D11.2
REPORT ON FAIRNESS IMPLEMENTATION ACTIVITIES IN THE BIODIVERSITY AND ECOSYSTEM SUBDOMAIN

Work Package	WP11
Lead partner	CREA
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	2020-12-31
Submission date	2021-04-12

Deliverable abstract

This Deliverable provides an overview of the ongoing data FAIRness implementation within the Biodiversity and Ecosystem subdomain, also in relation to the overall ENVRI FAIR project developments. Considering the central role assumed by the WP5 Task Forces within the project, the current compliance of the subdomain RIs with Task Forces recommendations has been analyzed, through a specific questionnaire administered to all subdomain RIs. In light of this survey results, single RI roadmaps are presented to fill the current gaps. A particular focus is finally reserved to the ongoing subdomain use cases, which are considered a key step in developing interoperability within the subdomain.

DELIVERY SLIP

	Name	Partner Organization	Date
Main Author	Flora De Natale	CREA (AnaEE)	
Contributing Authors	Monique Bossi	LifeWatch ERIC	
	André Chanzy	INRAe (AnaEE)	
	Dario De Nart	CREA (AnaEE)	
	Francesca De Pascalis	DANUBIUS	
	Laura Ferrighi	SIOS	
	Nicola Fiore	LifeWatch ERIC	
	P. Gasner	DANUBIUS	
	J. M. García Rodríguez	University of Seville	
	J.M. González-Aranda	LifeWatch ERIC	
	Sharif Islam	DiSSCo	
	Darius Ignatiuk	SIOS	
	Giovanni L'Abate	CREA (AnaEE)	
	Barbara Parisse	CREA (AnaEE)	
	J. Peterseil	eLTER	
	Chrisitan Pichot	INRAe (AnaEE)	
	Ilaria Rosati	CNR (LifeWatch ERIC)	
	A. J. Sáenz-Albanés	LifeWatch ERIC	
	F. M. Sánchez Cano	LifeWatch ERIC	
	Lucia Vaira	LifeWatch ERIC	
	Christoph Wohner	eLTER	
	Dario Papale	CMCC (ICOS)	
	Alberto Basset	LifeWatch ERIC	
Reviewer(s)	Giacinto Donvito	INFN	
	Margareta Hellström	ULUND	
Approver	Andreas Petzold	FZJ	2021-04-12

DELIVERY LOG

Issue	Date	Comment	Author
V 1.1	2021-01-14	Sent for review	Flora De Natale
V 2.0	2021-23-02	Sent for approval	Flora De Natale
V 2.0	2021-04-12	Submitted to EC Portal	Andreas Petzold

DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the Project Manager at manager@envri-fair.eu, as well as to the lead author (flora.denatale@crea.gov.it).

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.4471374.

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

D11.2 - REPORT ON FAIRNESS IMPLEMENTATION ACTIVITIES IN THE BIODIVERSITY AND ECOSYSTEM SUBDOMAIN

1 Introduction

The scientific community can greatly benefit by the implementation of the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) in the management of research data (Wisconsin et al., 2016).

Implementing the FAIR principles by the Environmental Research Infrastructures (ENVRIs) is the core activity of the ENVRI-FAIR project, which aims to connect the ENVRI cluster to the European Open Science Cloud (EOSC). This process is carried out in all four ENVRI subdomains: *Atmosphere, Marine, Solid Earth and Biodiversity and Ecosystem*, and at Research Infrastructure level.

After a first phase, which mainly consisted of an analysis of the overall FAIRness within the subdomains, the participant RIs are called to define a roadmap for implementing the FAIR principles within and across them, starting from the subdomain level.

This document focuses on the Biodiversity and Ecosystem subdomain and provides an overview on the current status and the main future steps of the FAIRness implementation.

The subdomain baseline

The first Deliverable of the ENVRI-FAIR Work Package 11 (WP11), about the "Biodiversity and ecosystem subdomain implementation short term plan" (Papale, 2020) provided a first analysis of the FAIRness of the Biodiversity and Ecosystem Research Infrastructures. The main strengths, weaknesses, priorities for each Research Infrastructure (RI) were investigated. The overall result is a very heterogeneous landscape, where RIs showed different maturity levels (mainly due to different starting years) concerning applying the FAIR principles, both in terms of policies and technologies adopted.

Moreover, during the first project year, a survey was carried out by the WP5 "Common requirements and testbed for (meta)data services, community standards and cataloguing" on the FAIRness maturity level of the different RIs. In this assessment, as reported in Deliverable 5.1 (D5.1) "Requirement analysis, technology review and gap analysis of environmental research infrastructures" (Magagna et al., 2020), the subdomain compliance with each FAIR principle has been analysed, allowing to highlight the main gaps to be filled in for each project subdomain.

A synthesis for the Biodiversity and Ecosystem subdomain is reported in table 1. The results are summarized in terms of FAIR compliant RIs' number, as reported in D5.1 Supplementary Material (in the protected project-internal Redmine environment).

Table 1. FAIRness maturity level within the subdomain. Data source: Deliverable 5.1 and D5.1 Supplementary Material

FAIR principles	Percentage of FAIR RIs
F1. (Meta)data are assigned a globally and persistent identifier	
– Identifier kind	29%
– Persistent Identifier (PID) provider	29%
F2. Data are described with rich metadata	
– Repositories with machine-readable metadata	29%
F3. (Meta)data clearly and explicitly include the identifier of the data they describe	
– Repositories with PIDs included in the metadata description	43%
F4. (Meta)data are registered or indexed in a searchable resource	
– Repositories which provide search on data	-
– Repositories registered in FAIR compliant registries	43%
A1. (Meta)data are retrievable by their identifier using a standardized protocol	
– A1.1 The protocol is open, free, and universally implementable	
o Repositories with FAIR access technologies	57%
– A1.2 The protocol allows for an authentication and authorization procedure, where necessary	
o Repositories with statement on access policy in metadata	57%
o Repositories with FAIR authentication and authorization protocol	43%
A2. Metadata are accessible, even when the data are no longer available	
– Longevity plan for metadata	0%
– Metadata openly available	0%
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	
– Machine-readable metadata exchange formats adopted	71%
I2. (Meta)data use vocabularies that follow FAIR principles	
– Metadata schemas adopted	57%
– Vocabularies adopted	71%
I3. (Meta)data include qualified references to other (meta)data	
– Repositories with schemas defined in open registries	43%
R1. Meta(data) are richly described with a plurality of accurate and relevant attributes	

E N V R I
FAIR

– R1.1. (Meta)data are released with a clear and accessible data usage licence		
o FAIR compliant licences		57%
– R1.2. (Meta)data are associated with detailed provenance		
o Machine-readable provenance information provided by RI		0%

Starting from this baseline, all the subdomain RIs have developed or planned new actions towards a better FAIRness implementation in their data management.

The FAIRness implementation plans presented by WP11 RIs except DiSSCo and eLTER (as they were not engaged in this project activity) in June 2020 show that in the current year most of the RIs, excluding ICOS which is already at a higher level of maturity, are mainly focused on two aspects of FAIRness:

1. data policies/Data Management Plan (DMP)
2. catalogues.

In addition, other activities carried out are related to the AAI (Authentication and Authorisation Infrastructure) protocol (AnaEE and Danubius), semantics and interoperability in general (AnaEE, ICOS ERIC, SIOS, LifeWatch ERIC), provenance (LifeWatch ERIC) and some first steps for case studies (ICOS ERIC).

Table 2. RIs participating in WP11 activities

Involved Research Infrastructure		Website
AnaEE	ANalysis And Experimentation on Ecosystems	www.anaee.eu
DANUBIUS-RI	International Center for Advanced Studies on River - Delta - Sea Systems	www.danubius-ri.eu
DiSSCo	Distributed System of Scientific Collections	www.dissco.eu
eLTER RI	European Long-Term Ecosystem, Critical Zone and Socio-ecological Research Infrastructure	www.lter-europe.net
ICOS ERIC	Integrated Carbon Observation System - European Research Infrastructure Consortium	www.icos-cp.eu
LifeWatch ERIC	e-Science and Technology European Infrastructure for Biodiversity and Ecosystem Research - European Research Infrastructure Consortium	www.lifewatch.eu
SIOS	Svalbard Integrated Arctic Earth Observing System	sios-svalbard.org

In such a variegated picture, the availability of shareable tools and the selection of cross-RIs use cases have been considered a good starting point to build a common background, to test the FAIRness and the interoperability among the RIs of the subdomain.

In particular, **four Use Cases** have been identified and shaped:
1. FAIR access to Soil Water Content (SWC) measurements across RIs,
2. Common system for Species Scientific Names (SSN) identification and reporting,
3. Common system for Site Documentation Interoperability (SDI),
4. Common layer of Core MetaData (CMD) in the Biodiversity and Ecosystems subdomain.

The document is organized as follows: in Chapter 2, we will summarise the ongoing discussion within the ENVRI-FAIR, so-called, WP5 Task Forces and the contributions provided by WP11 RIs. In Chapter 3 we will give a brief description of the current situation of WP11 RIs concerning critical topics highlighted by the task forces and in Chapter 4 we will present a roadmap for the next months of the ENVRI-FAIR project, outlining the planned activities and the expected results.

2 Task Forces Overview

In light of the gaps found within the different subdomains, resulting from the above mentioned WP5 FAIRness assessment (Magagna et al., 2020), **six cross-domain thematic groups** (also mentioned as WP5 Task Forces -TFs) have been set up. The TFs, with representatives from all sub-domains, aim to define and harmonise the necessary common solutions and ensure the coherent evolution of the four subdomains[1].

The task forces provide support in **designing and implementing the ENVRI catalogue to be integrated into European Open Science Cloud[2] (EOSC) (TF1) and recommendations for validating the ENVRI services in the framework of the ENVRI-hub (TF6)**. The topics discussed within the different task forces are related to the main requirements to make the ENVRI catalogue of services FAIR compliant: accessibility through a proper **Authentication and Authorization Infrastructure (AAI) protocol (TF2)**, findability ensured by the **use of PIDs (TF3)**, interoperability by mapping rich metadata standards, adopting a Resource Description Framework (RDF) schema for **knowledge representation and the relevant vocabularies (TF4), reusability, tracking licence information and provenance (TF5)** (Adamaki and Vermeulen, 2020).

The task forces' activity was launched during the 2020 ENVRI Week. The preliminary outcomes, progress, and future work have been presented during the ENVRI FAIR - WP5 Task Forces - Workshop on November 25th/26th, 2020. WP11 representatives participate in all six task forces to ensure a connection between the thematic groups and the Biodiversity and Ecosystem subdomain, as shown in table 2.

