



# D11.5

## Final Report on Implementation activities in the Biodiversity and Ecosystem subdomain

Work Package	WP11
Lead partner	CREA
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	2022-01-31
Submission date	2023-05-22

### 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 analysed, through a specific questionnaire administered to all subdomain RIs. Considering 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	Luca Cervone	CREA (AnaEE)	2023-03-10
Contributing Authors	Dario De Nart Cristian Cudalbu Pierpaolo Culurciello Lara Ferrighi Nicola Fiore Paul Gasner J. M. García Rodríguez J.M. González-Aranda Darius Ignatiuk Giovanni L'Abate Barbara Parisse J. Peterseil Christian Pichot Ilaria Rosati A. J. Sáenz-Albanés F. M. Sánchez Cano Lucia Vaira Christoph Wohner Alberto Basset Xeni Kechagioglou Sharif Islam André Chanzy Dario Papale Francesca De Pascalis Rudolf Denkmann	CREA (AnaEE) DANUBIUS-RI DANUBIUS-RI SIOS LifeWatch ERIC DANUBIUS-RI University of Seville LifeWatch ERIC SIOS CREA (AnaEE) CREA (AnaEE) eLTER INRAe (AnaEE) CNR (LifeWatch ERIC) LifeWatch ERIC LifeWatch ERIC LifeWatch ERIC eLTER LifeWatch ERIC LifeWatch ERIC DiSSCo INRAe (AnaEE) CMCC (ICOS) DANUBIUS-RI SIOS	2023-03-10
Reviewer(s)	Sharif Islam André Chanzy Dario Papale Francesca De Pascalis Rudolf Denkmann J. Peterseil Christoph Wohner Lucia Vaira	DiSSCo INRAe (AnaEE) CMCC (ICOS) DANUBIUS-RI SIOS eLTER eLTER LifeWatch ERIC	2023-03-17
Approver	Andreas Petzold	FZJ	2023-05-22

## DELIVERY LOG

Issue	Date	Comment	Author
V 1.1		Sent for review	2023-03-10
V2.0		Sent for approval	2023-03-17
V2.0		Sent for submission to PO	2023-05-05

## DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the Project Manager at [manager@envri-fair.eu](mailto: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.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

D11.5 - Final Report on Implementation activities in the Biodiversity and Ecosystem subdomain.....	5
1 Introduction.....	5
2 Participation of WP 11 in WP 5 Task Forces.....	8
2.1 TF1 - ENVRI Catalogue of services .....	8
2.2 TF2 - AAI implementation.....	9
2.3 TF3 -PIDs, identification types and registries .....	9
2.4 TF4 -Triple stores and data storage certification .....	9
2.5 TF5 -Licences, citation and usage tracking .....	9
2.6 TF6 -ENVRI-hub design and architecture .....	9
3 Current state of the art.....	11
3.1 Situation as of D 11.2.....	11
3.2 Data catalogue federation readiness .....	12
3.3 User federation readiness .....	13
3.4 Technologies and Certifications .....	13
3.5 Policies and Licenses .....	14
3.6 Implementation Progress Overview .....	15
4 Roadmap .....	17
4.1.1 AnaEE-RI .....	17
4.1.2 DANUBIUS-RI .....	19
4.1.3 DiSSCo .....	21
4.1.4 eLTER-RI .....	23
4.1.5 ICOS ERIC .....	25
4.1.6 LifeWatch ERIC .....	27
4.1.7 SIOS .....	29
4.2 User-oriented cross-RIs demonstration cases.....	32
4.2.1 Soil Water Content (SWC) .....	33
4.2.2 Species Scientific Names Identification (SNI) .....	35
4.2.3 - Site Documentation Interoperability (SDI).....	37
4.2.4 - Common layer of Core MetaData (CMD) .....	39
5 Conclusions.....	40
6 Impact on the project.....	41
7 Impact on stakeholders.....	41
8 Acknowledgements.....	41
9 References.....	41
10 Appendices.....	43
10.1 Appendix A: Glossary and terminology .....	43

# D11.5 - Final Report on 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 (Wilkinson 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	
○ Repositories with FAIR access technologies	
– A1.2 The protocol allows for an authentication and authorization procedure, where necessary	57%
○ Repositories with statement on access policy in metadata	
○ Repositories with FAIR authentication and authorization protocol	57%
	43%
A2. Metadata are accessible, even when the data are no longer available	
– Longevity plan for metadata	
– Metadata openly available	0%
	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	
– R1.1. (Meta)data are released with a clear and accessible data usage licence	
○ FAIR compliant licences	57%
– R1.2. (Meta)data are associated with detailed provenance	
○ 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-RI), 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	<a href="http://www.anaee.eu">www.anaee.eu</a>
DANUBIUS-RI	International Center for Advanced Studies on River - Delta - Sea Systems	<a href="http://www.danubius-ri.eu">www.danubius-ri.eu</a>
DiSSCo	Distributed System of Scientific Collections	<a href="http://www.dissco.eu">www.dissco.eu</a>
eLTER RI	European Long-Term Ecosystem, Critical Zone and Socio-ecological Research Infrastructure	<a href="http://www.lter-europe.net">www.lter-europe.net</a>
ICOS ERIC	Integrated Carbon Observation System - European Research Infrastructure Consortium	<a href="http://www.icos-cp.eu">www.icos-cp.eu</a>
LifeWatch ERIC	e-Science and Technology European Infrastructure for Biodiversity and Ecosystem Research - European Research Infrastructure Consortium	<a href="http://www.lifewatch.eu">www.lifewatch.eu</a>
SIOS	Svalbard Integrated Arctic Earth Observing System	<a href="http://sios-svalbard.org">sios-svalbard.org</a>

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 Identification (SNI) 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 Participation of WP 11 in WP 5 Task Forces

Considering 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<sup>1</sup>.

The task forces provide support in **designing and implementing the ENVRI catalogue to be integrated into European Open Science Cloud<sup>2</sup> (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 25<sup>th</sup>/26<sup>th</sup>, 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)

### 2.1 TF1 - ENVRI Catalogue of services

Given the future connection with the EOSC service ecosystem, **TF1 aims at implementing a common metadata catalogue schema** with associated management services to make the resources of ENVRI RIs findable and accessible. WP 11 RIs actively contributed to the TF discussion and organized activities such as the population of the ENVRI Catalogue Proof of Concept, which now includes several services offered by WP11 RIs. WP 11 has fully received the guidelines produced by TF 1 and has instituted a Use Case dedicated to the implementation of TF 1 guidelines.

<sup>1</sup> 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/>

<sup>2</sup> <https://eosc-portal.eu/>



## 2.2 TF2 - AAI implementation

The ongoing discussion in ENVRI-FAIR highlighted the importance of establishing a chain of trust among RIs to provide federated access to ENVRI resources.

TF 2 recommends the usage of industry standard protocols for AAI, namely OAuth2 and OpenID Connect, a best practice that WP11 fully acknowledges, as documented in Deliverable 11.2 and in the remainder of this document. WP11 representatives took active role in TF2 discussion and contributed to the blueprint architecture outline.

