR principles and improving the interoperability in the ENVRI cluster and further. The approach for the solid earth subdomain organises the implementation into four stages that concern data, metadata, access and (re-)use (see D10.1[13] for more information). Based on the gap analysis for the FAIRness status of the solid earth subdomain, the Integrated Core Services Central hub (ICS-C for short) infrastructure will focus on:

- metadata issues,
- authentication and authorisation solutions,
- other developments to ensure the interoperability of the EPOS Thematic Core Services (TCS), as well as
- the preparation of a demonstrator within a Virtual Research Environment (VRE), in the form of a visualisation and analysis web platform.

The implementation plan for Solid Earth is described in the WP10 D10.2[14] document, where the subdomain activities are explained as they take place at different levels within the EPOS structure (e.g. at TCs, ICS-D etc.) and as defined by the tasks of the ENVRI-FAIR project. For example:

- data and metadata formats coming from different providers need to be in agreement among the thematic cores,
- convertors to and from CERIF are necessary to support interoperability with other systems using different metadata schemas,
- in depth comprehension of the required metadata elements to describe workflows and e-Infrastructures will contribute significantly to the automation of further workflow construction,
- regarding the authentication and authorisation services, additional metadata elements should be considered for each digital asset described in the catalogue, as well as the appropriate access control software at the ICS-C to authorize access to services in the TCS communities and other external e-Infrastructures mentioned as ICS-Distributed (ICS-D) (the ICS-D will allow workflows to be executed utilising other e-Infrastructures).

To achieve harmonisation with the EMSO services, the implementation priorities and technical activities concern:

- the implementation of a harmonisation abstraction layer (metadata and integration processes),
- the quality control support on seismological data and data product generation,
- the enrichment of metadata and establishment of an agreed workflow to enhance the integration and improve the visibility of EMSO contributions through its regional facilities,
- the development of an appropriate FAIRness assessment process ideally based on both qualitative and quantitative methods.

3.4 Biodiversity and Terrestrial Ecosystems

Seven RIs constitute the Biodiversity and Ecosystem subdomain: AnaEE, DANUBIUS-RI, DiSSCo, eLTER, ICOS (ecosystem), LifeWatch and SIOS (ecosystem). The subdomain is characterized by a diverse scientific orientation and a large number of variables of interest. The first analysis of the FAIRness level of the subdomain RIs revealed the heterogeneity of the ecosystem RIs, as those were found in a broad range of FAIRness. Information on all involved RIs and the subdomain short-term implementation plan has been given in the WP11 D11.1[15] document. The rather inhomogeneous maturity of the RIs, ranging from fully operational to newly introduced to the roadmap infrastructures,

[13] ENVRI-FAIR D10.1 DOI: 10.5281/zenodo.3885335
[14] ENVRI-FAIR D10.2 DOI: 10.5281/zenodo.3925503
[15] ENVRI-FAIR D11.1 DOI: 10.5281/zenodo.3885361

is an important factor the subdomain and the ENVRI cluster in general need to consider. The variations introduce more levels of complexity especially when the development of demonstration cases of data FAIRness is considered. The harmonisation of the strategies followed for data description and access within the ENVRI-FAIR project is the primary goal for the subdomain. Common activities will concentrate on producing demonstration cases through integrative approaches of the involved RIs, in the framework of addressing specific research questions.

As a starting point, three case studies will be orchestrated, where two variables/properties will be selected (soil water content and species, scientific names and identification), taking into account the description of sites which are common across the RIs. The improvement of FAIRness and cross-RI access will be tested in these cases.

4 ENVRI cluster: Cross-domain priorities

Following the gap analysis, the subdomains, together with the RIs, plan their actions to meet the FAIR requirements. The task of WP5 and WP7 is to support the common development targets and properly define and prepare the output ENVRI catalogue of services. The cluster, with help from WP5, should design and provide the guidelines for the validation of these services and work together with EOSC to formulate a strategic roadmap for future development. During the ENVRI-FAIR project, issues regarding the ENVRI-hub and how this will be built (e.g. as a federated virtual hub) are (and will be further) discussed among all subdomains, to define and harmonise the necessary common solutions.

As explained in the previous sections and documented in the deliverable D5.1, the readiness state of the participating RIs covers a wide range. In principle, there are technical solutions which are already available and the technical experts from WP7 will help the RIs to implement them[16] (see also the demonstrators grouped by FAIR principles in the Knowledge Base[17]). The starting communities will be benefited from the activities organized by the WP6 team, to train their personnel.

In M10 of the project (October 2019), a workshop was organised by WP5 to discuss the preliminary results of the FAIRness assessment conducted by the ENVRIs, identify the technical gaps which are common for all partners, and plan the required actions to bridge these gaps for the cluster as a whole. With representatives from all subdomains, as well as the technical and management WPs, the workshop proved to be a successful means of interaction and communication, giving the required space to the ENVRI partners to address the common gaps identified among the subdomains, get direct feedback from the technical experts, and discuss the basic lines of a common strategy for the continuation of the project.

The preliminary interpretation of the gap analysis results per RI and subdomain shows that there are places for improvement in all areas of FAIRness. The synchronisation of all efforts by the ENVRIs, both at subdomain and cluster level, introduces a higher level of complexity due to the heterogeneity described in the previous sections. To build on a common ground, rather working in directions that differ per subdomain, the project partners prioritized the cluster goals in connection to the most common gaps that need to be bridged (**Figure 1**). One of the main outcomes of this first Workshop was the set-up of cross-domain thematic groups (also mentioned as WP5 Task Forces, **TFs**). The interaction between the working groups and the subdomains, ensuring the bidirectional information flow, is crucial for the Task Forces to achieve their goals and assist the ENVRIs in becoming more interoperable. The TF target audience includes the data centre staff, people who setup and maintain the RI services and the internal data management, as well as managerial positions for discussions on the strategy and goals of each RI. Therefore, as cross-domain working groups, the TFs engage representatives from all participating RIs, provide the required communication channels to enhance the collaboration among RIs from different subdomains, investigate the most crucial gaps and required solutions for building substantial foundations at a cluster level and coordinate the design of an ENVRI catalogue of services to integrate with EOSC services.