Table 3. WP11 participation in the WP5 task forces

WP5 - Task Forces	Participant WP11-RIs
TF1 ENVRI Catalogue	AnaEE, LifeWatch ERIC, ICOS ERIC (WP5-WP6)
TF2 AAI Implementation	AnaEE, SIOS
TF3 PIDs, identification types and registries	AnaEE, ICOS ERIC (WP6), LifeWatch ERIC (WP9)
TF4 Triple stores and data storage certification	AnaEE, ELTER-RI, ICOS ERIC (WP5-WP8), LifeWatch ERIC (WP9), SIOS
TF5 Licences, citation and usage tracking	AnaEE, ICOS ERIC, LifeWatch ERIC (WP9)
TF6 ENVRI-hub design and architecture	AnaEE, LifeWatch ERIC, ICOS ERIC (WP8)

---

[1] Atmosphere subdomain - WP8, Marine subdomain - WP9, Solid Earth - WP10 and Biodiversity and ecosystem subdomain implementation - WP11. Further information on WPs and WP-Leaders are available here:
https://envri.eu/work-packages/
[2] https://eosc-portal.eu/

ENVRI FAIR

2.1 TF1 - ENVRI Catalogue of services

Given the future connection with the EOSC, **TF1 aims at implementing a common metadata catalogue schema** with associated management services to make the resources of ENVRI RIs findable and accessible.

The resources mapped by the ENVRI catalogue have to meet two main requirements:
1. to be machine-readable assets,
2. to be described with rich metadata.

The resources can include datasets, web services, Graphical User Interfaces (GUIs) or portals, software, computational services, equipment, workflows and other. RIs participating in the TF1 have been required to propose some services as use cases for the ENVRI Catalogue, as representative of their own subdomain. A common schema of rich metadata is needed regarding metadata, which must represent information in a machine-readable and machine-understandable way. This schema is necessary to achieve the required interoperability within the ENVRI cluster, both from ENVRI-FAIR to the EOSC catalogue and vice versa. The TF1 has started to adopt the EPOS Data Catalogue Vocabulary Application Profile (EPOS_DCAT-AP) in RDF to interoperate with many other standards (including openAPI[3]), to produce an ENVRI catalogue from the metadata catalogues of different ENVRIs.

A tutorial will be prepared on mapping from heterogeneous metadata formats used in the RIs' catalogues to EPOS-DCAT-AP[4].

2.2 TF2 - AAI implementation

The recent discussion in ENVRI-FAIR highlighted the importance of establishing a chain of trust among RIs to provide federated access to ENVRI resources. For this purpose, the **TF2 activity focuses on defining a common protocol of Authentication and Authorization** to implement an AAI federation within the ENVRI cluster, potentially to be integrated into the EOSC. The RIs will agree a "minimal level" of permissions to access other ENVRIs.

The BluePrint Architecture (BPA, suggested by AARC[5]) has been recommended as a reference for setting up individual AAI systems to support the ENVRI-FAIR federation. As regards authentication, a token-based authentication system using external Identity providers (IdPs) is recommended. IdPs from RIs need to be registered using software such as Unity[6] that enables their integration.

The authorization procedure discussion is still ongoing because of the high heterogeneity in the user authorization schemas adopted so far by the ENVRIs and the low level of existing user federation across them. Over the following months, the discussion will address how to harmonize its recommendations with the TF4 ones and make them compliant with the policies defined by WP4 "Common FAIR Policies" and aligned with General Data Protection Regulation (GDPR)[7].

---

[3] See https://github.com/OAI/OpenAPI-Specification
[4] Documentation and materials on this standard are available at https://github.com/epos-eu/EPOS-DCAT-AP
[5] https://aarc-project.eu/architecture/
[6] https://www.unity-idm.eu/
[7] Please visit https://gdpr-info.eu for information on the GDPR - General Data Protection Regulation

2.3 TF3 -PIDs, identification types and registries

Persistent identifiers (PIDs) are key attributes to ensure findability of ENVRI resources. The metadata records of a data set (itself assigned a PID) can include:

- Open Researcher and Contributor ID (ORCID) identifiers of all related people,
- persistent URLs pointing to Linked-Open-Data-definitions of all included variables,
- Handle[8] PIDs of instruments used for the underlying observations,
- GitHub/Zenodo[9] Digital Object Identifiers (DOIs) for the applied analysis software code,
- DataCite[10] DOIs for associated journal articles describing the measurement and quality assurance protocols.

Trustable PIDs providers, compliant with the EOSC **Persistent Identifier Policies[11] have to be chosen by the ENVRIs, also referring to the guidelines provided for data citation by the TF5**. Sustainability, performance, granularity, and scalability are the basis for an optimal architecture dealing with the different PID systems used and supporting machine-actionability and workflows.

2.4 TF4 -Triple stores and data storage certification

Triple stores for knowledge representation play a key role in interoperability. The current standard language for knowledge representation is the Web Ontology Language (OWL), where the OWL ontologies are encoded as RDF. **The TF4 aims to map the ENVRI needs related to the triple stores and certification schemes and develop practical guidelines** (e.g., for creating triples, linking datasets to RDF). A triple store hosting RDF data from the site description use case was set up within the framework of TF4.

2.5 TF5 -Licences, citation and usage tracking

The TF5 aims to provide recommendations for documenting licences and data policy within the metadata, indicating which metadata items should be used. The main results regarding licences for metadata and data are a compiled list of metadata items needed to document licence, including legal viewpoint; determined mapping and gaps compared to existing metadata standards. As regards data citation, a double use of DOIs is recommended:

1. primary identifiers (DOIs), for all data products with granularity homogeneous and fixed per repository, are a basis for data use tracking. Granularity determines properties possible to resolve in data use tracking (platform, principal investigator, funding agency, framework, etc.).
2. data collection identifiers (DOIs) to identify a collection of data resources, following the Research Data Alliance[12] (RDA) Recommendation on Research Data Collections.

Besides, a scheme will be defined for identifying, citing data and for tracking its use.

---

[8] https://www.handle.net/
[9] https://github.com/, https://zenodo.org/
[10] https://datacite.org/
[11] https://zenodo.org/record/3780423
[12] https://www.rd-alliance.org/

2.6 TF6 -ENVRI-hub design and architecture

The TF6 aims to develop the architecture and the functionalities of the ENVRI-hub. The latter will be based on three main components:
1. the ENVRI Knowledge Base (human interface),
2. the ENVRI Catalogue of Services (machine-actionable interface),
3. the cross-domain demonstrator services and use cases for testing interoperability both at ENVRI cluster and EOSC level.

Basing on task forces inputs (mainly TF1 and TF4), the agreed ENVRI-hub architecture will be decentralized at RI/Subdomain level and will adopt the Data Catalogue Vocabulary (DCAT) standard as harmonized metadata schema and exchange protocol and SPARQL Protocol and RDF Query Language (SPARQL) endpoints as a search tool. The development of hub functionalities will be driven by the Knowledge Base and the use cases' implementation. The TF6 coordinates the design and implementation of use cases to be considered as executable demonstrators (e.g., via Jupyter notebooks). It aims to test legal, organizational, semantic and technical interoperability among the ENVRI subdomains.

Table 4. Task forces contact persons

WP5 - Task Forces	Leaders
TF1. ENVRI Catalogue	Daniele Bailo (Italian National Institute of Geophysics and Volcanology)
TF2. AAI Implementation	Daniele Bailo (Italian National Institute of Geophysics and Volcanology)
TF3. PIDs, identification types and registries	Margareta Hellström (Lund University)
TF4. Triple stores and data storage certification	Markus Stocker (Leibniz Information Centre for Science and Technology)
TF5. Licences, citation and usage tracking	Markus Fiebig (Norwegian Institute for Air Research)
TF6. ENVRI-hub design and architecture	Andreas Petzold (Jülich Research Centre)

3 Current state of the art

The ongoing discussion within ENVRI task forces has highlighted a pressing need for precise specifications.

We identify four critical aspects that at this stage of the ENVRI-FAIR project have to be surveyed to provide meaningful insights to guide the development of WP11 RIs in the context of the ENVRI FAIR main goals:

1. **Data catalogue federation readiness:** a key goal in improving data and resource access is the federation of catalogues across different RIs. However, to achieve this goal, all RIs must expose their catalogues on the Web and adopt a shared set of technologies outlined by the ENVRI community to access their data products and services.

2. **User federation readiness**: as outlined by the ENVRI community, user access federation across different RIs' systems is a substantial step towards better data accessibility. To this goal, ENVRI RIs must provide proven Provider systems that support user identity federation as a mandatory requirement, and user authorization federation as an optional one.

3. **Technologies and certifications**: aside from the specifications concerning the creation of an ENVRI catalogue and user community, the ENVRI-FAIR project has produced other practical recommendations regarding various aspects of data management. We are hereby assessing the compliance of the RIs involved in WP11 concerning technologies, standards, and certifications suggested by the ENVRI community so far.

4. **Policies and licences**: the landscape of open data is populated with a wealth of licences ranging from extremely liberal to viral ones that propagate themselves to every derivative product. Several of these licences are not compatible with each other, and a poor choice of licences can prevent data and services usage. Therefore, it is essential to assess the compatibility among the licences chosen by the various RIs to allow for data reuse and the development of cross-domain, multidisciplinary datasets, and to attract new stakeholders.

3.1 FAIRness implementation Survey

A specific survey has been carried out to collect information on technical details relevant for assessing how much the WP11 RIs are ready to implement the recommendations provided by the WP5 Task Forces and all the other requirements and best practices indicated so far by the ENVRI-FAIR project.

3.1.1 The questionnaire

The survey has been based on an online form[13], composed of four sections, which deal with the above mentioned four critical aspects for the FAIRness implementation. Overall, 32 questions have been proposed to WP11 RIs; most of them are rather technical and close-ended.

The questionnaire is organized into four sections:

---

[13] The online survey is available here: https://cutt.ly/OhNKUgB

ENVRI
FAIR

1. **Data catalogue federation readiness**: following the insights provided by Task Force 1, the WP11 RIs' situation was surveyed concerning data catalogues and their readiness to federate such data catalogues into the ENVRI catalogue.
2. **User federation readiness**: following the insights provided by Task Force 2, the WP11 RIs' current status was assessed for Authentication and Authorization Infrastructure (AAI) technologies and their readiness to enter an ENVRI user federation.
3. **Technologies and Certifications**: following the insights provided by Task Force 3 and 4, WP11 RIs' compliance was assessed to what appear to be the requirements emerging from the ongoing discussion.
4. **Policies and Licences**: following the insights provided by the ongoing discussion in the ENVRI community on data management policies and data licensing, the current WP11 RIs' situation was investigated concerning the operational and legal framework in which data is shared and published.