## 2.3 TF3 -PIDs, identification types and registries

Persistent identifiers (PIDs) are key attributes to ensure findability of ENVRI resources and TF 3 has provided the ENVRI community with a venue to discuss the benefits of different PID schemes and paradigms. WP 11 RIs participated into the TF 3 discussion actively and acknowledge the recommendation of adopting DOI as PID of choice within the ENVRI community.

## 2.4 TF4 -Triple stores and data storage certification

Triple stores for knowledge representation play a key role in interoperability, and TF 4 provided the ENVRI community with a venue to discuss experiences and best practices in Triple store management. Although the TF did not produce best practices and guidelines, WP 11, due to its active participation, would benefit from the insights thereby provided by other ENVRI RIs, and are using such insights in the development of Use Cases 1, 2, and 3 described in the reminder of this document.

## 2.5 TF5 -Licences, citation and usage tracking

Data Licensing is a key aspect of data FAIRness, as it determines how much reusable a data asset really is. Data citation and user tracking, on the other hand are topics of great interest for data publishers as they allow them to quantify the impact of their data assets. TF 5 was established to provide RIs with a venue to discuss licensing and usage tracking related issues and to share know-how with the aim of providing a set of common recommendations for documenting licences and data policy within the metadata, indicating which metadata items should be used.

The Task Force identified Creative Commons licenses as preferable as they can fit a wide variety of intellectual property management requirements and suggested double DOI usage as a best practice for data citation, following RDA's (Research Data Alliance<sup>3</sup>) recommendations.

## 2.6 TF6 -ENVRI-hub design and architecture

The TF6 coordinated the works on the ENVRI-hub, a set of components designed to expose the data and the knowledge produced by the ENVRI-FAIR project. The architecture of the ENVRI-hub was designed to be modular and decentralized, and it was designed by starting by requirements exposed by the RIs in a set of use-cases. Thus, the TF6 produced a set of documents describing the ENVRI-hub architecture, a set of proof of concepts showing how the components of the hub can be implemented, a set of use cases, and a set of data catalogue and software to access them, providing a detailed blueprint for future implementations. WP11 has contributed to the Task Force activities by presenting its requirements and implementation progress, thereby including WP11 RIs data catalogues and their related user stories, and it contributed to the creation of the Soil Water Content use case.

---

<sup>3</sup><https://www.rd-alliance.org/>

**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

In D 11.2 we identified four critical aspects that at this stage of the ENVRI-FAIR project must 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.

In the present document we assess the improvements in these four key areas made by WP11 RIs. The assessment of such improvements has been made leveraging the questionnaires administered to the ENVRI-FAIR participants by the ENVRI TASK FORCES, which consists of a superset of the questions included in the WP11 questionnaire presented in D 11.2.

#### 3.1 Situation as of D 11.2

The questionnaire presented in WP 11.2 highlighted a vastly heterogeneous readiness landscape among WP 11 RIs. Most notably, it was evident a stark contrast between young and prospect RIs such as AnaEE, DANUBIUS-RI, and Dissco, and more established ones, such as ICOS and Lifewatch, with the former group presenting both infrastructural and organizational gaps to fill and the latter providing an overall higher degree of readiness with respect to TF requirements. Another evident distinction in WP11 RIs concerns user identification and authorisation policies, as AnaEE, ICOS, and Lifewatch have policies about user registration and permissions and are already provided with the tools to implement such policies and support the future ENVRI user federation according to TF2 guidelines, while DANUBIUS-RI, Dissco, eLTER, and SIOS adopt a more liberal user policy.

Such an implementation situation is summarized in the following figure.

Table 5: Summary of the Implementation situation

		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

### 3.2 Data catalogue federation readiness

As a major step-up from the situation described in D 11.2, we can now state that most WP11 RIs are ready for ENVRI catalogue federation, with most involved RIs exposing on the Web a catalogue system fully or partially compliant with the requirements expressed so far by TF1.

Table 6. Overview of current Data catalogue federation readiness

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

All RIs opted for Open Source solutions for catalogue management, with GeoNetwork<sup>4</sup> being the most widespread product, except for DANUBIUS-RI which decided to adopt CKAN solution.

### 3.3 User federation readiness

The most notable implementation advancement is the growing adoption of Identity Providers within RIs. WP11 RIs are taking significant steps forward in user policies, and although DANUBIUS-RI and SIOS retain their account-less user management policy, other RIs decided to host an Identity Provider.

**Table 7.** Overview of user federation readiness

	Identity Provider	OpenID connect <sup>5</sup>	SAML 2.0	OAUTH <sup>6</sup> tokens	multi-factor authentication
AnaEE	Yes	Yes	Yes	Yes	Yes
DANUBIUS-RI	No	Not Applicable	Not Applicable	Not Applicable	Not Applicable
DiSSCo	Yes	Yes	Yes	No	No
eLTER RI	Yes	Yes	Yes	No	No
ICOS ERIC	Yes	Yes	Yes	Yes	No
LifeWatch ERIC	Yes	Yes	Yes	Yes	Yes
SIOS	No	Not Applicable	Not Applicable	Not Applicable	Not Applicable

With respect to the implementation situation presented in D11.2, DiSSCo and eLTER RI opted to adopt an identity provider as well as AnaEE, ICOS, and LifeWatch, and the adopted solutions are compliant with the OpenID and SAML protocols, allowing user federation across all WP 11 RIs that manage user identities. The lack of support for OAuth tokens however could pose a risk to the implementation of future user authorization common policies across RIs.

Furthermore, the RIs that use an Identity provider still 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, that do not requires any permission or authentication to access data and services.

### 3.4 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, while six of them have some degree of PID coverage for their data resources. In contrast, no RI has a

<sup>4</sup> <https://geonetwork-opensource.org/>

<sup>5</sup> Open ID (<https://openid.net/>)

<sup>6</sup> Open authorization protocol (<https://oauth.net/>)

data repository certification, hence at the time of writing no WP11 RI has a trusted repository among its services.

**Table 8.** Overview of Technologies and Certifications readiness

	SPARQL endpoint	PID coverage	Repository certification
AnaEE	Yes	Partial	Interested
DANUBIUS-RI	No	None	Not Interested
DiSSCo	No	Partial	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 and, except for AnaEE, they already have a framework agreement with a DOI provider. AnaEE and ICOS currently host data repositories of their own and are both oriented on the CoreTrustSeal<sup>7</sup> certification. We can safely state that all WP11 RIs have received the ENVRI recommendations, and they are actively working on achieving full compliance.

### 3.5 Policies and Licenses

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.

**Table 9.** Overview of Policies and Licenses readiness

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

<sup>7</sup> <https://www.coretrustseal.org/>

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<sup>8</sup>). Also, ICOS has a single eligible licence (CC-BY) for all types of data. All RIs, except for DiSSCo, can account for the machine readability of their eligible licences. Finally, WP11 RIs in general are not publishing data as a Service, except for eLTER and ICOS. The rationale behind this choice consists in the high initial cost of setting up such services; said costs are currently not justified by the demand, as researchers currently prefer accessing data as a discrete asset through catalogue services, which is viable due to the relatively low volume of said data assets. Currently, the development of such services appears to be driven top-down by means of framework initiatives and international projects rather than bottom-up by research community request. The emergence of Big Science and the ever-growing adoption of IoT technologies to monitor experiments however are likely to change this situation, making future data assets unmanageable in a monolithic fashion, and, on the other end, the ever-growing adoption of VRE (Virtual Research Environment) solutions may promote the usage of such services within the research community. WP 11 RIs acknowledge this need and hope that new digital initiatives pushed by the EU in the EOSC context would mitigate the startup costs of new data services.