---

[16] Common Implementation Plan, by WP7 (based on the reported subdomain implementation plans): https://docs.google.com/spreadsheets/d/1JINwEC1gSLZmJlimSBMQg37t610kahiBvKj_ClosxWA/edit#gid=1964688825

[17] ENVRI Knowledge Base http://envri-fair.github.io/knowledge-base-ui/, still work in progress

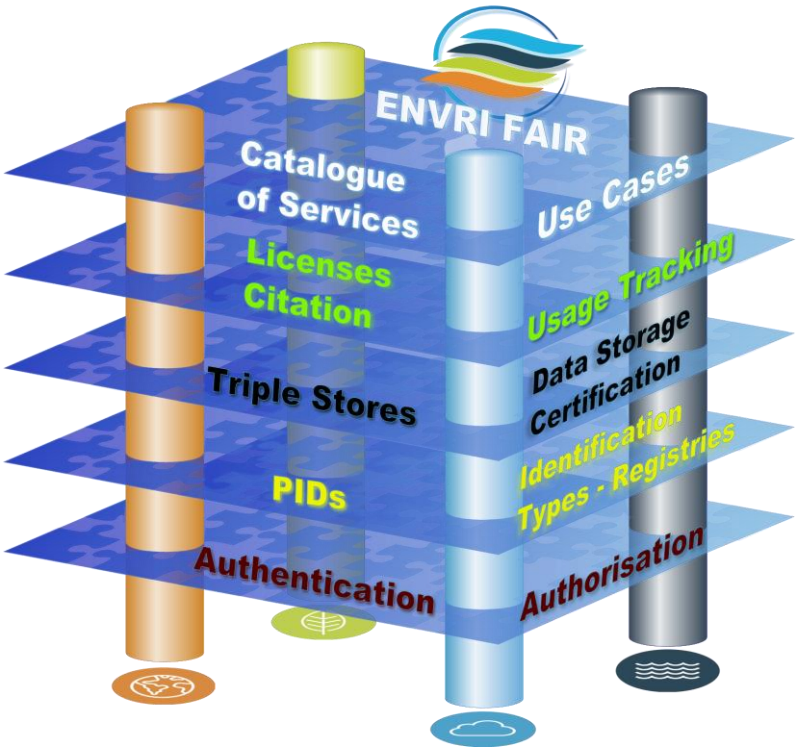

Figure 1. The newly introduced cross-domain Task Forces correspond to thematic groups which will focus on topics (mentioned in the illustration) of common interest, engaging representatives from all participating RIs, aiming to investigate the technical solutions which are required for the ENVRIs common strategy towards more FAIR data and services.

4.1 Designing the ENVRI Catalogue of Services

One of the main targets for the ENVRI community is the design of the ENVRI catalogue of services, which should be constructed in a way that also supports the integration into EOSC. The ENVRI service catalogue together with the Knowledge Base (KB), currently constructed by the ENVRI community, will need to result in a closely coordinated and mutually dependent set of tools by the end of the project. While the KB is focused on the semantic description of each RI and their FAIR data and services, aiming to resolve issues as e.g. finding an available catalogue (using search tools based on the KB or connected to the individual catalogues), the ENVRI catalogue of services will contain more practical technical solutions, concerning datasets, web services, Graphical User Interfaces (GUIs) or portals, software, computational services, equipment, workflows and others. Whether the research products, individual datasets etc. will also be included in the catalogue, is still under consideration.

4.1.1 A catalogue of services

To design the ENVRI service catalogue architecture, the community first needs to define the type of services which will be the assets of the ENVRI catalogue, before facing the challenge of finding the suitable schema to describe them. *Can for example sub-setting a big dataset be considered as a service?* In some cases this is indeed a computational service the RIs have made available to their end-users. Due to the heterogeneous landscape mapped within the cluster, it is crucial to converge to some commonly accepted and comprehended definitions for interoperable services. A practical solution is to combine use cases that include available services provided by the individual RIs, which will be integrated at cluster level. It is suggested to use the term “service” to mean:

a. the technical meaning of web services e.g. REST services etc., which are machine accessible,
b. from a user’s point of view, anything that provides functionalities, e.g. web portals which are accessible by humans.

To build a FAIR service, the ENVRI community approaches the catalogue architecture from both a human and a machine perspective. To be compliant with the FAIR principles, the ENVRI catalogue of services should combine some well-defined characteristics, e.g. use of PIDs, support the mapping of rich metadata standards, represent the relevant vocabularies, track licence information and provenance. Synchronizing different catalogues available at RIs in the environmental cluster, considering the high level of heterogeneity among them, is quite difficult. There are differences in the data formats and semantics that describe the datasets, as well as in the way of accessing the data and services.The way out here is to map the metadata of the available services. For example, the entry point where someone will get access to assets provided by the different scientific communities can be designed to combine the different types of services the communities make available, either those are web services or access points to individual data files etc. The metadata included in the catalogue will not necessarily describe datasets (findable at RI level), but instead describe services and/or resources that give access to the actual environmental data. A schema which allows rich metadata of discoverable services to be included will enable the ENVRI catalogue to be integrated in the EOSC and serve both humans and machines.