RIs have been required to provide a unique answer. In case of multiple solutions, the most representative one has been included in the analysis. In the remainder of this Chapter, we will present the main outcomes and findings of the survey.

3.1.2 Data catalogue federation readiness

In general, WP11 RIs are not ready for ENVRI catalogue federation, with the notable exception of LifeWatch ERIC that appears to be fully compliant with the requirements expressed so far by TF1.

Table 5. Overview of Data catalogue federation readiness

	Catalogue Online	Dataset harvesting support	DCAT metadata coverage	Programmatic data access with APIs
AnaEE	No	Yes	None	Under Development
DANUBIUS-RI	No	To Be Designed	To Be Assessed	Partial
DiSSCo	No	To Be Designed	To Be Assessed	Under Development
eLTER RI	Yes	Yes	None	Complete
ICOS ERIC	Yes	Yes	None	Complete
LifeWatch ERIC	Yes	Yes	Complete	Complete
SIOS	Under Development	Yes	None	Complete

The most significant issues highlighted by the survey are:
- the adoption of the DCAT vocabulary to provide a common presentation layer shared across all infrastructures,
- the online availability of catalogues, which is now partial because four out of seven RIs are either in the development or in the design phase of their catalogue systems.

Regarding DCAT metadata, it should be noted that the questionnaire investigated if these metadata are currently included in the data catalogue. These results are not exhaustive about the semantic developments ongoing within the RIs, but rather a snapshot of the current readiness for catalogue federation purposes. However, the situation is encouraging: WP11 RIs appear to have ongoing and planned activities to close the existing gaps. They all agree on supporting dataset harvesting (i.e., exposing endpoints to allow metadata ingestion from third parties) and offering wide Application Programming Interface (API) access to data.

All RIs opted for Open Source solutions for catalogue management, with GeoNetwork[14] being the most widespread product, except for DANUBIUS-RI which has not decided yet which solution to adopt.

---

[14] https://geonetwork-opensource.org/

3.1.3 User federation readiness

The most notable point highlighted by the survey is that four RIs out of seven do not use an Identity Provider to authenticate their users. This draws a clear line between RIs who can be interested in an ENVRI user federation (AnaEE, LifeWatch, and ICOS) and those that are not (DiSSCo, DANUBIUS, SIOS, and eLTER). For the three RIs that manage their users' identities, compatibility with the main AAI protocols involved in user federation appears to be a commodity. Hence, we can state that these RIs are ready for user federation with the rest of the ENVRI community.

Table 6. Overview of user federation readiness

	Identity Provider	OpenID connect [15]	SAML 2.0	OAUTH [16] tokens	multi-factor authentication
AnaEE	Yes	Yes	Yes	Yes	Yes
DANUBIUS-RI	No	Not Applicable	Not Applicable	Not Applicable	Not Applicable
DiSSCo	No	Not Applicable	Not Applicable	Not Applicable	Not Applicable
eLTER RI	No	Not Applicable	Not Applicable	Not Applicable	Not Applicable
ICOS ERIC	Yes	Yes	Yes	Yes	No
LifeWatch ERIC	Yes	Yes	Yes	Yes	Yes
SIOS	No	Not Applicable	Not Applicable	Not Applicable	Not Applicable

The survey has highlighted how different WP11 RIs have opted for other technical solutions to manage their AAI, including commercial products, Identity as a Service providers, and bespoke systems meant to fit their specific requirements.

Furthermore, the three RIs that use an Identity provider do not have a shared authorization scheme, with AnaEE still defining its authorization policies, LifeWatch ERIC using a stereotype-based system (hence grouping users according to predefined privilege levels), and ICOS managing permissions on an individual basis.

---

[15] Open ID (https://openid.net/)
[16] Open authorization protocol (https://oauth.net/)

ENVRI
FAIR

3.1.4 Technologies and Certifications

In this section, the compliance to SPARQL endpoints, PIDs, and repository has been investigated.

Out of seven RIs involved, only two (ICOS and AnaEE) appear to have a public SPARQL endpoint and a complete PID coverage for its data resources. In contrast, no RI has a data repository certification, hence at the time of writing no WP11 RI has a trusted repository among its services.

Table 7. Overview of Technologies and Certifications readiness

	SPARQL endpoint	PID coverage	Repository certification
AnaEE	Yes	Agreement Pending	Interested
DANUBIUS-RI	No	None	Not Interested
DiSSCo	No	None	Not Interested
eLTER RI	No	Partial	Not Interested
ICOS ERIC	Yes	Complete	Interested
LifeWatch ERIC	No	Partial	Not Interested
SIOS	No	Partial	Not Interested

All WP11 RIs agree on using DOIs as PIDs (except for AnaEE, all of them already have an agreement with a DOI provider). The RIs that have repositories are both oriented on the CoreTrustSeal[17] certification. Hence, we can state that all WP11 RIs have received the ENVRI recommendations. It is also noteworthy how most RIs are not interested in maintaining repositories. The only two of them who have their repositories are interested in obtaining the trusted repository certification.

3.1.5 Policies and Licences

Our survey highlighted how the WP11 RIs are in the process of regulating their data management procedures, and, except for DiSSCo, they appear to have a clear picture of their licensing framework.

---

[17] https://www.coretrustseal.org/

Table 8. Overview of Policies and Licences readiness

	DMP status	Suggested Data licences	Machine readability of licences	Data as a service
AnaEE	Draft	A Set Of Licences	Partial	No
DANUBIUS-RI	Under Definition	A Set Of Licences	Partial	No
DiSSCo	Draft	A Set Of Licences	To Be Assessed	To Be Planned
eLTER RI	Under Definition	Any Licence	Partial	Yes
ICOS ERIC	Final	Single Licence	Complete	Yes
LifeWatch ERIC	Under Definition	A Set Of Licences	Partial	No
SIOS	Final	A Set Of Licences	Partial	No

The overall situation appears to be encouraging with four out of seven RIs having defined a complete DMP (either final or still open to revisions), and the remaining committed to completing their current drafts. As for the licensing policies, the most widespread choice is to adopt a restricted set of eligible licences, with the notable exception of eLTER. Currently, eLTER does not impose any specific licence for data shared from the national networks; it only provides a recommendation to apply an open licence (e.g., CC-BY-NC 4.0 International[18]). Also, ICOS has a single eligible licence. All RIs, except for DiSSCo, can account for the machine readability of their eligible licences. Finally, WP11 RIs in general right now are not publishing data as a Service, except for eLTER and ICOS.

---

[18] https://creativecommons.org/licenses/by-nc/4.0/

4 Roadmap

The FAIR implementation roadmaps presented by the WP11 RIs as Project internal Milestones in June 2020 focused on the compliance with all FAIR principles. These plans need to be assessed and better addressed considering the current Task Forces' recommendations. Now the focus is on the **interoperability** within the subdomain and the ENVRI cluster. The aim is to evaluate how much the planned actions are in line with the Task Forces' common path and if it is needed to integrate the roadmaps with other targeted activities. This chapter includes two different sections.

The first section refers to the discussion of the general RI roadmaps, in the light of the four critical aspects considered for the status description (see Chapter 3), as summarised in figure 1, notably:

1. ongoing and planned implementations (in terms of actions and times) related to the RI catalogues (management system, harvesting, metadata standards, APIs),
2. ongoing and planned implementations (in terms of actions and times) related to the RI Authentication and Authorization protocol),
3. ongoing and planned implementations (in terms of actions and times) about semantic metadata, access, PIDs, repositories' certification),
4. ongoing and planned implementations (in terms of actions and times) in relation to DMP, licensing policy, licences adopted).

For each critical aspect, every RI provided an overview table including time (yearly quarters: Q1-Q4), actions, and the corresponding FAIRness level by indicating with an "X" which FAIR principle(s) is(are) covered with each specific action.

The second section concerns the implementation plan related to the User-oriented cross-RIs demonstration cases, where the joint activities are described, by highlighting the role of the different subdomain RIs.

The time horizon of the roadmap is the end of 2021.

Figure 1. Visual representation of the current sub-domain FAIRness status

		AnaEE	DANUB.	DiSSCo	eLTER	ICOS	LW	SIOS
Data Catalogue federation	Catalogue Online	No	No	No	Yes	Yes	Yes	UD
	Dataset harvesting support	Yes	TBD	TBD	Yes	Yes	Yes	Yes
	DCAT metadata coverage	None	TBA	TBA	None	None	Complete	None
	Programmatic data access with APIs	UD	Partial	UD	Complete	Complete	Complete	Complete
User federation	Identity Provider, Open Id connect, SAML 2.0, OAUTH tokens, multifactor authentication	Yes	No	No	No	Yes	Yes	No
Tech.gies and Certifications	SparQL endpoint	Yes	No	No	No	Yes	No	No
	PID coverage	Agreem. pending	None	None	Partial	Complete	Partial	Partial
	Repository certification	Int	Not Int	Not Int	Not Int	Int	Not Int	Not Int
Policies and Licenses	DMP status	Draft	UD	Draft	UD	Final	UD	Final
	Suggested Data licenses	A set of licenses	A set of licenses	A set of licenses	Any license	Single license	A set of licenses	A set of licenses
	Machine readability of licenses	Partial	Partial	TBA	Partial	Complete	Partial	Partial
	Data as a service	No	No	To be planned	Yes	Yes	No	No

4.1 General Roadmap

4.1.1 AnaEE-RI

Data Catalogue federation

A preliminary design of the AnaEE Data Discovery Portal (based on the Comprehensive Knowledge Archive Network -Ckan) has been sketched, and a first version will be delivered by June 2021. A beta release of a public access API portal has been developed, able to host a catalogue of APIs and to allow its users to request and manage APIs' access keys[19].

A survey on the data management conducted by the AnaEE platforms has been shaped. It will carried out in the framework of the SoftwarE for InfraStructure adMinistration (SEISM) application devoted to the AnaEE platforms, which the AnaEE Central HUB has developed.

Table 9. AnaEE timeline for Data Catalogue federation

Timeline	Action	Sub-Action	F	A	I	R
2019 (or before)	AnaEE platforms' catalog	Questionnaire drafting for Research Platforms	x			
2020-Q1	API portal	β-release	x	x	x	
2020-Q2	AnaEE platforms' catalog	β-release	x	x		
2020-Q4	API Portal	Training for resource publishers	x	x	x	x
2020-Q4	API Portal	Training for users	x	x	x	x
2021-Q1	API Portal	Full release	x	x	x	x
2021-Q2	Data Catalog	Metadata harvest (starting)	x			

---

[19] https://developer.anaee.eu/

User federation

AnaEE adopts an Identity Provider based on Active Directory technology to federate access to its services. AnaEE users, for instance, can already register to the AnaEE API portal with their institution's account. Although the federation of AnaEE platforms and centres is not complete yet, the existing solution can fit federation with any third-party identity provider, including the ENVRIs. The AnaEE Data and Modelling Centre manage the identity provider.