### 3.6 Implementation Progress Overview

WP 11 RIs have achieved over the past months considerable advancements in all the areas related to data Findability and Accessibility, therein including user base federation. WP11 RIs are laying the foundations of a federated and distributed ecosystem wherein services, products, and users although having a RI of origin can seamlessly communicate and interact with each other. Such an ecosystem is not to be mistaken for a concurrent initiative or a substitute for EOSC in any way, but rather as an additional facility for future EOSC users and developers within the Biodiversity and Ecosystem domain. Most notably, WP11 RIs currently have an online catalogue of data and services (DANUBIUS-RI is going to build it in its Implementation Phase project), and either full or partial API support for catalogue access.

The overall implementation progress appears to be driven by community recommendations, implying that WP 11 RIs have promptly received WP5 Task Forces' specifications wherever they were produced. The co-creation framework established within ENVRI-FAIR appears therefore to be effective in producing guidelines that participating RIs can implement in reasonable time windows. On the other hand, WP11 RIs had at the beginning of the project a considerable gap to fill in terms of IT infrastructure and expertise, a gap that some of them are still working to close.

Task Forces, acting as a distributed regulating body, offered WP11 RIs an adequate venue to bring up their instances to the ENVRI community and to make sure that standard adoption and best practices implementation would be sustainable within their own implementation roadmaps.

---

<sup>8</sup> <https://creativecommons.org/licenses/by-nc/4.0/>

Table 10: Overview on the implementation status

	AnaEE	Danub	DiSSCo	eLTER	ICOS	LW	SIOS
Catalogue Online	Blue	Orange	Blue	Blue	Blue	Blue	Blue
Harvesting support	Blue	Orange	Orange	Blue	Blue	Blue	Blue
DCAT coverage	Blue	Orange	Orange	Blue	Yellow	Blue	Green
API Coverage	Blue	Yellow	Yellow	Blue	Blue	Blue	Blue
OPEN-ID identity provider	Blue	Orange	Blue	Blue	Blue	Blue	Orange
SPARQL endpoint	Blue	Orange	Orange	Orange	Blue	Orange	Orange
PID coverage	Yellow	Orange	Yellow	Yellow	Blue	Yellow	Yellow
Repository certification	Orange	Grey	Grey	Grey	Orange	Grey	Grey
DMP status	Yellow	Blue	Yellow	Yellow	Blue	Yellow	Blue
Suggested data licenses	Yellow	Yellow	Yellow	Orange	Blue	Yellow	Yellow
Machine readable licenses	Yellow	Yellow	Orange	Yellow	Blue	Yellow	Yellow
Data as a Service	Orange	Orange	Orange	Blue	Blue	Orange	Orange



## 4 Roadmap

The FAIR implementation roadmaps were originally presented by the WP11 RIs as Project internal Milestones in June 2020 and were focused on the compliance with all FAIR principles. In D 11.3 a revised roadmap was presented, with more focus on interoperability and with planned actions broken down into four sections to mirror the survey therein presented and summarized in Section 3.3 of the present document.

In this section we will present, for all WP 11 RIs, their roadmap as presented in D 11.3 and the current status of each roadmap item. General Roadmap

### 4.1.1 AnaEE-RI

#### Data Catalogue federation

AnaEE has fulfilled most of its planned roadmap concerning Data Catalogue and it is now ready for catalogue federation. User training and metadata harvesting among its research platforms to populate the online catalogue are ongoing activities that are likely to continue over the next months and probably even after the project's end.



**Table 11.** AnaEE timeline for Data Catalogue federation

D 11.3 Timeline	Action	Sub-Action	Status
2019 (or before)	AnaEE platforms' catalog	Questionnaire drafting for Research Platforms	Completed
2020-Q1	API portal	β-release	Completed
2020-Q2	AnaEE platforms' catalog	β-release	Completed
2020-Q4	API Portal	Training for resource publishers	Completed
2020-Q4	API Portal	Training for users	In progress
2021-Q1	API Portal	Full release	Completed
2021-Q2	Data Catalog	Metadata harvest (starting)	In progress

#### User federation

AnaEE adopts an Identity Provider based on Active Directory technology to federate access to its services. Said Identity Provider is managed by the AnaEE Data and Modelling Centre and it is based on Active Directory technology, which also supports OpenID Connect, as recommended by Task Force 2. Expected deadlines were met with success, and the AnaEE team is now moving on to organize internal user federation within AnaEE platforms.

**Table 12. AnaEE timeline for user federation**

D 11.3 Timeline	Action	Sub-Action	Status
2021-Q1	AAI	federated identity with LDAP <sup>9</sup> protocol	Completed
2021-Q1	AAI	federated authorization with OAUTH2 <sup>10</sup> protocol	Completed

### Technologies and Certifications

The adoption of the DOI system for assigning PIDs to the RI datasets has been decided, however, while some research platforms already have access to DOIs, a framework agreement with a DOI provider at RI level is still missing. Regarding semantics, most scheduled tasks were put on hold as Task Force 4 has provided no authoritative indications on vocabularies and triple stores, hence data asset annotation will be carried out as usual. A triple store with a SPARQL endpoint is deployed at the AnaEE French national node.

**Table 13. AnaEE timeline for Technologies and Certifications**

Timeline	Action	Sub-Action	Status
2021-Q4	Semantic Pipelines	Deployment	In Progress
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	Suppressed
2021-Q4	PID implementation	DOI adoption	In Progress
2022-on	Vocabulary adoption		Suppressed
2020-Q4	SPARQL endpoint	developed for the AnaEE platforms, as an “experience report” in the framework of Task Force 4	Completed

### 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. 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<sup>11</sup>.

<sup>9</sup> Lightweight Directory Access Protocol (<https://ldap.com/>)

<sup>10</sup> Industrial standard-protocol for authorization (<https://oauth.net/2/>)

<sup>11</sup> for Creative Commons – Not rights reserved (CC-0) and Public Domain Mark (CC-PDM), see <https://creativecommons.org/publicdomain/>; for Open Data Commons Public Domain Dedication and License (ODC-PDDL), see <https://opendatacommons.org/licenses/pddl/1-0/>

**Table 14.** AnaEE timeline for Policies and licences

D 11.3 Timeline	Action	Sub-Action	Status
2019 (or before)	Licences	Adoption of CC-BY	Completed
2020-Q2	DMP and Data Policies	1 <sup>st</sup> version	Completed

#### 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 funded project. Originally DANUBIUS-RI planned to outline an internal data and services landscape by means of structured questionnaires in early 2021, however individual platforms requirements and constrains proved to be too diverse to fit a standardized questionnaire; DANUBIUS has therefore moved to a more agile and informal process, using free-form interviews with adhering platforms and data managers to achieve a detailed landscape analysis. Currently DANUBIUS-RI has moved on to a Use Case analysis phase, complemented by a market analysis to identify adequate tools to fit its requirements. This phase will eventually lead to the development of a DANUBIUS-RI data catalogue. In October 2022 the poorly funded Implementation Phase project DANUBIS-IP was started. This will be the framework for the further steps of the FAIRNESS Roadmap.