4.1.2 Interaction with EOSC

Ideally the ENVRI catalogue should be designed from the beginning in a way that suits the interaction with the EOSC, and at the same time serves the needs and goals of the ENVRIs and the interoperability of their services. Allowing EOSC to "harvest" the entry point of the ENVRI catalogue of services, an EOSC end-user will get access to the ENVRI assets through the ENVRI catalogue, which will serve as the interaction layer between the ENVRI community and EOSC. Direct interactions between EOSC and individual RIs (apart from the cluster level) are still an open question. For the ENVRI community, following the EOSC recommendations seems crucial. Some of these recommendations might be already clarified and others develop through the ongoing projects, which indicates a quite dynamic process. The EOSC projects have set up some minimum requirements for the service providers for the EOSC service catalogue and the marketplace, which can be the starting points for this interaction.

EOSC-hub, one of central projects in the EOSC landscape, has established and operates several of the EOSC central services, including the EOSC Portal and an Authentication and Authorisation system. In the last version of the "EOSC-hub Integration Handbook for Service Providers"[18], the minimum criteria of becoming a service provider in EOSC are defined. For a service to be onboarded in EOSC:

1. It should bring value to users and facilitate them to implement Open Science
2. It is either an online or a human service (web API, web services or training etc.), but not datasets or service artefacts (those should be deposited in data and software repositories registered in EOSC)
3. It is mature, reaching the appropriate Technology Readiness Level (TRL7) defined by the European Commission[19], and has been used by early adopter scientists ("System prototype demonstration in operational environment")
4. The service description template compulsory fields are filled.

As stated in the Integration Handbook, after successful validation, the service entry is publicly available and accessible in the EOSC Portal and Marketplace, and this corresponds to the minimum level of EOSC integration. If e.g. a service requires an application for access, the user requests will be sent to the service provider through the Marketplace. Other integration services provided by the EOSC-hub, aiming to simplify users' access or interaction, are also described in the EOSC-Hub Integration Handbook.

4.1.3 The initial approach

From an end-user's perspective, when meeting the entry point to the ENVRI catalogue it should be clear what access is provided. There are different scenarios. The ENVRI catalogue of services is not

[18] EOSC-hub Integration Handbook for Service Providers, May 2020, DOI: 10.5281/zenodo.3826907
[19] Technology Readiness Levels defined by the European Commission
https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

meant to duplicate the already existing RI catalogues, or include all metadata available at RI level. The catalogue can include information on the way the RIs give access to their assets, which can be portals (through URLs for humans) as well as application APIs described with the appropriate metadata. It will also provide access to the RI resources for both humans and machines.

The types of ENVRI services available often differ among the scientific communities. The resources, data, metadata, software and other digital objects are provided through web services. Their descriptive information needs to be mapped, and the metadata will be introduced into a catalogue. To achieve interoperability among the ENVRIs, ensuring rich metadata and implementing the necessary technologies is essential. The metadata can also be exported into e.g. the EOSC catalogue of services. The scientific communities need to first select their services and then provide the necessary description. Starting with some simple use cases, all RIs will contribute to building the common catalogue and deal with further complexities. E.g. EPOS provides a catalogue of waveform data which is accessible through web APIs. What will be reported in the ENVRI catalogue by the Solid Earth subdomain is a sufficiently mapped computational service which will give access to these datasets.

One of the first questions to be answered concerns the access to the ENVRI catalogue from an external service (e.g. through EOSC). A web API or a GUI gives access to this catalogue of information. The community has agreed to start with the web APIs, as those can provide solutions towards a federated environment. Supporting the machine-to-machine access to (and use of) the ENVRI services will enhance the interoperability within the cluster. The cross-domain working group set up to work on the design of the catalogue (Task Force 1, TF1) mapped the current technical status of each participating RI regarding the availability of APIs. At the moment there are a number of RIs that allow access to their resources (or sometimes access to datasets in other repositories) through a GUI. In many cases machine-readable solutions are also under development (as web APIs).

The processes of mapping the metadata and the use of metadata schemas are currently discussed in the project, and particularly in TF1 where the project partners are represented. To form the required integrating layer for the community, the use of an existing solution is recommended, instead of introducing a new metadata schema. During the ENVRIplus project some of the schemas were tested. CERIF that was used as metadata standard within EPOS is a good example of an existing solution that can be adopted, as it can convert from various metadata formats, map many resources and also support provenance. The metadata provided by the communities will be mapped in the main catalogue. *How is the metadata transferred and how is interoperability achieved? How will the community store and manage the metadata semantics?* Practical training will be required, to help the RIs adopt the suggested solutions, as there are many steps to be followed before the metadata reach the main catalogue: backend processes that are required to e.g. harvest the metadata properly from the repositories, have them transferred and converted to be included into the metadata schema, and eventually make them findable and accessible at the frontend services. Other issues, e.g. concerning security, should also be discussed further.

4.2 Improving FAIRness with Authentication and Authorisation services

According to the Authentication and Authorisation for Research and Collaboration (AARC) initiative[20], to manage an application someone needs to guarantee it is accessed only by people that have the right to use it. To achieve successfully this goal, two distinct processes need to be implemented:

a) the user needs to authenticate to a system with existing credentials, and
b) the system owner wants to check user's permission and give appropriate access

The AA services refer to the accessibility of the environmental data and services (F**A**IR). Within the environmental cluster, most of the participating RIs already use authentication and authorisation protocols. The objective is to harmonise the authentication and authorisation schemes across all ENVRIs and find an AAI (Authentication and Authorisation Infrastructure) prototype as a common solution for the cluster. So far it has been clarified that the ENVRI-FAIR project is not aiming at

[20] AARC Project https://aarc-project.eu

building a centralized ENVRI-FAIR identity provider. The goal for the cross-domain group which has been set up to work on the topic (Task Force 2, TF2) is to negotiate among the participating ENVRIs a protocol that allows RIs to federate. The end-users of the ENVRI services will be able to get authenticated (i.e. to log in) and access all shared resources. The permissions associated with resources and datasets from the environmental infrastructures are quite heterogeneous. Within a governance framework the RIs also need to consider what information will be shared. By e.g. agreeing on a "minimal level" of permissions to access other ENVRIs (so that the individual RIs can still manage the Authorisation rules for accessing their applications), an AAI federation can be implemented and tested within the ENVRI cluster, potentially to be extended to the services integrated with the EOSC services.