Table 10. AnaEE timeline for user federation

Timeline	Action	Sub-Action	F	A	I	R
2021-Q1	AAI	federated identity with LDAP[20] protocol		x		
2021-Q1	AAI	federated authorization with OAUTH2[21] protocol		x		

Technologies and Certifications

The adoption of the DOI system for assigning PIDs to the RI datasets has been decided. Workflows for data publication and mining of DOIs will be created.

Regarding semantics, the main activities are related to the implementation of semantic pipelines (developed within the ENVRI-plus[22] portfolio) for the interoperability of data/metadata. A deployment test of these pipelines has been carried out on an external soil database, as a first step for their deployment by the AnaEE Data and Modelling Centre. Additional developments are related to implementing a triple store SPARQL endpoint, which has been firstly fed by information about the AnaEE sites' description (WP11 use case) generated by the semantic pipelines. This work is connected to the WP5 Task Force 4 (Triple store and data storage certification) and the ENVRI-FAIR knowledge base[23] activity (WP7 T7.2). The endpoint is publicly accessible, and currently exposes a preliminary set of triples describing AnaEE sites.

A specific activity has been undertaken to integrate this information in the metadata sets concerning provenance.

---

[20] Lightweight Directory Access Protocol (https://ldap.com/)
[21] Industrial standard-protocol for authorization (https://oauth.net/2/)
[22] http://www.envriplus.eu/
[23] The Knowledge Base provides Knowledge-as-a-Service for the RI development communities to document the development and operation of RI services and to address engineering problems

Table 11. AnaEE timeline for Technologies and Certifications

Timeline	Action	Sub-Action	F	A	I	R
2021-Q4	Semantic Pipelines	Deployment	x		x	x
2021-Q4	Data Provenance description	Identification and integration of provenance elements for the enrichment of metadata type information in the AnaEE Information Systems, metadata records and datasets	x			x
2021-Q4	PID implementation	DOI adoption	x		x	
2022-on	Vocabulary adoption				x	
2020-Q4	SPARQL endpoint	developed for the AnaEE platforms, as an "experience report" in the framework of Task Force 4	x	x	x	

Policies and licences

A first version of the Data Management Plan has been completed, which also defines the AnaEE data policy. The AnaEE assembly has approved such Data and Management Plan of members. It can be considered the ruling document for the near future regarding data management, data quality, and data licences. AnaEE will support its adhering bodies in creating better data sets, providing them with recommendations, automatic tools for data curation, and expert advice in a peer-review fashion. The aim is to help platforms maximizing the impact of their data on research and business alike. On the licensing side, as stated in the DMP, AnaEE strongly supports Open Access publication in all of its forms and deems acceptable all the major open data licences. In general, AnaEE encourages using the most liberal licences, which are CC-0, CC-PDM, and ODC-PDDL[24].

Table 12. AnaEE timeline for Policies and licences

Timeline	Action	Sub-Action	F	A	I	R
2019 (or before)	Licences	Adoption of CC-BY				x
2020-Q2	DMP and Data Policies	1st version		x		x

4.1.2 Danubius-RI

Data Catalogue federation

DANUBIUS-RI Preparatory Phase project ended in December 2019, and the infrastructure started its implementation phase without a formal project yet. During 2020, a proto data catalogue was planned; therefore, a questionnaire, for all the RI components, was developed to collect the data. The data typologies that DANUBIUS-RI Data Centre will ingest will be heterogeneous, so different approaches would be possible.

Table 13. DANUBIUS RI timeline for Data Catalogue federation

Timeline	Action	Sub-Action	F	A	I	R
2020-Q3	Data Catalogue	Creation of a questionnaire to investigate data availability for populating DANUBIUS-RI FAIR data catalogue	x			
2021-Q1	Data Catalogue	Start filling questionnaire for populating the DANUBIUS-RI FAIR data catalogue	x			
2021-Q1	Data Catalogue	Identification of use cases and metadata harvest: Internal check between the FAIR Tech team and the DANUBIUS-RI data producers, to identify needed tools for data handling in a FAIR perspective. Info to fill DMP	x			
2022-on	Data Catalogue	prototype of data catalogue	x			

User federation

For what concerns AAI activities, DANUBIUS-RI evaluates to connect to the EDUcation Global Authentication Infrastructure[25] (eduGAIN) federated authentication service using an LDAP server as backend and OAUTH2 or OpenID connect protocols.

---

[25] https://edugain.org/

Table 14. DANUBIUS RI timeline for user federation

Timeline	Action	Sub-Action	F	A	I	R
2020-Q4	AAI (Authentication, Authorization, and Infrastructure).	Development of the RI central data storage with authentication protocol	x	x		

Technologies and Certifications

The actions of DANUBIUS-RI in terms of metadata, access, PIDs, repositories certifications are strictly connected to the first outcomes of the survey on the proto-catalogue and to the possibility of implementing solutions already accepted at ENVRI level.

Table 15. DANUBIUS RI timeline for Technologies and Certifications

Timeline	Action	Sub-Action	F	A	I	R
2021-Q2	FAIR Training	Organization of DANUBIUS-RI FAIR training sessions – training of trainers				x
2021-Q2	Prototype service for "Soil Water Content"	Data access across RIs version 1	x	x	x	
2022-on	PID implementation	Test on use cases for PID	x	x		
2022-on	Data curation: Quality Control (QC), data standardization	Standardized and actionable procedures for QC			x	x

Policies and licences

DANUBIUS-RI already prepared a version of DMP that will be revised during 2021 depending on the other activities' outcomes.

Table 16. DANUBIUS RI timeline for Policies and licences

Timeline	Action	Sub-Action	F	A	I	R
2020-Q3	DMP and data policies	State of the art Management Plan within ENVRI-FAIR and outside: – best practices (e.g., US Geological Survey, …) – to define the first list of needs		x		x
2021-Q2	DMP and data policies	Minimum conditions of the Data Management Plan - first draft		x		
2022-on	DMP and data policies	Minimum conditions of the Data Management Plan - final draft		x		x

4.1.3 DiSSCo

Disclaimer: DiSSCo is not involved in Task 11.3 "Implementation of FAIR roadmap in the Biodiversity and Ecosystem subdomain" as it is still in its early Preparatory Phase. However, it decided to participate voluntarily to the survey exercise in a simplified way.

Data Catalogue federation

Table 17. DiSSCo timeline for Data Catalogue federation

Timeline	Action	Sub-Action	F	A	I	R
2020-Q3	Create a data catalogue for Digital Specimens.	Test data type registries and experiment with DiSSCo data types	x	x	x	
2022 (or after)	Integration with EOSC service catalogue	Create detailed descriptions of various DiSSCo e-services.			x	
2022 (or after)	A generalized set of API specifications	API compliance with the Digital Object Interface Protocol (DOIP) specification requirements published by the DONA[26] Foundation.		x	x	

---

[26] https://www.dona.net/

<u>User federation</u>

Table 18. DiSSCo timeline for user federation

Timeline	Action	Sub-Action	F	A	I	R
2022 (or after)	AAI	Piloting access through an AAI infrastructure (based on AARC blueprint, also support single sign-on and ORCID integration)		x		

<u>Technologies and Certifications</u>

Table 19. DiSSCo timeline for Technologies and Certifications

Timeline	Action	Sub-Action	F	A	I	R
2021-Q1	PID implementation	Designing a PID schema	x			
2021-Q1	Data Standard Compilation	First version	x		x	x
2021-Q2	Prepare a data modelling framework	Specifications of the Digital Specimen Objects	x	x	x	
2021-Q2	Semantics	A best practice guide for semantic enhancement	x		x	
2022 (or after)	Data infrastructure	Evaluation of FAIR Digital Object technical infrastructure	x	x		x
2022 (or after)	Repository	Museum collection Management systems interoperability investigation		x	x	

Policies and licences

Table 20. DiSSCo timeline for Technologies and Certifications

Timeline	Action	Sub-Action	F	A	I	R
2019 (or before)	DMP	Periodic evaluation of the provisional version		x	x	x
2020-Q4	Software licences	Recommendations/guidelines for DiSSCo members				x

4.1.4 eLTER-RI

<u>Data Catalogue federation</u>

eLTER currently uses Dynamic Ecological Information Management System - Site and dataset registry (DEIMS-SDR) as the central catalogue to document research sites and related datasets generated at these sites and platforms.

Within eLTER PLUS,[27] the Digital Asset registry will be set up as central catalogue for eLTER Core Data from the emerging eLTER RI. This catalogue provides and will provide mapping to different metadata standards (F, A, I) enabling the integration to other catalogues (e.g., GEOSS[28]).

Table 21. eLTER timeline for Data Catalogue federation

Timeline	Action	Sub-Action	F	A	I	R
2021-Q2	Create a data catalogue for eLTER Core Data	Create Digital Asset Registry as central catalogue to document and host eLTER Core Datasets	x	x	x	
2021-Q4	Metadata interoperability	Map eLTER sites dataset metadata on related eLTER datasets, with basic information for discovery	x	x		
2020-Q3	Data interoperability	Update of the eLTER Data Specification for data submission to the network, including method description and reference lists			x	x
2022	Data interoperability	Agreement on community standards for eLTER Core Data			x	x
2021-Q2	Data interoperability	Update of the site and platform documentation to implement the core fields for site description	x		x	

<u>User federation</u>

Federated AAI among the different components of the eLTER Information System is a requirement which will be tackled in the coming years. The activities have not started yet.

---

[27] https://www.lter-europe.net/projects/PLUS
[28] https://catalogue.nextgeoss.eu/

<u>Technologies and Certifications</u>

Currently, PID (e.g., DOIs) are issued using EUDAT B2SHARE[29] repository for sharing eLTER data. This needs to be extended to apply a policy for eLTER Core Data. This is also necessary to implement provenance tracking for the data. Also, the use of DOI for sites and implemented instruments is currently under evaluation. The integration of semantic artefacts in the documentation and annotation will be fostered. This will also include further development of Environmental Thesaurus (EnvThes[30]) based on the RDA I-Adopt[31] working group's recommendations for the description of environmental observations. Provenance will build on the developments and experiences of the ENVRIplus project[32] and implemented to track workflows done in the DataLab which is currently under development.