**Table 15.** DANUBIUS RI timeline for Data Catalogue federation

D 11.3 Timeline	Action	Sub-Action	Status
2020-Q3	Data Catalogue	Creation of a questionnaire to investigate data availability for populating DANUBIUS-RI FAIR data catalogue	Suppressed
2021-Q1	Data Catalogue	Start filling questionnaire for populating the DANUBIUS-RI FAIR data catalogue	Suppressed
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	In Progress
2022-on	Data Catalogue	prototype of data catalogue	Scheduled

##### User federation

DANUBIUS-RI has planned to connect to the EDUcation Global Authentication Infrastructure<sup>12</sup> (eduGAIN) federated authentication service using an LDAP server as backend and OAUTH2 or OpenID connect protocols. Although this activity was originally scheduled for late 2020, it had to be rescheduled to cope with the above-described analysis activities.

<sup>12</sup> <https://edugain.org/>

**Table 16. DANUBIUS-RI timeline for user federation**

D 11.3 Timeline	Action	Sub-Action	Status
2020-Q4	AAI (Authentication, Authorization, and Infrastructure).	Development of the RI central data storage with authentication protocol	Scheduled

#### Technologies and Certifications

DANUBIUS-RI has developed its infrastructural framework, which led to an active contribution to UC1 (Soil Water Content), allowing external researchers to access its resources, and to the establishment of internal data Quality Control procedures. Due to delays in catalogue analysis and development tasks concerning PIDs and user training had to be rescheduled and will be completed after the deployment of the DANUBIUS-RI data catalogue.

**Table 17. DANUBIUS RI timeline for Technologies and Certifications**

D 11.3 Timeline	Action	Sub-Action	Status
2021-Q2	FAIR Training	Organization of DANUBIUS-RI FAIR training sessions – training of trainers	Scheduled
2021-Q2	Prototype service for “Soil Water Content”	Data access across RIs version 1	Completed
2022-on	PID implementation	Test on use cases for PID	Scheduled
2022-on	Data curation: Quality Control (QC), data standardization	Standardized and actionable procedures for QC	Completed

#### Policies and licences

DANUBIUS-RI has successfully implemented its Policies and Licences roadmap, developing a comprehensive DMP compliant with Task Forces’ recommendations.

**Table 18. DANUBIUS RI timeline for Policies and licences**

D 11.3 Timeline	Action	Sub-Action	Status
2020-Q3	DMP and data policies	State of the art Management Plan within ENVRI-FAIR and outside: <ul style="list-style-type: none"> <li>– best practices (e.g., US Geological Survey, ...)</li> <li>– to define the first list of needs</li> </ul>	Completed
2021-Q2	DMP and data policies	Minimum conditions of the Data Management Plan - first draft	Completed
2022-on	DMP and data policies	Minimum conditions of the Data Management Plan - final draft	Completed

### 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

DiSSCo is currently following the roadmap presented in D 11.3 with success, and although it is not ready yet for catalogue federation, the necessary actions are ongoing and begun on schedule. Even though FAIR Digital Object specification is still under development, DiSSCo is making incremental progress towards Digital Specimen catalogue development.

**Table 19. DiSSCo timeline for Data Catalogue federation**

D 11.3 Timeline	Action	Sub-Action	Status
2020-Q3	Create a data catalogue for Digital Specimens.	Test data type registries and experiment with DiSSCo data types	Suppressed
2022 (or after)	Integration with EOSC service catalogue	Create detailed descriptions of various DiSSCo e-services.	In Progress
2022 (or after)	A generalized set of API specifications	API compliance with the Digital Object Interface Protocol (DOIP) specification requirements published by the DONA <sup>13</sup> Foundation.	In Progress

#### User federation

DiSSCo is following the roadmap presented in D 11.3 and is currently working to establish an Identity Provider compliant to Task Force 2 recommendations.

**Table 20. DiSSCo timeline for user federation**

D 11.3 Timeline	Action	Sub-Action	Status
2022 (or after)	AAI	Piloting access through an AAI infrastructure (based on AARC blueprint, also support single sign-on and ORCID integration)	In progress

#### Technologies and Certifications

DiSSCo is following the planned roadmap with minor delays. As many other RIs, DiSSCo acknowledges DOI as the de facto standard for PIDs and is already using such identifiers to improve the findability of its assets.

Notably, DiSSCo has successfully completed ahead of schedule the development of a Museum Collection Manager tool that will facilitate their internal repository management, such an activity has produced a publication titled "DiSSCo Prepare report D6.1 Harmonization and migration plan for the integration of CMSs into the coherent DiSSCo Research Infrastructure".

<sup>13</sup> <https://www.dona.net/>

**Table 21. DiSSCo timeline for Technologies and Certifications**

D 11.3 Timeline	Action	Sub-Action	Status
2021-Q1	PID implementation	Designing a PID schema	Completed
2021-Q1	Data Standard Compilation	First version	In progress
2021-Q2	Prepare a data modelling framework	Specifications of the Digital Specimen Objects	In progress
2021-Q2	Semantics	A best practice guide for semantic enhancement	Completed
2022 (or after)	Data infrastructure	Evaluation of FAIR Digital Object technical infrastructure	In progress
2022 (or after)	Repository	Museum collection Management systems interoperability investigation	Completed

Policies and licences

DiSSCo has successfully established a licensing policy for its data assets, however the DMP development has been put on hold as its periodic revision process has not been established yet.

**Table 22. DiSSCo timeline for Technologies and Certifications**

D 11.3 Timeline	Action	Sub-Action	Status
2019 (or before)	DMP	Periodic evaluation of the provisional version	Suppressed
2020-Q4	Software licences	Recommendations/guidelines for DiSSCo members	Completed

#### 4.1.4 eLTER-RI

##### Data Catalogue federation

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,<sup>14</sup> 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<sup>15</sup>).



**Table 23. eLTER timeline for Data Catalogue federation**

D 11.3 Timeline	Action	Sub-Action	Status
2021-Q2	Create a data catalogue for eLTER Core Data	Create Digital Asset Registry as central catalogue to document and host eLTER Core Datasets	In Progress
2021-Q4	Metadata interoperability	Map eLTER sites dataset metadata on related eLTER datasets, with basic information for discovery	In Progress
2020-Q3	Data interoperability	Update of the eLTER Data Specification for data submission to the network, including method description and reference lists	In Progress
2022	Data interoperability	Agreement on community standards for eLTER Core Data	In Progress
2021-Q2	Data interoperability	Update of the site and platform documentation to implement the core fields for site description	Completed

##### User federation

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.

##### Technologies and Certifications

Currently, PID (e.g., DOIs) are issued using EUDAT B2SHARE<sup>16</sup> 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<sup>17</sup>) based on the RDA I-Adopt<sup>18</sup> working group's recommendations for the description of environmental observations. Provenance will build on the developments and experiences of the ENVRIplus project<sup>19</sup> and implemented to track workflows done in the DataLab which is currently under development.