At first, the ENVRIs need to discuss the optimal mechanisms, the technical solutions and policies which will grant authorized access to services within the environmental cluster when an end-user is authenticated at one of the RIs. Currently, various authentication mechanisms are in use targeting a token-based authentication system using external Identity Providers (IdPs) and software such as Unity[21] to permit integration across authenticated identities from multiple IdPs. In such case, IdPs from RIs need to be registered to a Unity central node, and RIs need to check the token (OAuth2 proxy might be required, to be investigated).

The BluePrint Architecture[22] (BPA) by AARC2 provides a meaningful starting point and thus it is recommended as a reference for setting up individual AAI systems for the ENVRIs, to support the ENVRI-FAIR federation. AARC is an initiative which was first launched in May 2015 to address the increased need for federated access and the mechanisms in research and e-infrastructures for authentication and authorisation. In May 2017 the project entered its second phase to continue with the integration of a cross-discipline authentication and authorisation framework, using the already existing AAIs.

The AARC2 project suggests the use of the BluePrint Architecture which is a set of software building blocks. Following this, software architects can use tried and tested components to build customised solutions. **Figure 2** presents the last version of the BPA, which includes the user identity, the community attribute services which are related to the management of the information about the users (e.g. the community roles) which might be provided from the user identity layer, the access protocol translation, the authorisation and the end-services. The BPA solution introduces the concept of a “Proxy”, meaning that the community or the infrastructure runs an “Infrastructure Proxy” and all services connect only to that Proxy, considering that services within an infrastructure usually have common requirements that can be deployed at a central point**.** The BPA has already been used in other domains (e.g. the LIGO Collaboration) and there are partners in the environmental cluster which have participated in AARC pilot projects and can contribute to building a BPA.

[21] Unity, A solution for identity, federation and inter-federation management, https://www.unity-idm.eu
[22] AARC Blueprint Architecture https://aarc-project.eu/architecture/

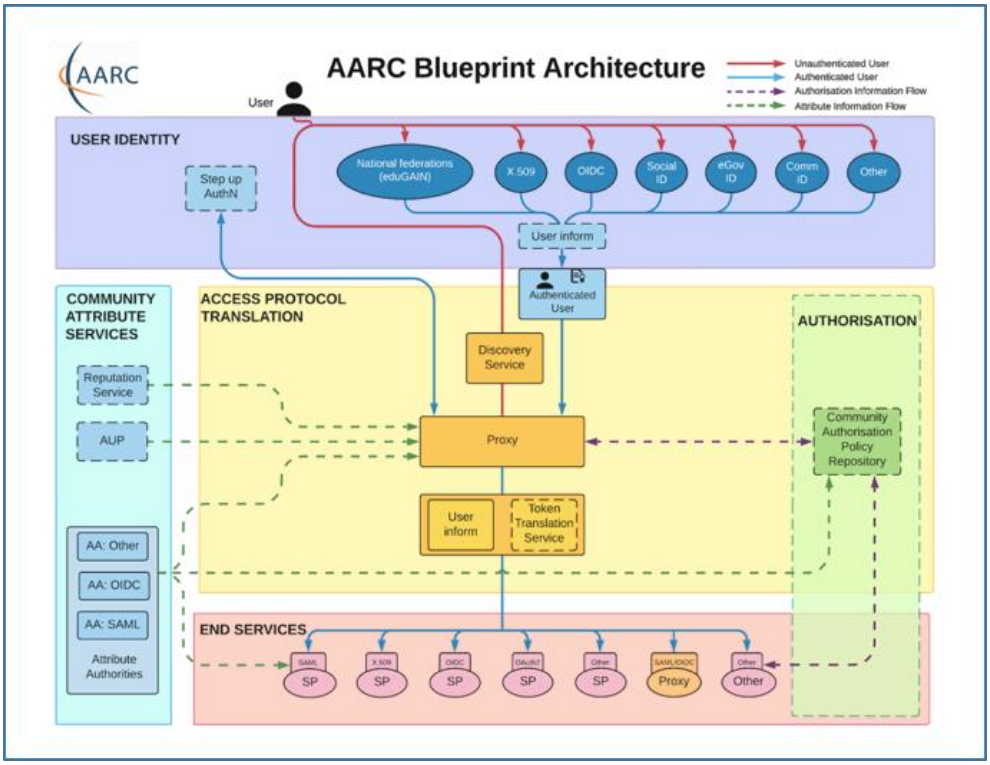

Figure 2. The last version of the AARC Blueprint Architecture (BPA). It includes the user identity, the community attribute services which are related to the management of the information about the users (e.g. the community roles) and might be provided from the user identity layer, the access protocol translation, the authorisation and the end-services.

4.3 The role of persistent identifiers

Globally unique and resolvable persistent digital identifiers (GUR-PIDs) form a core component of FAIR data repositories. Persistent identifiers can be applied to data, metadata, instruments, measurement sites, people, organisations, physical samples and any entity that can be given a digital representation – i.e. a Digital Object (DO). A GUR-PID not only serves as a means to assign a unique "identity tag" to any DO, but simultaneously provides a mechanism to direct a user to either the entity itself or to a "landing page" with relevant metadata.

By comprehensively and consistently implementing the use of persistent identifiers throughout their data processing and curation activities, ENVRIs are taking a big step towards making both their data and any services built on them FAIR. As an example, the metadata record of a data set (itself assigned a PID) can include ORCiD identifiers of all related people, persistent URLs pointing to Linked-Open-Data-definitions of all included variables, Handle PIDs of instruments used for the underlying observations, GitHub/Zenodo DOIs for the applied analysis software code, and DataCite DOIs for associated journal articles describing the measurement and quality assurance protocols.