Table 22. eLTER timeline for Technologies and Certifications

Timeline	Action	Sub-Action	F	A	I	R
2021-Q4	PID implementation	Design and concept to apply PIDs for sites and platforms documentation in DEIMS-SDR	x		x	
2021-Q4 (or after)	PID implementation	Design PID schema for eLTER Core Data	x		x	
2021-Q4 (and after)	Data repository	Extend eLTER Central Data Node to other data types (including spatial data)		x	x	
2021-Q2	Semantics	Concept for applying I-Adopt recommendations for EnvThes to extend the use as semantic backbone for the eLTER Information System			x	x
2022 (and after)	Semantics	Implement semantic annotation and mapping for variable description			x	x

---

[29] B2SHARE *is the EUDAT user-friendly, reliable and trustworthy service for researchers, scientific communities and citizen scientists to store and publish research data from diverse contexts.* (https://eudat.eu/services/userdoc/b2share-usage)
[30] https://github.com/LTER-Europe/EnvThes
[31] RDA - InteroperAble Descriptions of Observable Property Terminology (I-Adopt), https://github.com/i-adopt
[32] https://www.envriplus.eu/

| 2021-Q4 | Data provenance | First version of metadata needed to ensure data provenance and workflow descriptions (e.g., provenance templating service) | x | | | x |
| 2022 | Data provenance | Implementation of provenance and integration to metadata for selected workflows from DataLabs | x | | | x |

Policies and licences

A first version of the Data Management Plan (DMP) and Data Policy was developed within the eLTER PLUS project, which will be the basis for the emerging eLTER RI. This will be further defined and implemented within the eLTER PPP project[33] defining the scope and organization of the RI. eLTER promotes the use of open data and science, recommending CC-BY-NC as a common licence for data shared in the eLTER context. Despite the recommendation currently licences and data policies are defined on data provider level resulting in a range of different licences. This will be tackled in the eLTER PPP project defining eLTER Core Data and the related common licence for these data products.

Table 23. eLTER timeline for Policies and licences

Timeline	Action	Sub-Action	F	A	I	R
2023 (or before)	Data licence	Selection of appropriate licence (e.g., CC-BY-NC) for eLTER Core Data				x
2023	DMP and Data Policy	First draft of the DMP and Data Policy for the emerging eLTER RI		x		x

---

[33] https://www.lter-europe.net/projects/PPP

4.1.5 ICOS ERIC

Data Catalogue federation

Table 24. ICOS ERIC timeline for Data Catalogue federation

Timeline	Action	Sub-Action	F	A	I	R
2020-Q3	Metadata interoperability	First version of the metadata for translation, with basic information	x			x
2020-Q4	Data interoperability	First version of the data for netcdf preparation			x	x
2021-Q2	Metadata interoperability	Second version of the metadata for translation, with more detailed information			x	
2021-Q3	Data interoperability	Agreement on community standard using netcdf			x	x
2022 (or after)	Metadata interoperability	Third version of the metadata for translation with complete information	x		x	
2022 (or after)	Data interoperability	Second version of the data for netcdf preparation			x	x

User federation

No activity has been carried out since the beginning of ENVRI FAIR, nor it is planned in 2021

Technologies and Certifications

Table 25. ICOS ERIC timeline for Technologies and Certifications

Timeline	Action	Sub-Action	F	A	I	R
2021-Q1	Data provenance	First version of metadata needed to ensure provenance and workflow description - selected variables	x			x
2021-Q2	Metadata collection	First version of the tool for standard collection of complex metadata	x			
2021-Q3	Data provenance	Second version of metadata needed to ensure provenance and workflow description - explicit association with data	x			x
2021-Q4	Metadata collection	Second version of the tool for standard collection of complex metadata	x			
2022 (or after)	Data provenance	Third version of metadata needed to ensure provenance and workflow description - all variables	x			x

Policies and licences

No activity has been carried out since the beginning of ENVRI FAIR, nor it is planned in 2021

4.1.6 LifeWatch ERIC

<u>Data Catalogue federation</u>

A beta release of the LifeWatch **ERIC Metadata Catalogue** has been developed to host Datasets, Virtual Research Environment, Services, Workflows and Research Sites (F, A, I). The solution is based on the open-source software GeoNetwork[34] .

Next implementations:
- Enrich the APIs to query the catalogue
- Management of other resources (e.g., Instruments; Software; Projects; Manuals, guidelines, standards and best practices; etc.)
- AAAI (Authentication, Authorization and Accounting Infrastructure) mechanisms
- A prototype of the Dataset exposer is being developed adopting the Ecological Metadata Language (EML) 2.2.0 profile

Next steps:
- Catalogue population involvement of all national nodes and external providers (Research Infrastructures, initiatives, institutions, etc.)

The **ENVRI-FAIR Training Catalogue**[35] allows to manage metadata related to training materials using a customized profile of the IEEE Standard for Learning Object Metadata[36] (IEEE LOM 2002), consisting of only 27 metadata elements. The Catalogue allows to manage multi-value metadata (e.g., keywords, language, URLs, etc.). REST APIs[37] to query the catalogue have been developed (Synthetic and Detailed List of Resources; Single Resource Detail; LOM Definition; Simple and Advanced Search).

A beta release of EcoPortal[38], a Semantic Resources Catalogue focused on vocabularies and ontologies in the ecological domain, has been developed and deployed.

Next implementations:
- Extend already existing vocabularies
- LifeWatch ERIC top-level ontology
- LifeWatch ERIC domain-specific ontology
- Migration to OntoPortal-3.0-lirmm39
- Use of semantic resources for annotation in LifeWatch ERIC catalogues and national data catalogues
- AAAI mechanisms

Next steps:
- Repository population
- Harmonize/align already existing semantic resources
- Documentation

---

[34] https://metadatacatalogue.lifewatch.eu

[35] https://trainingcatalogue.envri.eu

[36] https://standards.ieee.org/standard/1484_12_1-2002.html

[37] Representational State Transfer – Application Programming Interface

[38] http://ecoportal.lifewatch.eu/

[39] https://ontoportal.org/ontoportal-virtual-appliance-release-3-0-is-here/

Table 26. LifeWatch ERIC timeline for Data Catalogue federation

Timeline	Action	Sub-Action	F	A	I	R
2019 (or before)	LifeWatch ERIC metadata (data) catalogues	Analysis of the state of the art of the different tools and framework available on the market	x	x	x	
2019 (or before)	LifeWatch ERIC metadata (data) catalogues	EcoPortal Publication			x	
2019	ENVRI-FAIR Training Catalogue	Development and deployment of the training catalogue	x	x	x	x
2020-Q1	Metadata Catalogue	β-release based on GeoNetwork	x	x	x	
2020-Q2	Exposer/Connector	First prototype based on Metacat[40] and EML 2.1	x	x	x	
2020-Q3	Metadata Catalogue	Develop EML 2.2.0 profile and Candidate release based on GeoNetwork	x	x	x	
2020-Q3	ENVRI-FAIR Training Catalogue	Updates and second release of the training catalogue	x	x	x	x
2020-Q3	EcoPortal	Updates and second release of EcoPortal	x	x	x	x
2020-Q3	Exposer/Connector	Second prototype based on Metacat and EML 2.2.0	x	x	x	
2020-Q3	Metadata Catalogue	GeoNetwork Metacat harvesting	x	x	x	
2020-Q3	Metadata Catalogue	Integration in the Virtual Research Environment (VRE)	x	x	x	
2020-Q4	Metadata Catalogue	User Training	x	x	x	x
2021-Q1	Metadata Catalogue	Full release	x	x	x	
2021-Q2	Exposer/Connector	Full release	x	x	x	

---

[40] https://github.com/NCEAS/metacat

ENVRI
FAIR

User federation

A federated AAAI protocol among LifeWatch ERIC participants has been defined, and an RI Identity provider has been developed, but not yet fully implemented. However, LifeWatch ERIC ICT Team has established an AAAI Task Force that aims to define an access policy for the data and the formal process for accessing restricted data, and external authorizations. The policy will be available in machine-readable, and LifeWatch. ERIC blockchain-based technology platform (LifeBlock) will provide a clear framework regarding data citation.

Identity Management solution: in this moment, we have a bespoke solution based on a traditional login mechanism that follows robust cybersecurity best practices.

- Next steps: we are evolving the Identity Management (IM) infrastructure by deploying a solution that allows the integration of external identity providers (e.g., Google, ORCID, etc.) and provides multi-factor authentication.

Authorization policies: at this moment, we are using traditional strategies based on closed-role management (guest, researcher, specialist, etc.). Such methods are neither easily scalable nor flexible for all the use-cases that LifeWatch ERIC needs to manage.

- Next steps: We evaluate an additional Service Level Agreement (SLA) based contract-checking framework that considers accountability issues, a more suitable approach for all the use-cases that LifeWatch ERIC needs to manage.

Technologies and Certifications

The DOI system for assigning persistent identifiers to the RI resources has been developed using the DataCite tools and already integrated in the LifeWatch ERIC Metadata Catalogue and EcoPortal. As the next step, we define a validation process to provide DOIs (criteria, working groups for assignment/validation, etc.) as a first study to identify aspects for the Provenance Management (LifeBlock). On top of this, an overall objective is to supplement those practices using the capabilities provided by LifeWatch ERIC Blockchain system, also known as LifeBlock (full link traceability coverage and provenance, among other features). In the planning, the datasets generated in the infrastructure can be given both persistent ID provided by third parties, as well as a LifeWatch ERIC generated PID. Also, they will be linked together through the LifeBlock system. To this extent, there is an active collaboration with the GEDE-RDA[41] Biodiversity task-force. A survey on the data management/curation conducted by the LifeWatch ERIC has been shaped:

---

[41] Group of European Data Experts in RDA

Table 27. LifeWatch ERIC timeline for Technologies and Certifications

Timeline	Action	Sub-Action	F	A	I	R
2020-Q2	DataCite integration into LifeWatch ERIC Metadata Catalogue	GeoNetwork - DataCite connection to provide DOIs	x		x	
2020-Q2	DataCite integration into EcoPortal	EcoPortal - DataCite connection to provide DOIs	x		x	
2020-Q4	Provenance Management (LifeBlock)	Identification of the provenance aspects in the VRE	x			x
2021-Q4	Provenance Management (LifeBlock, Tesseract)	Design of the integration of the provenance aspect in the VRE	x			x

Policies and licences

A first version of Data Management Plan (F, R) is currently being developed by LifeWatch ERIC, built on the top of the Data Management Plans already put in place by some of the LifeWatch ERIC National Nodes. LifeWatch ERIC, as a backer of general open-access policies, is designing guidelines and mechanisms (i.e., machine-readable). This can encourage selecting the most suitable OSI[42]-backed open-source and Creative-Commons licences. The data policies already put in place by some of the LifeWatch ERIC National Nodes.