<sup>14</sup> <https://www.lter-europe.net/projects/PLUS>

<sup>15</sup> <https://catalogue.nextgeoss.eu/>

<sup>16</sup> 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>)

<sup>17</sup> <https://github.com/LTER-Europe/EnvThes>

<sup>18</sup> RDA - Interoperable Descriptions of Observable Property Terminology (I-Adopt), <https://github.com/i-adopt>

<sup>19</sup> <https://www.envriplus.eu/>

**Table 24. eLTER timeline for Technologies and Certifications**

D 11.3 Timeline	Action	Sub-Action	Status
2021-Q4	PID implementation	Design and concept to apply PIDs for sites and platforms documentation in DEIMS-SDR	Scheduled
2021-Q4 (or after)	PID implementation	Design PID schema for eLTER Core Data	Scheduled
2021-Q4 (and after)	Data repository	Extend eLTER Central Data Node to other data types (including spatial data)	Scheduled
2021-Q2	Semantics	Concept for applying I-Adopt recommendations for EnvThes to extend the use as semantic backbone for the eLTER Information System	Scheduled
2022 (and after)	Semantics	Implement semantic annotation and mapping for variable description	In Progress
2021-Q4	Data provenance	First version of metadata needed to ensure data provenance and workflow descriptions (e.g., provenance templating service)	In Progress
2022	Data provenance	Implementation of provenance and integration to metadata for selected workflows from DataLabs	In Progress

#### 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<sup>20</sup> 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 25. eLTER timeline for Policies and licences**

D 11.3 Timeline	Action	Sub-Action	Status
2023 (or before)	Data licence	Selection of appropriate licence (e.g., CC-BY-NC) for eLTER Core Data	In Progress
2023	DMP and Data Policy	First draft of the DMP and Data Policy for the emerging eLTER RI	In Progress

<sup>20</sup> <https://www.lter-europe.net/projects/PPP>



#### 4.1.5 ICOS ERIC



##### Data Catalogue federation

ICOS has experienced delays in its implementation roadmap concerning data and metadata interoperability, however the integration process is ongoing, and the RI as a whole is compliant with catalogue federation requirements suggested by Task Force 1. In particular the netCDF format preparation required a number of coordination actions with other global networks and the process is not yet completed. A first version for the netCDF data format has been prepared and delivered for verification to the modelling community but it is not ready to be shared as official product.

**Table 26. ICOS ERIC timeline for Data Catalogue federation**

D 11.3 Timeline	Action	Sub-Action	Status
2020-Q3	Metadata interoperability	First version of the metadata for translation, with basic information	Completed
2020-Q4	Data interoperability	First version of the data for netcdf preparation	Completed
2021-Q2	Metadata interoperability	Second version of the metadata for translation, with more detailed information	Completed
2021-Q3	Data interoperability	Agreement on community standard using netcdf	In Progress
2022 (or after)	Metadata interoperability	Third version of the metadata for translation with complete information	In Progress
2022 (or after)	Data interoperability	Second version of the data for netcdf preparation	In Progress

##### User federation

ICOS allows users to access its data products anonymously, hence no user federation is required to access ICOS data products; this policy is unlikely to change in the foreseeable future. Although not planned in the roadmap presented in D 11.3, ICOS has set up an Identity Provider compliant with Task Force 2 recommendations to allow its users and personnel to access other assets from the ENVRI community.

##### Technologies and Certifications

ICOS is successfully carrying out its development of metadata standardization tools, with a first version of the proposed tool already available to its users and a second iteration under development. Activities concerning data provenance, however, have been put on hold, and though remaining within the scope of ICOS' roadmap, they and have been rescheduled.

**Table 27.** ICOS ERIC timeline for Technologies and Certifications

D 11.3 Timeline	Action	Sub-Action	Status
2021-Q1	Data provenance	First version of metadata needed to ensure provenance and workflow description - selected variables	Suppressed
2021-Q2	Metadata collection	First version of the tool for standard collection of complex metadata	Completed
2021-Q3	Data provenance	Second version of metadata needed to ensure provenance and workflow description - explicit association with data	In Progress
2021-Q4	Metadata collection	Second version of the tool for standard collection of complex metadata	In Progress
2022 (or after)	Data provenance	Third version of metadata needed to ensure provenance and workflow description - all variables	Scheduled

Policies and licences

ICOS has no intention of modifying its policies in the foreseeable future.

#### 4.1.6 LifeWatch ERIC

##### Data Catalogue federation

LifeWatch ERIC has successfully implemented its Data Catalogue development roadmap with minor delays concerning catalogue integration with the LifeWatch Virtual Research Environment.



**Table 28.** LifeWatch ERIC timeline for Data Catalogue federation

D 11.3 Timeline	Action	Sub-Action	Status
2019 (or before)	LifeWatch metadata catalogues	ERIC (data) Analysis of the state of the art of the different tools and framework available on the market	Completed
2019 (or before)	LifeWatch metadata catalogues	ERIC (data) EcoPortal Publication	Completed
2019	ENVRI-FAIR Training Catalogue	Development and deployment of the training catalogue	Completed
2020-Q1	Metadata Catalogue	β-release based on GeoNetwork	Completed
2020-Q2	Exposer/Connector	First prototype based on Metacat <sup>21</sup> and EML 2.1	Completed
2020-Q3	Metadata Catalogue	Develop EML 2.2.0 profile and Candidate release based on GeoNetwork	Completed
2020-Q3	ENVRI-FAIR Training Catalogue	Updates and second release of the training catalogue	Completed
2020-Q3	EcoPortal	Updates and second release of EcoPortal	Completed
2020-Q3	Exposer/Connector	Second prototype based on Metacat and EML 2.2.0	Completed
2020-Q3	Metadata Catalogue	GeoNetwork Metacat harvesting	Completed
2020-Q3	Metadata Catalogue	Integration in the Virtual Research Environment (VRE)	In Progress
2020-Q4	Metadata Catalogue	User Training	Completed
2021-Q1	Metadata Catalogue	Full release	Completed
2021-Q2	Exposer/Connector	Full release	Completed

<sup>21</sup> <https://github.com/NCEAS/metacat>

#### User federation

LifeWatch ERIC already had an adequate Identity Provider, i.e. compliant with Task Force 2 recommendations before the beginning of WP11's implementation activities, hence no actions were planned in the framework of AAAI implementation. Anyway, the LifeWatch ERIC IT team is working on the LifeWatch ERIC IDP to improve it and allow further integration with other IDPs than the ones hosted by ENVRI community members.

#### Technologies and Certifications

LifeWatch ERIC has successfully implemented its roadmap on schedule except for the aforementioned Virtual Research Environment integrations, which are still ongoing activities.

**Table 29. LifeWatch ERIC timeline for Technologies and Certifications**

D 11.3 Timeline	Action	Sub-Action	Status
2020-Q2	DataCite integration into LifeWatch ERIC Metadata Catalogue	GeoNetwork - DataCite connection to provide DOIs	Completed
2020-Q2	DataCite integration into EcoPortal	EcoPortal - DataCite connection to provide DOIs	Completed
2020-Q4	Provenance Management (LifeBlock)	Identification of the provenance aspects in the VRE	In Progress
2021-Q4	Provenance Management (LifeBlock, Tesseract)	Design of the integration of the provenance aspect in the VRE	In Progress

#### Policies and licences

LifeWatch ERIC has experienced substantial delays in the development of its DMP, as the integration of the various national nodes data management systems has proven a challenging task, however all of LifeWatch's national nodes acknowledge how liberal licensing is a fundamental FAIRness requirement and CC-BY has therefore been adopted at ERIC level as preferred license for data assets.