It is important the ENVRIs take care to base their operations on GUR-PID systems that are backed by organisations and service providers that are trustable and operate in a sustainable way. For most applications, the basic functionalities (registration, resolution, maintenance of records) are sufficient, but as the needs and requirements of both RIs and user communities develop and change, it will be necessary to enter into a dialogue with relevant actors (e.g., PID providers, SMEs and other data clusters) to propose extensions to existing services, or even collaborate to develop new ones.

There are several initiatives and working groups active at European and Global level which provide their expertise on the PID applications and thus are followed by the ENVRI partners (e.g. RDA,

FREYA, FAIRsFAIR, EOSC working groups). In 2015, a survey (Atkinson et. al, 2016)[23] by ENVRIplus project asked participant RIs to supply information on their research data management (RDM) practices and related technology requirements. The survey revealed a variability of organisational maturity among the respondents, ranging from RIs that were just starting and didn't yet have detailed data management procedures in place, while other RIs were more advanced. As for identification and citation, a majority of RIs indicated that they had started to assign PIDs to final "publishable" data products, but only a few were applying identifiers to e.g. raw or intermediate-level data, or to non-data objects. The results of this survey are also available through the ENVRI Knowledge Base.

The ENVRIplus work package on Identification and Citation mapped out the landscape of existing PID services and applications, and based on the technology review also documented the identification best practices (Hellström et. al, 2017)[24] for the RIs, e.g. which data objects and entities should be assigned a PID, the importance of factors like the practical issues of granularity for data and the cost of assigning/maintaining PIDs etc., the need of assigning PIDs to the metadata of the data objects, PIDs for documents, people and organisations, instruments, physical samples, software, workflows. Examples of relevant applications at RI level have been given also in other ENVRIplus deliverables (Hellström et. al, 2019)[25], with information on how the ENVRIs inform their end-users regarding identifications, how the citation statistics can be used etc.

The newly formed cross-domain working group within the ENVRI community (Task Force 3, TF3) will build on this earlier work, revisiting current PID usage and identifying new challenges and solutions, in order to update best practices and policies. A new report can give more insights into the current data management plans, reveal possible similarities that might exist among the RI practices and use them to build the common solutions for attaching PIDs to data objects and ways to leverage them. By collecting use cases from (mature) RIs, and analyzing these to map out existing PID metadata models and strategies for assigning identifiers to their data, their preferred workflows for machine interaction etc. will provide valuable knowledge, especially for RIs that are in early stages of implementation. Documenting the end-users' needs will also be useful. Collaboration with the task force that focuses on data citation (see section 4.5) is also required.

The EOSC FAIR[26] and EOSC Architecture[27] working groups, where ENVRI-FAIR is represented, have put together a Persistent Identifier policy[28] which mainly targets the organisations/data centres/e-service providers who will provide PID services and will be connected to the EOSC landscape (the topics discussed in the PID Policy Document are listed in **Figure 3**). The ENVRI-FAIR TF3 will take this policy into account from the RI point of view, as it provides definitions, usage requirements and guidelines on how to build systems based on the available specific technologies and their functionality. ENVRI-FAIR is also participating in the formulation of a document outlining the architecture of PID services that will be required for supporting the EOSC. This report[29] will describe common PID systems, the landscape of actors and their interactions, and specify minimal requirements for e.g. sustainability, security, capacity and functionality.

---

[23] M. Atkinson, A. Hardisty, R. Filgueira, C. Alexandru, A. Vermeulen, K. Jeffery, T. Loubrieu, L. Candela, B. Magagna, P. Martin, Y. Chen and M. Hellström: A consistent characterisation of existing and planned RIs. ENVRIplus Deliverable 5.1, submitted on April 30, 2016. Available at http://www.envriplus.eu/wp-content/uploads/2016/06/A-consistent-characterisation-of-RIs.pdf

[24] M. Hellström, M. Lassi, A. Vermeulen, R. Huber, M. Stocker, F. Toussaint, M. Atkinson and M. Fiebig: A system design for data identifier and citation services for environmental RIs projects to prepare an ENVRIPLUS strategy to negotiate with external organisations. ENVRIplus Deliverable D6.1, submitted on January 31, 2017. Available at http://www.envriplus.eu/wp-content/uploads/2015/08/D6.1-A-system-design-for-data-identifier-and-citation-services-for-environmental-RIs.pdf

[25] M. Hellström, M. Johnsson, A. Vermeulen, D. Lear and M. Fiebig: Report on identification and citation service case studies. ENVRIplus Deliverable D6.3, submitted on February 28, 2019. Available at http://www.envriplus.eu/wp-content/uploads/2015/08/D6.3.pdf

[26] EOSC FAIR WG https://www.eoscsecretariat.eu/working-groups/fair-working-group

[27] EOSC Architecture WG https://www.eoscsecretariat.eu/working-groups/architecture-working-group

[28] EOSC Persistent Identifier Policy, see https://zenodo.org/record/3780423 (second draft version)

[29] U. Schwardmann et al., PID Architecture for the EOSC (in progress)