Table 28. LifeWatch ERIC timeline for Policies and licences

Timeline	Action	Sub-Action	F	A	I	R
2019 (or before)	Licences	Adoption of CC-BY				x
2020-Q2	DMP and Data Policies	1st Draft		x		x

---

[42] Open Source Initiative

ENVRI
FAIR

4.1.7 SIOS

Data Catalogue federation

SIOS has been working to integrate contributing data centres in the central node - SIOS members operate data centres that offer machine interfaces to discovery metadata. Through the utilization of these interfaces and tweaking of the metadata harvested, unified dataset discovery is achieved. Now, more than 1500 datasets are available in SIOS Data Access Point. The SIOS Data Access Point (DAP) is the entry point to SIOS datasets, including the annual State of Environmental Science in Svalbard (SESS) reports' datasets and collected by the SIOS Access Programme. It offers a web interface that contains information about datasets (metadata). The search and results interface have been upgraded at the end of June 2020[43]. The new GUI[44] includes a new WMS[45] client based on OpenLayers (v6)[46] which shows all results on a base map, with different projections and interactive behaviours. It also contains a new interface for the time series visualization. This new visualization tool is now offering dynamical access to data values and plotting, supporting time series and vertical profiles. It is based on Python, using the Bokeh plotting library[47]. A new version of the search interface will be implemented to ensure more dynamical filtering of datasets. For the machine-readable catalogue, SIOS implements a CSW[48] endpoint that will expose metadata in Directory Interchange Format (DIF) and ISO 19115.

Table 29. SIOS timeline for Data Catalogue federation

Timeline	Action	Sub-Action	F	A	I	R
2021-Q1	Improving SIOS Data Management System (SDMS) tools enabling FAIR data	Adaptation of the Rosetta[49] implementation converting unstructured data from CSV to NetCDF/CF[50]	x		x	x
2021-Q2	Improving SDMS tools enabling FAIR data	Improving an Excel template generator for registration of data.			x	

---

[43] https://sios-svalbard.org/metadata_search
[44] Graphical User Interface
[45] Web Map service
[46] https://openlayers.org/
[47] https://bokeh.org/
[48] Catalogue Service for the Web
[49] http://tomcat.nersc.no/rosetta/
[50] CF (Climate and Forecast) metadata (https://cfconventions.org/)

Timeline	Action	Sub-Action	F	A	I	R
2021-Q2	Case study 3	Improving and updating the SIOS Observation Facility Catalogue according to the standards from the WIGOS[51].	x	x	x	x
2021-Q4	CSW endpoint	A CSW endpoint will be exposed including metadata records which are findable on the SIOS data access	x	x	x	

<u>User federation</u>

SIOS does not require authentication/authorization to search and access data within the portal. Authorization is only needed for additional services, such as using Basket[52], i.e., to perform actions on multiple products.

Table 30. SIOS timeline for User federation

Timeline	Action	Sub-Action	F	A	I	R
2021-Q4	evaluation of eduGAIN	Possible integration of eduGAIN authentication will be evaluated to provide login to the SIOS portal		x		

<u>Technologies and Certifications</u>

SIOS has started to implement pyCSW [53]for machine-to-machine interfaces to metadata. The development is under a testing phase, but the CSW catalogue will expose DIF and ISO 19115 output profiles

Interoperability Guidelines for the SIOS Data Management System (SDMS) is nearing completion. The guidelines are used to define how contributing data centres can connect to the SDMS[54]

SIOS does not assign or manage PIDs for the contributing centres, which have full ownership of the datasets provided via the SIOS data portal. On the other hand, SIOS recommends having DOIs in place, and the different data centres are aligning to this recommendation. In this context, SIOS is also working on identifying core data, with particular relevance for the critical questions within Arctic research, which shall require a DOI.

---

[51] WMO Integrated Global Observing System (WIGOS) metadata standard. (World Meteorological Organization, 2019)

[52]See https://owwl.org/help/my-lists

[53]https://github.com/geopython/pycsw

[54] https://github.com/SIOS-Svalbard/SDMSInteroperabilityGuidelines

Table 31. SIOS timeline for Technologies and Certifications

Timeline	Action	Sub-Action	F	A	I	R
2021-Q2	Improving FAIRness of biodiversity data managed by SIOS and served to SDMS	Transforming existing data into Darwin Core Archives using the Darwin Core glossary of terms - Establishing a machine service transforming existing terminology for the relevant data to Darwin Core terminology	x	x	x	x
2020-Q2	Improvements in how FAIR biodiversity data are handled in SDMS	Updating the SIOS Data Management System (SDMS) Interoperability Guideline Document with more detailed information on how biodiversity and ecosystem data are documented and exchanged between contributing data centres in SDMS			x	
2020-Q4	Improvements in how FAIR biodiversity data are handled in SDMS	Training in how to generate a Darwin Core Archive using metadata according to the Ecological Metadata Language (EML)	x	x	x	x
2020-Q4	Improvements in how FAIR biodiversity data are handled in SDMS	Adding support for mapping to/from EML and other metadata standards - Enabling support for indexing EML powered datasets in SDMS	x	x	x	
2021-Q2	Improvements in how FAIR biodiversity data are handled in SDMS	Initiating semantic search based on various semantic frameworks. - Testing only, no operational deployment	x		x	

Policies and licences

SIOS has delivered a revised SIOS Data Policy document. Beneficiaries of the SIOS services and activities follow the recommendations and regulations for data sharing and improving the data FAIRness based on SIOS Data Policy.

SIOS has been working on the update of the internal metadata representation to support URLs for licences in line with TF5 (through SPDX[55], which helps machine-readability) and to link metadata items for platforms and instruments to WIGOS metadata standard. The new documentation for the internal metadata standard MET Norway Metadata Format (MMD) is compliant with ISO and DIF and available online[56].

Table 32. SIOS timeline for Policies and licences

Timeline	Action	Sub-Action	F	A	I	R
2021-Q4	URL for licence	Licences will be exposed in a linked-data approach using SPDX URLs				x

---

[55] Software Package Data Exchange
[56] https://github.com/metno/mmd/blob/master/doc/mmd-specification.pdf

## 4.2   User-oriented cross-RIs demonstration cases

Due to the high heterogeneity of the WP11 RIs, the definition and implementation of common use cases are key to developing interoperability within the subdomain. Therefore, the FAIRness implementation will focus on three subdomain topics defined as showcases:

1. **Soil Water Content** (SWC)**,**
2. **Species Scientific Names Identification** (SNI)**,**
3. **Site Documentation Interoperability** (SDI)**.**

Also, a fourth transversal Use case,

4. **Common layer of Core MetaData** (CMD)

will ensure that the services developed by the others can be integrated in the ENVRI catalogue.
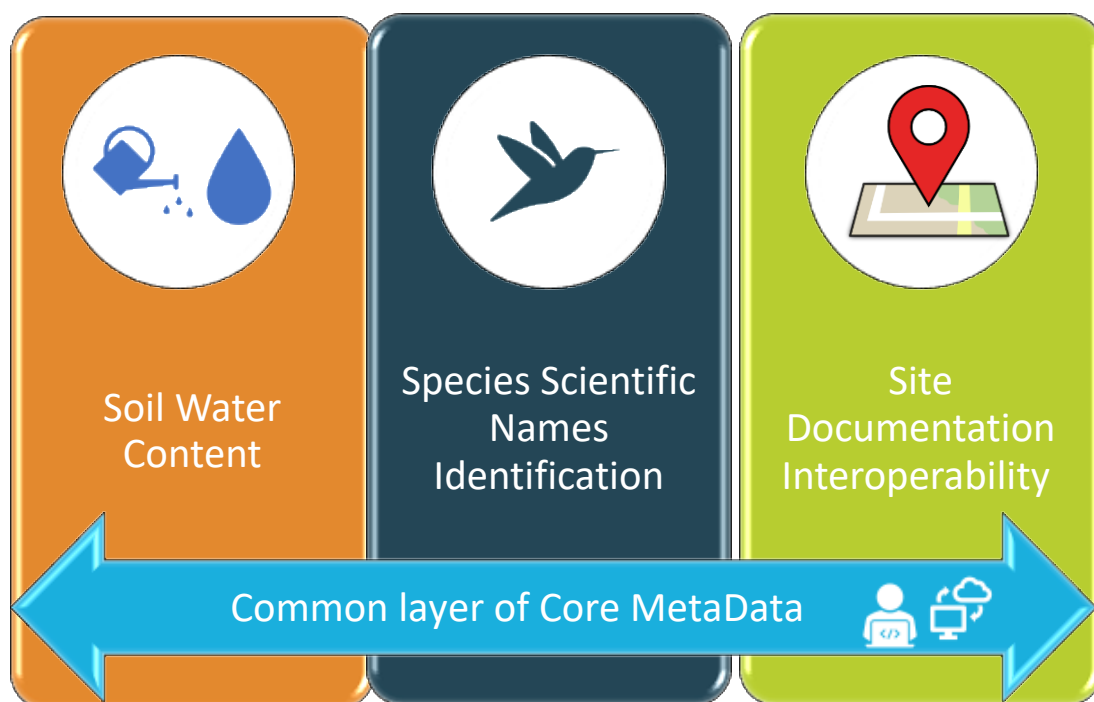
Figure 2. Visual representation of the four WP11 Use Cases, including their interaction

The selection of cross-RIs' use cases has been considered a good starting point to build a common background and to test the FAIRness and the interoperability among the RIs of the subdomain.

The current composition of the UCs is as follows:

Table 33. Coordinators and RIs involved in the WP11 Use Cases

	UC1 - SWC	UC2 - SNI	UC3 - SDI	UC4 - CMD
Coordinators	A. Chanzy G. L'Abate	S. Islam	C. Wohner C. Pichot	D. De Nart
Involved RIs	**AnaEE** Danubius eLTER ICOS LifeWatch SIOS	AnaEE **DiSSCo** eLTER ICOS LifeWatch	**AnaEE** **eLTER** ICOS LifeWatch SIOS	**AnaEE** Danubius DiSSCo eLTER ICOS LifeWatch SIOS

The UCs are deeply involved in the Task Forces' activities[57] as a cross-feeding process, as shown in Table 34.