**Table 30. LifeWatch ERIC timeline for Policies and licences**

D 11.3 Timeline	Action	Sub-Action	Status
2019 (or before)	Licences	Adoption of CC-BY	Completed
2020-Q2	DMP and Data Policies	1 <sup>st</sup> Draft	In progress

#### 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 170,000 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)<sup>22</sup>. The GUI<sup>23</sup> includes a new WMS<sup>24</sup> client based on OpenLayers (v6)<sup>25</sup> 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<sup>26</sup>. A new version of the search interface has been implemented to ensure more dynamical filtering of datasets. For the machine-readable catalogue, SIOS has implemented a CSW endpoint that exposes metadata records through standard protocols and different metadata formats.



A vocabulary server is now available, and it is hosting vocabularies used for the internal metadata standard, including relevant mapping to other vocabularies. The SIOS data access portal has been upgrade to resolve controlled vocabularies directly on this vocabulary server. Additionally, CF standard names, when provided, are resolved to the NERC vocabulary server.

**Table 31. SIOS timeline for Data Catalogue federation**

D 11.3 Timeline	Action	Sub-Action	Status
2021-Q1	<del>Improving SIOS Data Management System (SDMS) tools enabling FAIR data</del>	<del>Adaptation of the Rosetta<sup>27</sup> implementation converting unstructured data from CSV to NetCDF/CF<sup>28</sup></del>	Adaptation not "compatible" with Biodiversity DataSet
2021-Q2	Improving SDMS tools enabling FAIR data	Improving an Excel template generator for registration of data.	In Progress
2021-Q2	Case study 3	Improving and updating the SIOS Observation Facility Catalogue according to the standards from the WIGOS <sup>29</sup> .	In Progress
2021-Q4	CSW endpoint	A CSW endpoint will be exposed including metadata records which are findable on the SIOS data access	Completed

<sup>22</sup> [https://sios-svalbard.org/metadata\\_search](https://sios-svalbard.org/metadata_search)

<sup>23</sup> Graphical User Interface

<sup>24</sup> Web Map service

<sup>25</sup> <https://openlayers.org/>

<sup>26</sup> <https://bokeh.org/>

<sup>27</sup> <http://tomcat.nerc.no/rosetta/>

<sup>28</sup> CF (Climate and Forecast) metadata (<https://cfconventions.org/>)

<sup>29</sup> WMO Integrated Global Observing System (WIGOS) metadata standard. (World Meteorological Organization, 2019)

### User federation

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<sup>30</sup>, i.e., to perform actions on multiple products, access specific internal pages. SIOS has activated on the webportal a module which provides a client implementation for the OpenID Connect protocol. Identification supports google, github, microsoft, orcid and feide which is member of eduGAIN.

**Table 32. SIOS timeline for User federation**

D 11.3 Timeline	Action	Sub-Action	Status
2021-Q4	evaluation of eduGAIN	Possible integration of eduGAIN authentication will be evaluated to provide login to the SIOS portal	Completed

### Technologies and Certifications

SIOS has implemented pyCSW for machine-to-machine interfaces to metadata. The catalogue is now available and exposes all data indexed in the SIOS Data Portal. It provides OGC CSW, OAI-PMH, OpenSearch protocols as well as different metadata schemas, such as ISO 19115 and DIF. 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<sup>31</sup>

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.

<sup>30</sup>See <https://owwl.org/help/my-lists>

<sup>31</sup> <https://github.com/SIOS-Svalbard/SDMSInteroperabilityGuidelines>, see also the [Technical documentation \(sios-svalbard.org\)](https://sios-svalbard.org)

**Table 33. SIOS timeline for Technologies and Certifications**

D 11.3 Timeline	Action	Sub-Action	Status
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	In Progress
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	In Progress
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)	Completed
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	In Progress
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	In Progress

#### Policies and licences

SIOS has revised its policy documents within the planned schedule.

SIOS is now compliant with TF5 recommendations (through SPDX<sup>32</sup>, which helps machine-readability) and able 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<sup>33</sup>.

**Table 34. SIOS timeline for Policies and licences**

D 11.3 Timeline	Action	Sub-Action	Status
2021-Q4	URL for licence	Licences will be exposed in a linked-data approach using SPDX URLs	Completed

<sup>32</sup> Software Package Data Exchange

<sup>33</sup> <https://github.com/metno/mmd/blob/master/doc/mmd-specification.pdf>

## 4.2 User-oriented cross-RIs demonstration cases

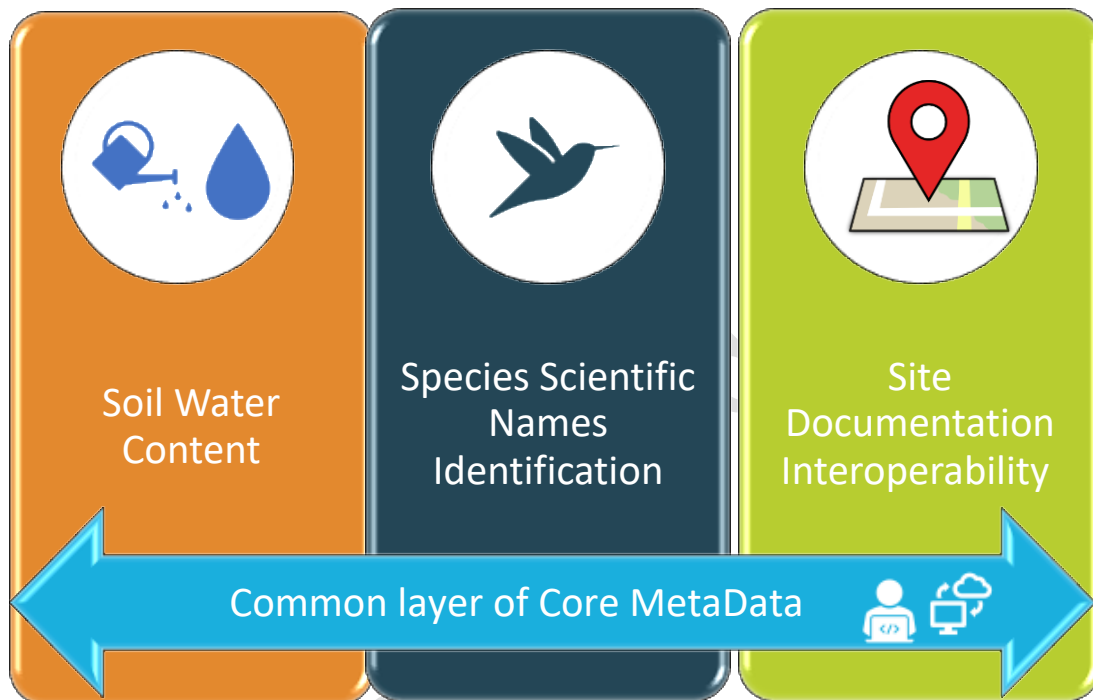
Due to the high heterogeneity of the WP11 RIs, the definition and implementation of common use cases are key to developing interoperability within the subdomain. Therefore, the FAIRness implementation will focus on three subdomain topics defined as showcases:

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 35.** 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	<b>AnaEE</b> DANUBIUS-RI eLTER ICOS LifeWatch SIOS	AnaEE <b>DiSSCo</b> eLTER ICOS LifeWatch	<b>AnaEE</b> <b>eLTER</b> ICOS LifeWatch SIOS	<b>AnaEE</b> DANUBIUS-RI DiSSCo eLTER ICOS LifeWatch SIOS

The UCs are deeply involved in the Task Forces' activities<sup>34</sup> as a cross-feeding process, as shown in Table 34.