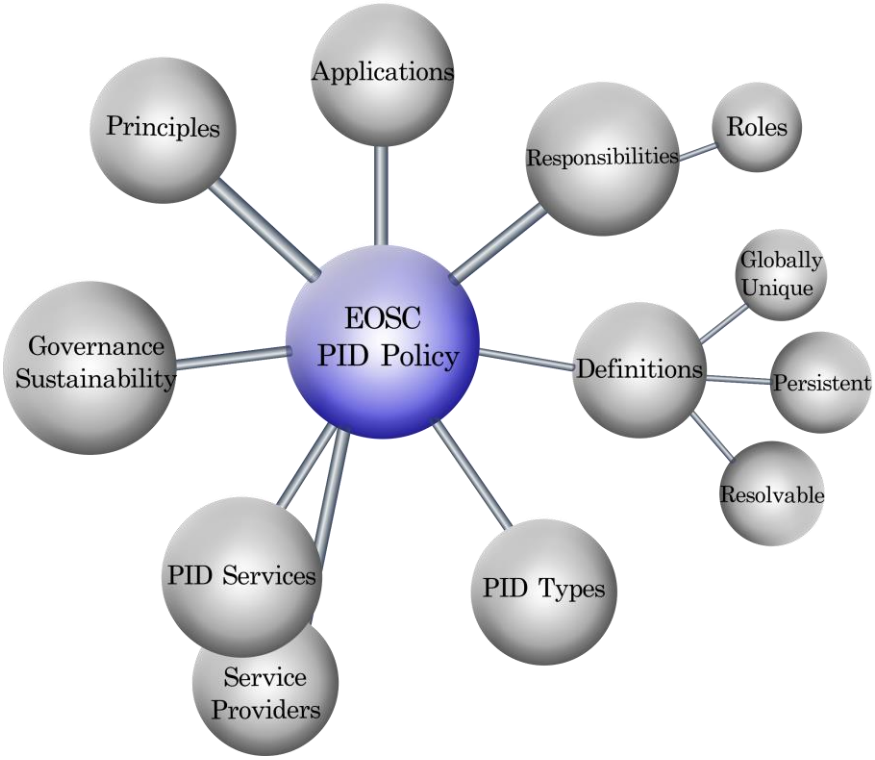

Figure 3. The topics discussed in the Persistent Identifier (PID) Policy for EOSC[30] (second draft version), authored by representatives of the EOSC FAIR WG and EOSC Architecture WG.

TF3 has already planned several activities based on the implementation plans and targets of the environmental subdomains. The already existing material related to PIDs from the ENVRIplus project deliverables and the various RDA working group outputs provide solid starting points for this work. The focus of attention will be on designs and solutions that support machine-actionability and workflows, while still enabling human interaction with FAIR digital objects. The community needs to take into account internal aspects, e.g. what happens at the PID registry level, if there is an optimal architecture which can deal with the fact that different PID systems are used (handles, persistent URLs etc) and will be able to successfully resolve any incoming PID. Performance and scalability are also important aspects for the ENVRIs, considering that e.g. for some RIs there are large amounts of data objects (as in daily observations) and thus they need systems that can deal with requests for many PIDs, update the registry metadata, resolve properly etc. Training will be required, on topics which are common priorities at cluster level, such as PIDs of instruments[31], the FAIR Digital Object (FDO) framework[32,33] (**Figure 4**), the use of PIDs to support provenance tracing etc.

[30] Second draft Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC), May 2020, DOI: 10.5281/zenodo.3780423
[31] RDA PIDs for instruments WG https://www.rd-alliance.org/groups/persistent-identification-instruments-wg
[32] GO FAIR FAIR Digital Object Framework https://www.go-fair.org/today/fair-digital-framework/
[33] GEDE FDO https://github.com/GEDE-RDA-Europe/GEDE/tree/master/FAIR%20Digital%20Objects

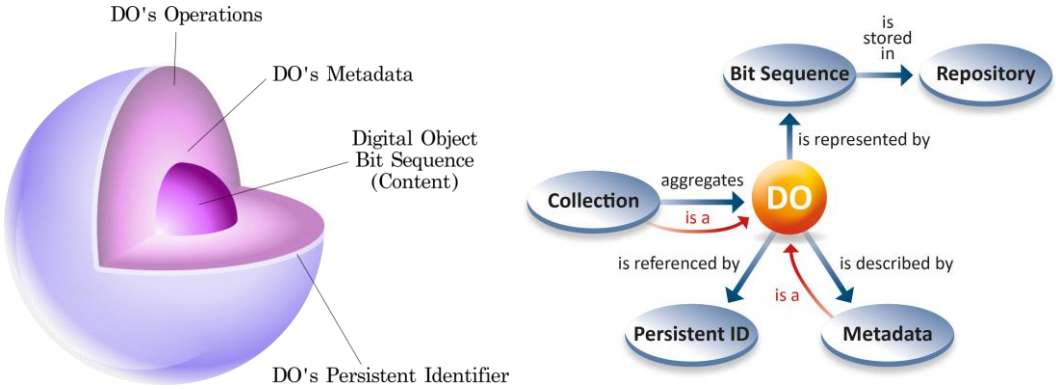

Figure 4. *(Left)* Graphical representation of the Digital Object (DO) concept and *(right)* representation of the Core Data Model architecture (RDA Data Foundation and Terminology[34] group; Wittenburg et al., 2019[35]).

4.4 Triple stores and data storage certifications

Triple stores constitute a popular database technology for linked graph data and were developed to manage (store, query, inference) the growing amount of Resource Description Framework (RDF) data. RDF data is a set of statements, each of them being a triple consisting of a subject (URI), a predicate (URI) and object (URI, Literal). They are part of the so-called semantic web technology stack. Databases for RDF data are examples of triple stores (e.g. Apache Jena/Fuseki, Virtuoso, Stardog, Allegro Graph, D2RQ). What is included in the database can be represented in various formats, depending on the human or machine requesting the content. There are some common languages used to express queries across the Web (ontology languages - RDFs, OWL, SPARQL query language). Triple stores are a relevant technology to FAIR data in particular because of the Principles' reference to languages for knowledge representation (Interoperability). In state-of-the-art information systems, the Web Ontology Language (OWL) is the standard language for knowledge representation. OWL ontologies are encoded as RDF. Hence, triple stores can be used to manage OWL ontologies as well as data that instantiate OWL ontologies. With their support for formal (i.e. machine readable) semantics and inference, semantic web technologies are also relevant to (semantic) findability of data as well as (machine) re-usability.