Table 34. Contribution of WP11 use cases to WP5 Task Forces

Task Force	UC1 SWC	UC2 SNI	UC3 SDI	UC4 CMD
TF1. ENVRI Catalogue	✓	✓	✓	✓
TF2. AAI Implementation	✓	✓		
TF3. PIDs, identification types and registries		✓		
TF4. Triple stores and data storage certification			✓	✓
TF5. Licences, citation and usage tracking				
TF6. User-oriented cross-domain demonstration cases	✓		✓	

---

[57] The presentation given during the TFs workshop (Nov, 25th-26th 2020) is available here: https://iagos-comm.iek.fz-juelich.de/dmsf/files/5557/view

ENVRI
FAIR

4.2.1 Soil Water Content (SWC)[58]

Objective

Soil water content is a crucial environmental variable, essential to, e.g., farmers, meteorologists, water resource management and disaster management units (Bauer-Marschallinger et al., 2019; Shi et al., 2006). SWC presents extreme Spatio-temporalization linked to the soil's heterogeneity, vegetation and climatic conditions, and the relief that can contribute to water redistribution by runoff. Global soil moisture products (remote sensing, SWC monitoring network - International Soil Moisture Network -ISMN) remain limited to the soil surface and/or provide little information on contextual data that drive SWC temporal dynamic.

RIs offer networks with highly instrumented sites including SWC and most of the contextual data which are required by:

- the science community (to calibrate and implement models, soil moisture being attractive by itself or can be co-variable of other processes - erosion, compaction, biochemical cycle, soil biology …)
- space agency interested by developing advanced products (level 4)
- water resources management, risk assessment, climate change stakeholders

This use case will deliver comprehensive and interoperable data sets, including SWC and context data and metadata from different RIs.

Description of work

After a preliminary meeting at the Dresden ENVRI week, finalized to define the use case, two operational sessions took place in December 2020, finally defining participants and roles and drawing the roadmap.

As for the metadata standard, the ISMN template was recognized as a reference, to extend mandatory metadata according to:

i.   their availability within different RIs;
ii.   the results of a poll addressed to data users on the main criteria to find Soil Water Content data, related data and context data.[59]

Site description metadata will comply with specifications from Use Case on Site documentation and interoperability (SDI). Each RI will define at least one site to demonstrate interoperability for SWC.

Implementation plan

- Identification of data sets that will contribute (2020)
- Final Design of the service (2020)
- Technical specification of the different components (01-2021)
- Service architecture and technology selection (03-2021)
- Implementation (2021)

4.2.2 Species Scientific Names Identification (SNI)[60]

Objective

The goal of this use case is to explore how we can use the scientific names in a FAIR way across all the Research Infrastructures.

Species scientific names are basic information that all the RIs in the Biodiversity and Ecosystem domain have to handle. At the moment, there is no agreement upon procedures on how these names should be defined, assigned, used and distributed in different RIs.

Description of work

The following activities have been planned:

- − analysis of the current procedures regarding the usage of scientific names across different RIs. The outputs of this analysis will be a summary of the current methods and the identification of common elements and divergent practices.
- − setting up of a communication channel with Catalogue of Life (CoL)[61] to provide a summary of the analysis.
- − analysis, with various RIs and Catalogue of Life, on how the recent development of Catalogue of Life ChecklistBank[62] can provide FAIR access to scientific names.
- − development of FAIR guidelines for common usage of scientific names and taxonomic data sharing across RIs.

Implementation plan

The implementation plan of UC2 is moving its first steps as additional technical clarifications and status of the use case are needed. Anyway, participants in UC2 are providing some examples of CoL data usage (such as subspecies level data, link to synonyms and variants). They are assessing the possibility to extend the use of CoL data: CoL data are dataframe containing metadata about each sample. They should include a sample identifier as well as any relevant experimental factors. Besides, LifeWatch ERIC provides an example of dataset and services that uses taxonomic backbone to pave the way of the implementation plan.

---

[58]A living document on the use case activities is available at the following link:
https://docs.google.com/document/d/1pzczbVENA9PvHlyt3oBhODFv-qC8XNPQ/edit
[59]https://docs.google.com/forms/d/e/1FAIpQLSe2IWLyTQJURkVfcySjNnWjJcwra18MjO0i0FpQzYUQOFfOeg/viewform?vc=0&c=0&w=1&flr=0).
[60] A living document on the use case activities is available at the following link:
https://docs.google.com/document/d/14kx_6t-0pPzSKPOafa45f_9MFarT5SPXdERuUCO94_M/edit#heading=h.jptk0pgm43da
[61] https://www.catalogueoflife.org/
[62] https://data.catalogue.life/) data repository and API (https://api.catalogue.life/)

4.2.3 Site Documentation Interoperability (SDI)[63]

Objective

Information related to research sites is essential when describing the context of observations. It is a key element of site-based research infrastructures (RIs) and their catalogues (Wohner et al., 2019). Such information is needed to describe the actual extent of ecosystem research infrastructures (in Europe) and identify potential gaps and biases (Martin et al., 2012). This would foster analyses of the existing ecosystem research infrastructures on a national, regional and global scale and increase these infrastructures' ability to answer large-scale environmental questions (Wohner et al., 2020).

The main objective of this use case is to increase interoperability of research site descriptions by:

- proposing a set of minimum information and, if possible, for each participating RI implementing and exposing such information
- harmonizing terminology used across RIs to ease data aggregation

Another objective is, if possible, to develop a routine for selected RIs to serve site information through a (graph-based) common interface to distributed documentation of research sites for discovery.

Description of work

The following activities have been planned:

- compiling an overview of site documentation systems and formulating recommendations for increasing interoperability. This activity is based on a recent survey (Wohner et al., 2020) and ongoing activities of the involved Ris;
- having a workshop on December 9th, 2020) with all involved RIs about:
  - i) feasibility of implementing these recommendations, harmonizing terminology and definitions used across RIs, aggregating RI site information and making it discoverable through a centralized (graph-based) interface,
  - ii) defining the expected outcome of the use case
- achieving increased interoperability of research site documentation between the RIs (autumn 2021);
- developing and deploying a web-accessible prototype that allows to search site information across RIs (summer 2022);
- ensuring seamless feeding of the ENVRI-hub through the generation of DCAT formatted metadata based on the site documentation data provided by the Ris.

Implementation plan

Based on the December use case meeting the following first implementation steps have been agreed to start the implementation:

---
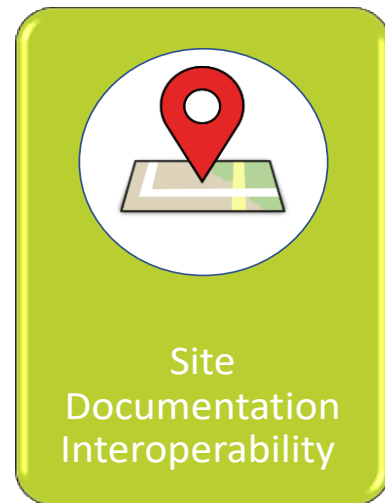
[63] A living document on the use case activities is available at the following link:
https://docs.google.com/document/d/10heAQ75WKnh9b6b_nO0pZ-An9z-b0ZOOoYp7icdknr0/edit?usp=sharing

- evaluation of the core fields and existing terminology for site documentation for the site catalogues of the participating RIs. Also, it has been agreed to work on a mapping to share site information via a standardized information model;
- showcase the mapping and harvesting of site documentation based on ISO19115/19139 for AnaEE, LifeWatch and eLTER (DEIMS-SDR) for the ENVRI hub DCAT.

If the results of the showcase implementation of AnaEE, LifeWatch and eLTER turn out to be satisfactory, the other RIs (ICOS, SIOS), would also implement these workflows as far as it is possible, with regards to partially fundamental differences in the setup of the systems, databases and information models.

4.2.4 Common layer of Core MetaData (CMD)[64]

Objective
The main goal is fostering inclusion of WP11 case studies services into the ENVRI catalogue. It will be reached by enforcing DCAT metadata over the services produced by the other three case studies.

Common layer of Core MetaData

Description of work
Schedule: one preliminary action and then a loop of feedback with the other case studies.

- Identification of mandatory and optional properties within the DCAT vocabulary and preparation of a handbook for using them.
- Production of RDF data by the other three use cases, with values for all mandatory properties at least (optional ones are welcome, though).
- Quality assessment on the RDF mentioned above data. If quality goals are not met, the previous step will be repeated until the quality goals will be met.
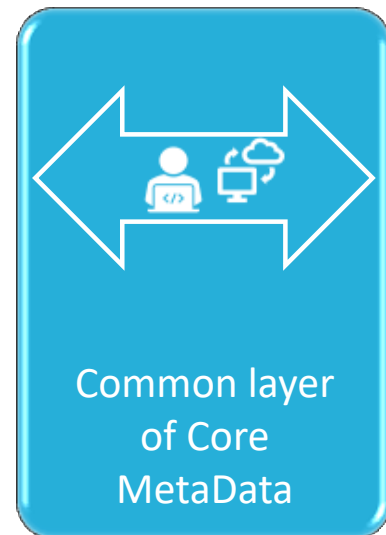
Quality goals: valid RDF data, proper usage of DCAT primitives, absence of loops in the RDF graph, reasonable degree of connection to the Linked Open Data cloud[65].

Implementation plan
The participants in UC4 are working on the preparation of DCAT compatible RDF data. Indeed, in UC4 the DCAT properties are used to describe the data produced by the demonstrators of the cases, or the services (as in UC2, as it is mostly a service and not the actual data). The implementation phase foresees:

1. identification of data produced by the demonstration cases (UC1-3),
2. identification of the reasonable properties to expose for this data or services,
3. collection of the values for the selected properties from the demonstration cases.

These values might be from controlled vocabularies and via a good enough template, that could be reused in the future. AnaEE will provide additional information about available template for metadata and inclusion of vocabularies, to allow the RDF production. Indeed, AnaEE uses a vocabulary based on the semantic approach, which will give information already formatted in DCAT. So, it is quite integrated into the semantic pipeline process.

---

[64] A living document on the use case activities is available at the following link:
https://docs.google.com/document/d/1SSvbOxPe4GUWyzl5pZcyFsJIxfGhPuzHeO7dKvOCs1A/edit
[65] https://www.w3.org/egov/wiki/Linked_Open_Data

5 Conclusions

This Deliverable describes the ongoing data FAIRness implementation within the Biodiversity and Ecosystem subdomain towards the improved FAIRness and the overall ENVRI FAIR project developments. In particular, the current compliance of the subdomain RIs with TF recommendations has been analysed considering the central role assumed by the WP5 Task Forces within the project.

The subdomain landscape is very heterogeneous[66], and the RIs' different maturity levels are mainly due to their specific scientific domain, knowledge and skill (e.g. current standards and IT support) and services offered as well as their lifecycle and organization. However, the situation is encouraging, as all subdomain RIs appear to have ongoing and planned activities to close their significant gaps, as reported in their roadmaps (see Chapter 4). In this context, the definition and implementation of cross-RIs' use cases have been considered a good starting point to build a common background and test the FAIRness and interoperability among the subdomain's RIs. Three subdomain topics have been selected as showcases: Soil water content, Species scientific names identification, Site documentation interoperability. All RIs are involved in at least one use case, and their activities started in Autumn 2020. Besides, a fourth transversal use case "Common layer of core metadata" will ensure that the services developed by the others can be integrated in the ENVRI catalogue. This latter is a key goal for the overall project. The ENVRI Catalogue of Services (machine-actionable interface) is one of the three main components of the ENVRI-hub, together with the ENVRI Knowledge Base (human interface) and the cross-domain demonstrator services.