**Table 36.** 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	✓		✓	

#### 4.2.1 Soil Water Content (SWC)<sup>35</sup>

##### Context Overview

As introduced in D 11.2 and D 11.3, Soil water content (SWC) is a relevant environmental variable to broad number of processes and task (Bauer-Marschallinger et al., 2019; Shi et al., 2006) as the vegetation dynamic, the biogeochemical cycles, water balance, soil physical properties. However, SWC presents very strong spatial variations linked to the heterogeneity of the soil, the vegetation and climatic conditions as well as the relief which can contribute to water redistribution by runoff. This variability is also temporal, linked in particular to the temporal dynamics of plant cover and climate. In situ

<sup>34</sup> The presentation given during the TFs workshop (Nov, 25<sup>th</sup>-26<sup>th</sup> 2020) is available here: <https://iagos-comm.iek.fz-juelich.de/dmsf/files/5557/view> (ENVRI-FAIR internal resource, available on request)

<sup>35</sup>A living document on the use case activities is available at the following link: <https://docs.google.com/document/d/1pzczbVENA9PvHlyt3oBhODFv-qC8XNPO/edit>

measurement methods are very local, requiring large, expensive and intrusive sampling. Remote sensing offers the possibility of spatialization, but the methods remain limited to a layer of a few centimetres. Modelling therefore remains an essential tool for providing a representation of soil moisture at scales of interest for many applications (hydrology, ecology, agronomy). Models are nevertheless relatively complex and require a large number of contextual variables, such as the local climate, plant cover and its rooting, soil properties and topography. It is therefore important to be able to surround the soil moisture measurements by the acquisition of contextual variables that allow the interpretation of the measurements and feed the models. For this purpose, research infrastructures offer a particularly favourable framework that is complementary to existing networks by producing both quality soil moisture measurements and by describing the context variables. The provision of soil moisture data and context variables in a standardised framework is undeniably a very important service for a wide range of users.

#### Objectives:

The objective of the SWC UC, defined in December 2020, is to prototype a service able to identify relevant datasets based on rich metadata annotation covering environmental characteristics, contextual variable availability and measurements characteristics. This prototype, developed by LifeWatch ERIC, gathers at least one dataset from each participating RI (AnaEE, eLTER, DANUBIUS-RI, ICOS, SIOS), to display how a user-oriented service can be developed on existing RI dataset resources and give feedback to RIs on their metadata scheme, the availability of the metadata and semantic standards used.

#### Original Implementation plan

In D11.3 the following implementation plan was presented:

- Identification of datasets 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)

#### Current Implementation Situation

Use Case implementation is ongoing as described below showing that most of the task are now completed.

- A first step of the use case was to identify users' needs, by defining criteria to identify relevant datasets and useful metadata to document the datasets. A survey was done with about 100 answers. The main foreseen uses are environmental model calibration, data assimilation, remote sensing product calibration, global change studies and environmental monitoring. To use SWC data, the main information expected by the users are: the soil characteristics, the geolocation and the ecosystem type. Concerning the availability of contextual variables, the climate, the soil physical characteristics and ecosystem managements were mentioned as the most important.
- From that survey, a semantic model was proposed to identify and specify the main metadata to be used by the querying tools and the datasets' description.
- The EML 2.2.0 standard was used for metadata discovery and the LifeWatch ERIC Metadata Catalogue was used to collect and publish a dataset from each involved RI and, in order to ensure the interoperability, an XSL Transformation has been developed to map the datasets metadata records from the ISO 19139 standard to the EML 2.2.0 standard. The list of harvested and mapped datasets is available on the LifeWatch ERIC Metadata Catalogue<sup>36</sup>.
- Even if the metadata fields provided by the EML standard sufficiently describe the SWC datasets, they have shown limitation for advanced queries on the existence of contextual variables or on the datasets exploitation metadata and thus a dedicated portal was developed by LifeWatch ERIC for advanced searches. DCAT model and its extension developed in the frame of ENVRI-FAIR was preferred whenever possible and all metadata were gathered in an RDF triple store.

<sup>36</sup> LifeWatch ERIC Metadata Catalogue: list of harvested datasets - [https://metadatalogue.lifewatch.eu/srv/eng/catalog\\_search#/search?facet.q=groupOwner%2F14150%26type%2Fdataset&resultType=details&sortBy=relevance&fast=index&content\\_type=json&from=1&to=20](https://metadatalogue.lifewatch.eu/srv/eng/catalog_search#/search?facet.q=groupOwner%2F14150%26type%2Fdataset&resultType=details&sortBy=relevance&fast=index&content_type=json&from=1&to=20)

- The semantic model is constructed as a mash-up of new and existing semantic entities and relationships. Entities of the graph were initially defined based on the netCDF structure of the AnaEE data series, a number of complex search queries submitted in text by the project participants, and a list of concepts to describe soil data context. The entities were then connected structurally, either with OWL and RDFS properties or with ad hoc relationships, working principally in agreement with the participants, who acted as domain experts, data producers, and data consumers. Eventually, entities were mapped to or replaced by externally defined ones from well-established and recognised semantic artefacts. The reused components of the graph include ontologies, controlled vocabularies for application interoperability and for domain expertise (e.g., the AnaEE thesaurus, DCAT, GeoSPARQL), and the SKOS model for linking domain thesauri.
- The final service is a desktop web application – a dashboard - built with the Angular framework and with a semantic graph database (GraphDB) at its backend and serves as an entry point for (meta)data in the model. Among others, users may search for datasets by type of soil texture or pedological class in the site where data was collected, or they may search using terms from the thesaurus of their preference. Spatial search and dataset location are also enabled, while results are aggregated in tables and histograms, and they are visually rendered on an interactive map.
- The availability of the metadata through machine-to-machine process was analysed and recommendations to data providers to annotate their datasets were also given. These concern the dataset description itself but also the site description that holds part of the useful information. An evaluation of the vocabulary alignment effort was assessed considering both automatic and the remaining manual alignments.

The design and specification phase of all software components required to implement the data harvesting procedures was successfully completed as well and now UC participants are ready to move to the operative phase that will produce the final set of data to expose in the ENVRI Catalogue of Services. The SWC UC completion is expected by the end of the project.