Data storage certification (World Data Systems[36], Data Seal of Approval - later replaced by the CoreTrustSeal[37], ISO etc.) is an important milestone for any RI, acting as a "trust label" for stakeholders depositing or consuming data. Getting a repository certified, however, is a quite complex process. According to CoreTrustSeal (CTS) requirements[38] for example, a repository which preserves data and provides ongoing access to it should also maintain all applicable licences covering data access, have a continuity plan, be compliant with disciplinary and ethical norms, have a clear system of governance and in general fulfil a series of requirements to guarantee the quality, integrity and authenticity of the data. Other requirements refer to the technical characteristics of the repositories (regarding hardware and software technologies), the necessary documentation of all procedures in data curation and the availability of all relevant information for the end-users. A certified repository is compliant with the FAIR principles through all stages of the data life cycle, protecting the data, products, services and end-users, while allowing them to discover data, properly cite the resources and have access to all required metadata to explicitly understand the data and services they have access to.

[34] RDA DFT Core Terms and Model http://hdl.handle.net/11304/5d760a3e-991d-11e5-9bb4- 2b0aad496318
[35] Wittenburg, P., Strawn, G., Mons, B., Bonino, L., Schultes, E., 2019. Digital Objects as Drivers towards Convergence in Data Infrastructures https://doi.org/10.23728/B2SHARE.B605D85809CA45679B110719B6C6CB11
[36] World Data Systems https://www.worlddatasystem.org
[37] Core Trust Seal Organisation https://www.coretrustseal.org/about/
[38] CTS Requirements for certified trustworthy repositories https://www.coretrustseal.org/why-certification/requirements/

The ENVRI-FAIR working group that is assigned to the investigation of triple stores and their role in the ENVRI community and the FAIR project (Task Force 4, TF4), first attempted to record the operational as well as the planned triple stores in the participating RIs, to map the ENVRI needs (for triple stores, certification schemes etc) and develop practical guidelines (e.g. for creating triples, linking datasets to RDF). TF4 documented the current status of the participating ENVRIs in experience reports which are available for all ENVRIs, giving information on the technologies and schemes used for triple stores and data storage certifications (e.g. by the end of 2020 some of the ENVRI repositories will apply for the CTS), the purpose they serve and other information. Together with the ENVRI-FAIR training WP, the community will also have the opportunity to follow webinars in the following months.

4.5 License, data citation and usage tracking

While one of the central ENVRI-FAIR WPs (WP4) focuses on the common FAIR policies for data and services provision which will be implemented by the participating RIs, the ENVRI community has recognized the need for a working group which will investigate the technical aspects of the licenses for data and metadata. Therefore, the Task Force 4 (TF4) will review how licenses are documented and how the license information is transferred within the existing metadata frameworks, considering e.g. that it is necessary to support both human- and machine-readable forms. The intention is to give the RIs recommendations on which metadata items should be used to document (in the metadata standards) information concerning the data licenses and policy (including the license of the metadata itself).

As for scientific publications, the use of data needs to be tracked by their identifier. The ENVRI community will monitor ongoing activities and collaborate when needed with other ongoing initiatives and indexing agencies for quantifying data. One of the main objectives for TF4 is to work on the data citation issues, having the valuable experience from earlier projects as a starting point. The community needs to investigate further the approach presented in the ENVRIplus project and evaluate the concept of double use of DOIs to support applications across all environmental subdomains, giving eventually the ENVRI end-users an easy way of citing datasets (or collections of data). The aim is to define a scheme for data identification, and thus citation, that is suitable for tracking the data use (down to the granularity of the individual PI, contributing organisation or framework). Having a description of the metadata items needed for data license, data policy, and metadata licenses, the ENVRIs will work towards the implementation of the TF4 recommendations for the documentation of license metadata.

4.6 Use cases in the ENVRI strategy

One of the long-term targets for the ENVRI community is the design of an ENVRI-hub, which is a concept of implementing a centralized access interface for the environmental data and services. To accomplish a virtual, federated machine-to-machine interface, the ENVRIs are currently building the foundations by working on the technical preconditions. There are different points of view to consider, such as the scientific use of the services, the information flow within the hub, the computational issues relevant to the service interaction in the hub and with EOSC, the architecture and technologies required to realize a hub for the environmental domain. An analysis of ENVRI-FAIR use cases will assist the gathering of usage requirements while designing the ENVRI hub. To achieve their goals the ENVRIs will use the ENVRIplus catalogue experience and build a rich metadata catalogue of ENVRI services, as described earlier in section 4.1. The appropriate technical solutions to be adopted at RI level are guided by the common requirements at cluster level, thus the ENVRI use cases will have to trace back to the specific requirements to contribute to the testing of the services functionality during a validation process at later steps of the project.

Cross-domain demonstration cases and services are requested as key outputs of the ENVRI-FAIR project to be integrated with the EOSC services. The concept and character of the domain demonstrators will be later harmonised with other ESFRI cluster projects. At the moment it is clear that having aligned services from all subdomains is crucial. A working group with representatives from all ENVRIs has been set up (Task Force 6, TF6) to coordinate the development and implementation of the cross-domain services for EOSC. The TF6 will operate within a broad and complex framework to specify selected science- and technology-based target subjects for the ENVRI-FAIR services, oversee the development activities in the contributing subdomains and guide the service prototype implementation.