To this extent, the Deliverable is not considered a final, complete and closed document. It is expected that the findings of developments emerging from the activities reported in this document can improve many aspects of cooperation in FAIRness within the Biodiversity and Ecosystem domain. However, a "final FAIRness stage" cannot be achieved; it is indeed a continuous effort with embedded risks and weaknesses. New technologies become available and need evaluation, new skills are required, and a sound interaction with the so-called big-players[67] is necessary.

FAIRness is a journey, and it is uncertain how much of this development work can be achieved within the resources and timescale of ENVRI-FAIR (Jeffery et al., 2020).

---

[66] See Deliverable 11.1 (Papale, 2020),
[67] In this context big players refer both to technologies developers and networks/initiatives (e.g RDA, EOSC…)

6   Impact on the project

WP11 considers this Deliverable a starting point for implementing FAIR principles in the Biodiversity and Ecosystem subdomain and paving the way towards the cooperation among the ENVRI FAIR project WPs and groups. It provides information for other project members in how the seven Research Infrastructures involved in the Biodiversity and Ecosystem subdomain are improving FAIRness and provides a basis for discussion on best practices. The actual cooperation and testing are carried out, at WP11 level, through the activities of the Use Cases and, at project level, through the exchange with the Task Forces (see Chapter 2). As a piece of example, the outcomes of the implementation survey (see Chapter 3.1) and more precisely the matrix (see Figure 1) can be proposed as a benchmarking tool and offered as a basis on which both WP11 and the other subdomains (**WP8, 9 and 10**) in the project can build a more structured discussion.

Furthermore, besides the cooperation among the four subdomains, it emerges from the report that WP11 has actively collaborated with other WPs. Notably, in this Deliverable the paramount role of **WP5** has been highlighted. Having identified gaps in FAIRness in collaboration with WP5 (Magagna et al. 2020) and laid the groundwork for the implementation plan in Task11.2 "Analysis of status and plan for development and harmonization for RIs data FAIRness" (Papale, 2020) an updated status together with an advanced plan for implementation is provided in this Deliverable.

WP6 "Training and capacity building" is also a key counterpart in terms of possible and desirable improved cooperation. The findings and needs of the Use Cases can easily be transferred to the WP6 to assess the possibility of future training activities and propose the topics that might be useful for the four subdomains (Hellström, 2019)

7   Impact on stakeholders

Stakeholders of the Biodiversity and Ecosystem subdomain – researchers, government agencies, commercial organizations, educators, interested citizens – represent a much larger group than the one involved in the ENVRI FAIR project. However, to a different degree, they can already benefit from the findability, accessibility, interoperability, and reusability of data implemented by the participating RIs. On one side, we can consider this a very initial step to set the stage for reliable interaction (in the broadest sense) among RIs in the subdomain. On the other side, the Biodiversity and Ecosystem subdomain aims to eventually integrate its results into the ENVRI-hub to provide a much wider group with appropriate services and asset access. Populating the EOSC asset catalogue from the ENVRI-hub will reach an even wider group of stakeholders, providing them with the services and assets to approach a wide range of environmental problems.

8   Acknowledgements

9 References

Adamaki A., Vermeulen A. (2020). ENVRI-FAIR D5.2 Implementation plan for common development goals (Version 2.1). Zenodo. https://doi.org/10.5281/zenodo.4061702

Bauer-Marschallinger B., Freeman V., Cao S., Paulik C., Paulik S., Stachl T., Modanesi S., Massari C., Ciabatta L., Brocca L., Wagner W. (2019). Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles. In IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 1, pp. 520-539, Jan. 2019, doi: 10.1109/TGRS.2018.2858004

Hellström M., Johnsson M., Konijn J., Fiore N., Quimbert E., Boulanger D., Baker G.R. (2019). ENVRI-FAIR D6.1 Inventory & gap analysis of FAIR training materials (Version 1). Zenodo. https://doi.org/10.5281/zenodo.3885122

Jeffery K.G., Bailo D., Rocquencourt J.B., Langeland T., Spinuso A., Trani L., Manunta M., Rodero I. (2020). ENVRI-FAIR D10.4: Report on Implementation Activities (Version 1). Zenodo. https://doi.org/10.5281/zenodo.4418953

Magagna, B., Adamaki, A., Liao, X., Rabissoni, R., Zhao, Z. (2020). ENVRI-FAIR D5.1 Requirement analysis, technology review and gap analysis of environmental RIs (Version 1). Zenodo. https://doi.org/10.5281/zenodo.3884997

Martin L.J., Blossey B., Ellis E., (2012). Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. Frontiers in Ecology and the Environment. 10. 195-201. https://doi.org/10.1890/110154

Papale, D. (2020). ENVRI-FAIR D11.1 Biodiversity and Ecosystem subdomain implementation short term plan (Version 1). Zenodo. https://doi.org/10.5281/zenodo.3885360

Shi J., Jiang L., Zhang L., Chen K.S., Wigneron J.P., Chanzy A., Jackson T.J. (2006). Physically Based Estimation of Bare-Surface Soil Moisture With the Passive Radiometers. In IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 11, pp. 3145-3153, Nov. 2006, doi: 10.1109/TGRS.2006.876706

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

Wohner C., Peterseil J., Poursanidis D., Kliment T., Wilson M., Mirtl M., Chrysoulakis N. (2019) WMO DEIMS-SDR – A web portal to document research sites and their associated data. Ecological Informatics. 51, 15-24. https://doi.org/10.1016/j.ecoinf.2019.01.005

Wohner C., Peterseil J., Genazzio M.A., Guru S., Hugo W., Klug H. (2020). Towards interoperable research site documentation – Recommendations for information models and data provision. Ecological Informatics. 60, 101158. https://doi.org/10.1016/j.ecoinf.2020.101158

World Meteorological Organization (2019). WIGOS Metadata Standard. WMO-No. 1192. ISBN 978-92-63-11192-0.
https://repository.oceanbestpractices.org/bitstream/handle/11329/1379/1192_en.pdf

10 Appendices

10.1 Appendix A: Glossary and terminology

The following is the list of acronyms and terms used in this deliverable[68]:

AAAI	Authentication, Authorization and Accounting Infrastructure
AAI	Authentication and Authorisation Infrastructure
AnaEE	Analysis and Experimentation on Ecosystems
API	Application Programming Interface
CC0	Creative Commons – Not rights reserved
CC-BY-NC 4.0	Creative Commons attribution non-commercial license
CC-PDM	Creative Commons - Public Domain Mark
CF	Climate and Forecast (semantics for NetCDF)
Ckan	Comprehensive Knowledge Archive Network. It is an open Source data portal platform
CMD	Common layer of core metadata
CoL	Catalogue of Life
CSW	Catalogue Service for the Web
DANUBIUS-RI	International Centre for Advanced Studies on River-Sea Systems
DAP	Data Access Point
DataCite	A leading global non-profit organization that provides persistent identifiers (DOIs) for research data and other research outputs
DCAT	Data Catalogue Vocabulary
DEIMS-SDR	Dynamic Ecological Information Management System - Site and dataset registry
DIF	Directory Interchange Format (National Aeronautics and Space Administration)
DiSSCo	Distributed System of Scientific Collections
DMP	Data Management Plan
DOI	Digital Object Identifier
EcoPortal	The LifeWatch ERIC comprehensive repository of ecological ontologies
eduGAIN	EDUcation Global Authentication INfrastructure
eLTER	Long-Term Ecosystem Research in Europe
eLTER PLUS	H2020 project on eLTER Advanced Community Project
eLTER PPP	eLTER Preparatory Phase Project (H2020 project)
EML	Ecological Metadata Language
ENVRI	Environment research infrastructures
ENVRI-hub	A federated machine-to-machine interface to access environmental data and services provided by the contributing ENVRIs
EnvThes	Environmental Thesaurus
EOSC	European Open Science Cloud

---

[68] The latest version of the master list of the glossary is available at
http://doi.org/10.5281/zenodo.3465753.

EPOS	European Plate Observing System
ERIC	European Research Infrastructure Consortium
FAIR	Findable Accessible Interoperable Reusable
GEDE-RDA	Group of European Data Experts in RDA
GeoNetwork	A catalogue application to manage spatially referenced resources
GEOSS	Global Earth Observation System of Systems
GUI	Graphical User Interface
I-ADOPT (RDA)	InteroperAble Descriptions of Observable Property Terminology (it is a RDA Working Group)
ICOS	Integrated Carbon Observation System
IM	Identity Management
ISMN	International Soil Moisture Network
ISO	International Organisation for Standardization
LDAP	Lightweight Directory Access Protocol
LifeBlock	LifeWatch. ERIC blockchain-based technology platform)
LifeWatch	LifeWatch European Research Infrastructure Consortium
LOM	Learning Object Metadata
Metacat	Metacat: Data Preservation and Discovery System
NetCDF	Network Common Data Format.
OAUTH	Open Authorisation (standard)
ODC-PDDL	Open Data Commons Public Domain Dedication and License
OntoPortal	Repository code for semantic content from any domain
OpenID	Open standard authentication protocol (it allows for signing into multiple websites with a unique account)
ORCID	Open Researcher and Contributor ID
OSI	Open Source Initiative
OWL	Web Ontology Language
PID	Persistent Identifier
pyCSW	An OGC CSW (Open Geospatial Consortium Catalogue Service for the Web) implementation written in Python
QC	Quality Control
RDA	Research Data Alliance
RDF	Resource Description Framework
REST	REpresentational State Transfer
RI	Research Infrastructure
SDMS	SIOS Data Management System
SEISM	SoftwarE for InfraStructure administration (France)
SESS	State of Environmental Science in Svalbard
SIOS	Svalbard Integrated Arctic Earth Observing System
SPARQL	SPARQL Protocol and RDF Query Language
SPDX	Software Package Data Exchange
SDI	Site Documentation Interoperability
SLA	Service Level Agreement
SNI	Species scientific Names Identification
SWC	Soil Water Content
TF	Task Force
UC	Use Case
URI	Uniform Resource Identifier
URL	Uniform Resource Locator

USGS	United States Geological Survey
VRE	Virtual Research Environment
WIGOS	WMO Integrated Global Observing System
WMS	Web Map service
WP	Work Package