#### 4.2.2 Species Scientific Names Identification (SNI)<sup>37</sup>

##### 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)<sup>38</sup> 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<sup>39</sup> 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

<sup>37</sup> A repository documenting the use case activities available here: <https://zenodo.org/record/7152139>

<sup>38</sup> <https://www.catalogueoflife.org/>

<sup>39</sup> <https://data.catalogue.life/> data repository and API (<https://api.catalogue.life/>)

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.

The main difficulties in this UC were due to the key role of organizations not directly involved in the project. However, the UC activities helped to clarify the needs and identify the gaps where CoL with the support of DiSSCo and LifeWatch is working to propose technical solutions to be implemented.

not yet approved by EC

### 4.2.3 - Site Documentation Interoperability (SDI)<sup>40</sup>

#### 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 9<sup>th</sup>, 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.

#### Original Implementation plan

In D11.3 the following implementation plan was presented:

- 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.

#### Current Implementation Situation

The SDI use case implementation is ongoing. Currently, the use case identifies the most important and common fields for describing sites and link them to existing DCAT classes. Ten DCAT site records were created for each of the five research infrastructures (AnaEE, eLTER, ICOS, and SIOS) and temporarily loaded into a SPARQL endpoint.

At writing date, the implementation progress can be summarized as follows:

- 5 of the 6 documentation sites, highlighted by literature research, are currently analyzed
- A Subset of 10 metadata records per each of the 5 RI is currently included in the UC

---

<sup>40</sup> 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](https://docs.google.com/document/d/10heAQ75WKNh9b6b_nO0pZ-An9z-b0ZOOoYp7icdknr0/edit?usp=sharing)

- 2 metadata conversion services are provided: the ISO to DCAT converter and the CSV to DCAT converter
- The metadata are available and queryable through a SPARQL endpoint, and a set of 5 example queries are supplied

A permanent repository for the above records and products should be considered in the future. Also, a potential work for future ENVRI projects could involve creating an ENVRI-Site ID to uniquely identify sites across different research infrastructures. This would make it easier to combine site records and reduce the number of duplicate records from sites that are listed in multiple site catalogues. The UC work is ongoing, and its implementation activities will continue for all the project's duration.

not yet approved by EC

#### 4.2.4 - Common layer of Core MetaData (CMD)<sup>41</sup>

##### 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.

##### Original Implementation plan

In D11.3 the implementation plan for UC4 was described as a feedback loop with other UCs participants to assist them in adopting the DCAT format and curating adequate metadata for their deliverables. Once a UC had a preliminary data asset to share, UC4 experts would have provided assistance to its DCAT annotation, verifying compliance with DCAT constraints and with the RDF data model in general.

##### Current Implementation Situation

UC4 has curated a brief [handbook](#) for Use Case participants to introduce them to the DCAT-AP vocabulary, its allowed namespaces, and its mandatory classes and properties. UC4 experts have participated so far in TF1 events, including the TF1-Hackaton, and several UC-related meetings, however, as UC1, UC2, and UC3 have not released data assets yet, the initially planned feedback loop has not been established yet. UC4 personnel will attend future TF1 events and assist WP11 RIs in improving their ENVRI catalogue integration.

---

<sup>41</sup> A living document on the use case activities is available at the following link:

<https://docs.google.com/document/d/1SSvbOxPe4GUWyzl5pZcyFsJlxfGhPuzHeO7dKvOCs1A/edit>

## 5 Conclusions

This Deliverable describes the data FAIRness implementation situation within the Biodiversity and Ecosystem subdomain near the end of the ENVRI-FAIR project. Although some additional developments may happen, this document summarizes the FAIRness implementation advancements carried out by WP11 RIs throughout the ENVRI-FAIR project. Over the course of the project, it has been noted on several occasions how the subdomain landscape is very heterogeneous<sup>42</sup>, and the RIs' maturity levels vary greatly, spanning multiple dimensions, including specific scientific domain, internal IT expertise, established services, lifecycle and organization. In such a context, knowledge and expertise transfer become challenging as the many different cultural and methodological differences in the subdomain are concrete risk factors. However, the encouraging situation described in D 11.3 has proven to be a solid foundation for the developments here described: as subdomain RIs appear to have fulfilled most of their implementation plans, albeit with delays, and to have ongoing and planned activities to further improve their data FAIRness.

Notably, the wide adoption of Web APIs based on the RESTful paradigm to provide data and metadata access in a machine-actionable way appears as a major convergence amidst such different organizations. Such a paradigm provides abstraction over data structure and persistence, allowing for the development of data transformation logic by third parties, which represents a breakthrough in interoperability as it allows any organization to develop its own procedures to access and integrate data in a programmatic way. Future implementation efforts should therefore focus on the improvement of such services and the promotion of their usage among researchers and stakeholders. On the other hand, semantic approaches towards data provenance and quality control, although theoretically sound, proved to be problematic to implement, as several actions concerning these topics were re-scheduled or put on hold. The motivations behind these issues include technological, organizational, and often human factors, and appear to be varied and too RI-specific to draw general conclusions. The open discussion format adopted by Task Force 4 allowed RIs to discuss their issues and foster expertise transfer, however, for the purpose of data provenance tracking, semantic technologies may be simply too formal and cumbersome to maintain and update to be successfully implemented at RI and cross-RI level.

On the more methodological side, cross-domain Task Forces established by WP5 played a pivotal role in promoting knowledge transfer, as they offered RIs an adequate venue to voice their requirements, specificities, and gaps both in terms of technology and methodology. Such an open discussion allowed for the production of sustainable recommendations that WP11 RIs have proven to be able to comply with. In this context, the definition and implementation of cross-RIs' use cases have also been instrumental in building a common framework to promote technology and expertise transfer, as they represent non-trivial demonstrations of practical FAIRness and interoperability among the subdomain's RIs. Although Use Cases did not produce their final data assets yet, they are expected to deliver before the end of ENVRI-FAIR, contributing to the development of the ENVRI catalogue.

Quoting D10.4, *FAIRness is a journey*(Jeffery et al., 2020), but after four years of ENVRI-FAIR it appears clear that a long distance has been travelled, and though there may be still much to be done to achieve FAIRness, the work hereby presented suggests how ENVRI-FAIR managed through its co-creation, cross-domain, approach to establish a functional framework for RI development, allowing young RIs to rapidly fill technological and organizational gaps, and more established ones to improve and expand their service portfolio.

---

<sup>42</sup> See Deliverable 11.1 (Papale, 2020),



## 6 Impact on the project

WP11 considers this Deliverable a milestone in the implementation of FAIR principles in the Biodiversity and Ecosystem subdomain and paving the way towards the cooperation among the ENVRI FAIR project WPs and groups. As a follow-up of D 11.3, it provides up-to-date information for other project members in how the seven Research Infrastructures involved in the Biodiversity and Ecosystem subdomain are improving FAIRness. The actual cooperation and testing have been 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) the outcomes of which are detailed 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

We gratefully acknowledge the valuable contributions we received from the research infrastructures' representatives within the Biodiversity and Ecosystem subdomain.

We thank the reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

## 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](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<sup>43</sup>:

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 ENVRI
EnvThes	Environmental Thesaurus
EOSC	European Open Science Cloud
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

<sup>43</sup> The latest version of the master list of the glossary is available at <http://doi.org/10.5281/zenodo.3465753>.

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