5 Future Steps

The FAIRness assessment of the ENVRIs identified some long-term and overarching issues that will need close coordination between the subdomains in order (the issues) to be properly addressed. The subdomains planned their respective actions which will help them meet the FAIR principles. This process needs to be harmonised at cluster level. The goal is to select the common development targets and translate them into requirements for WP8-WP11, aiming to eventually prepare the output ENVRI catalogue of services. Having an up-to-date analysis of the gap(s) each individual RI needs to bridge to meet the FAIR requirements, the WP5 Task Forces will operate as thematic cross cutting RI/subdomain groups, aiming to provide support with the commonly identified issues. With help from WP7, there will be recommendations for common development targets for (metadata and data) services that will need to be implemented in the 4 environmental subdomains. This process will contribute to the design, development and implementation of the ENVRI catalogue of EOSC services and will also provide guidelines to test and validate the ENVRI services and formulate a strategic roadmap for future development. The different groups operate with timelines defined by their participants and include representatives of all the involved RIs. Where there are more complex or longer lists of targets, the groups will prioritise their actions accordingly. The Task Forces have planned activities for at least one year. Before the end of this year, a virtual meeting will be arranged (in the M23 of the project, i.e. November 2020), to give an opportunity to the working groups to exchange experience and report on their first outcomes. This meeting will serve as a preparatory phase for the second ENVRI week (which is planned to take place in the beginning of 2021, M26), where the cross-domain Task Forces will report their findings and recommendations to the ENVRI community. **Table 4** summarises the actions currently planned within the ENVRI cluster, in connection to WP5, for the following project.

Table 4. Summary of the actions the ENVRI cluster is currently planning, in the form of short- and long-term goals.

Task Forces Reports – November 2020	Goals at Cluster level (2020-2021)
Map of RI heterogeneity	New FAIRness assessment(s)
Convergence on common definitions and ENVRI catalogue schema	Recommendations from technical experts and working groups
Initial approach of the ENVRI catalogue design	Record/Report FAIRness improvement per RI
Understanding ENVRI catalogue entries through use cases	Record/Report Implementation progress per subdomain
Current and best PID practices	Active dialog with the EOSC
Experience from tutorials and reports on existing solutions	Guidelines for the validation of the ENVRI FAIR services
Implications (technical or other)	Gradually shape the ENVRI-hub framework

The next FAIRness assessment is expected to be conducted in early 2021. With the experience gained from the first evaluation process, the technical experts have now more tools which can help the RIs perform their next assessment, with automatic machine-accessible processes. The collaboration with international initiatives (e.g. the relevant GO FAIR working groups[39]) will contribute significantly.

Software validation and demonstration, quality control and interoperability within the ENVRI cluster and with respect to other systems (e.g. EOSC) is a long-term task which will continue until the end of the project. The partners have started elaborating, in synergy with WP7, potential criteria to be used as validation guidelines for testing the functionality of the developed RI specific services, in order to achieve technical harmonisation of the ENVRI services at cluster level and the interoperability with EOSC. Some of the actions that have been discussed include the development of potential use cases for interoperable demonstrators, and the use of Hackathons to develop interoperable service prototypes. The community will continue working on the design of the ENVRI-hub, considering different points of view (e.g. RI end-users, EOSC end-users, service providers) and with the constructive feedback from the interaction with the EOSC projects and working groups. How the ENVRI cluster will use the EOSC services (e.g. the EOSC catalogue, computational facilities, storage) is still under discussion. The ENVRI use cases and the work of TF6 will clarify the different perspectives of the problem.

[39] GO FAIR Convergence Matrix WG https://www.go-fair.org/today/FAIR-matrix/

6 Appendix 1. Glossary – Acronyms

AAI	Authentication and Authorisation Infrastructure
AARC	Authentication and Authorisation for Research Collaborations
ACTRIS	Aerosol Cloud and Trace gas InfraStructure network
AnaEE	Analysis Experimentation on Ecosystems
API	Application Programming Interface
BPA	AARC BluePrint Architecture
CERIF	Common European Research Infrastructure Format
CTS	Core Trust Seal
DANUBIUS-RI	International Centre for Advanced Studies on River-Sea Systems
DiSSCo	The Distributed System of Scientific Collections
DO	Digital Object
DOI	Digital Object Identifier
EISCAT	European Incoherent Scatter radar system
eLTER	Long-Term Ecosystem Research in Europe
EMSO	European Multidisciplinary Seafloor and water column Observatory
ENVRI	Environment research infrastructures (in ESFRI level or upcoming) as a community
ENVRIplus	An environmental RI cluster H2020 project
EOSC	European Open Science Cloud
EPOS	European Plate Observation System
ESFRI	European Strategy Forum on Research Infrastructures
Euro-Argo	Argo European Research Infrastructure Consortium
FAIR	Findable Accessible Interoperable Reusable
FAIRsFAIR	Fostering FAIR Data Practices in Europe
FDO	FAIR Digital Object
FREYA	Connected Open Identifiers for Discovery, Access and Use of Research Resources
FTP	File Transfer Protocol
GO FAIR	An international programme on FAIR implementation
GUI	Graphical User Interface
HTTP	HyperText Transfer Protocol
IAGOS	In-service Aircraft for a Global Observing System
ICOS	Integrated Carbon Observation System
ICS-C	EPOS Integrated Core Services Central hub
ICS-D	EPOS ICS-Distributed
IdPs	Identity Providers
IRI	Internationalised Resource Identifier
ISO	International Organisation for Standardisation
KB	Knowledge Base
LifeWatch	LifeWatch European Research Infrastructure Consortium
LIGO	Laser Interferometer Gravitational-Wave Observatory
NERC	Natural Environment Research Council
OAUTH	Open Authorisation (standard)
ORCID	Open Researcher and Contributor ID
OWL	Web Ontology Language
PID	Persistent Identifier
Postgres	Post Ingres – Interactive Graphics Retrieval System
RDA	Research Data Alliance
RDF	Resource Description Framework
REST	REpresentational State Transfer
RI	Research Infrastructure
RoP	Rules of Participation
SEADATANET	SeaDataNet pan-European infrastructure for marine data management
SIOS	Svalbard Integrated Arctic Earth Observing System
SPARQL	SPARQL Protocol and RDF Query Language
TCS	EPOS Thematic Core Services
TF	Task Force
TNA	EPOS Trans-National Access

TRL	Technology Readiness Level
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VRE	Virtual Research Environment
WG	Working Group
WP	Work Package
YAML	Yet Another Markup Language