D11.6
Assessment of FAIRness in the Biodiversity and Ecosystem subdomain

Work Package	WP11
Lead partner	LifeWatch ERIC
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	30 June 2023
Submission date	30 June 2023

Deliverable abstract

This Deliverable provides an assessment of the level of FAIRness of the seven RIs in the Biodiversity and Ecosystem subdomain. The assessment has been done by taking advantage of the FAIR Implementation Profiles, resulting from the collaborative effort between the GO-FAIR Foundation and ENVRI-FAIR WP5. In this report the outcome for the FIPs compiled by the RIs belonging to WP11 is discussed in more detail. A final chapter is dedicated to the evolution of the FIPs along the four years of the project.

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

DELIVERY SLIP

	Name	Partner Organisation	Date
Main Author	Alberto Basset Andrea Tarallo	LifeWatch ERIC CNR-IRET / LifeWatch ERIC	26 June 2023
Contributing Authors	Francesca De Pascalis Sharif Islam J. Peterseil Christoph Wohner Dario Papale Rudolf Denkmann Dario De Nart Lucia Vaira	DANUBIUS-RI DiSSCo eLTER eLTER CMCC (ICOS) SIOS CREA (AnaEE) LifeWatch ERIC	
Reviewer(s)	Dario Papale Nicola Fiore	CMCC / ICOS LifeWatch ERIC	26 June 2023
Approver	Andreas Petzold	FZJ	30 June 2023

DELIVERY LOG

Issue	Date	Comment	Author
V 1.1	21 June 2023	Sent for review	Andrea Tarallo
V 2.0	27 June 2023	Sent for approval	Andrea Tarallo
V 2.0	28 June 2023	Sent for submission	Andrea Tarallo

DOCUMENT AMENDMENT PROCEDURE

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

GLOSSARY

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

D11.6 – Assessment of FAIRness in the Biodiversity and Ecosystem subdomain ... 4
1 Introduction ... 4
2 Findability ... 4
2.1 F1. What globally unique, persistent, resolvable identifiers do you use for (meta-)data records? 5
2.2 F2. Which metadata schemas do you use for findability? ... 6
2.3 F3. What is the technology that links the persistent identifiers of your data to the metadata description? ... 7
2.4 F4. In which search engines are your (meta)data indexed? ... 8
3 Accessibility ... 9
3.1 A1.1. Which standardised communication protocol do you use for (meta)data? ... 10
3.2 A1.2 Which authentication & authorisation technique do you use for (meta)data? ... 11
3.3 A2 Which metadata longevity plan do you use? ... 11
4 Interoperability ... 12
4.1 I1. Which knowledge representation languages (allowing machine interoperation) do you use for (meta)data? ... 12
4.2 I2. Which structured vocabularies do you use to annotate your metadata and encode your data? 13
4.3 I3. Which models, schema(s) do you use for your (meta)data? ... 14
5 Reusability ... 15
5.1 R1.1. Which usage license do you use for your (meta)data records? ... 15
5.2 R1.2. Which metadata schemas do you use for describing the provenance of your (meta)data? 16
6 Discussion ... 16
7 Acknowledgements ... 19
8 References ... 19
9 Appendices ... 19
9.1 Appendix A: Glossary and terminology ... 19

D11.6 – Assessment of FAIRness in the Biodiversity and Ecosystem subdomain

1 Introduction

The ENVRI-FAIR project's objective is to implement "FAIRness" for data produced in the European Research Infrastructures (RIs) in the Environmental Domani (ENVRI) connecting the RIs to the European Open Science Cloud (EOSC). In this context, "FAIR" is an acronym comprising the aspects of "Findable", "Accessible", "Interoperable", and "Reusable" as specified by the FORCE11 community. This report is the assessment of the level of FAIRness of the seven RIs in the Biodiversity and Ecosystem subdomain, i.e., ANAEE ERIC, Danubius-RI, DiSSCo, eLTER RI, ICOS ERIC, LifeWatch ERIC, SIOS. An assessment of FAIRness has been done for all the project participating RIs on a yearly basis by WP5 by means of FAIR Implementation Profile (FIP) assessment. A FIP is a compilation of technology choices, known as FAIR Enabling Resources (FERs), which are selected by a specific community of practice to implement one or more FAIR Guiding Principles. The FIP is the outcome of a collective decision made by the community members. The FIPs, resulting from the collaborative effort between the GO-FAIR Foundation and ENVRI-FAIR WP5, serve as a reflection of the implementation choices made by the different communities. They provide a means to track the progress of FAIR data services and the convergence among these communities. These profiles consist of 21 questions that are posed to the data steward of a community, assessing the FAIRness of their resources (Table 1) (Schultes et al., 2020). When a question is answered with an existing resource (or a resource that is planned to be adopted in the near future), it is considered a FER, indicating whether a FAIR principle is fulfilled or not. Multiple FERs can satisfy the same FAIR principle.
Taking advantage of the existing FIPs, we carried out an analysis of the level of FAIRness and the specific choices of the RIs belonging to the Biodiversity and Ecosystem subdomain.

Table 1: A schematic representation of the questionnaire used to build a FIP. From https://bit.ly/yourFIP

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	*e.g. PURL, DOI*
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	
F4	In which search engines are your metadata records indexed?	Search engines	
F4	In which search engines are your datasets indexed?	Search engines	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	

2 Findability

The Findability principle focuses on ensuring that the research (meta)data are easily discoverable and identifiable. Findability is crucial because if (meta)data cannot be found, it cannot be accessed, utilised, or built upon by others. F1 emphasises the need for research (meta)data to be assigned a globally unique and persistent identifier. This identifier serves as a permanent reference point, ensuring its long-term accessibility and citability. Persistent identifiers, such as DOIs (Digital Object Identifiers) or URLs, are designed to remain unchanged over time, even if the location or storage of the (meta)data changes. Overall, the assignment of persistent unique identifiers to research (meta)data is essential for its long-term findability and accessibility, providing a stable reference point for researchers and enabling the effective sharing and reuse of scientific data.

2.1 F1. What globally unique, persistent, resolvable identifiers do you use for (meta-)data records?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
F1-Metadata	DOI | Digital Object Identifier	0	0	0	3	3	3	0	3
	ePIC | Persistent Identifier Consortium for eResearch	0	0	0	0	3	0	0	1
	Handle System	0	2	3	3	3	0	0	3
	ORCID | Open Researcher and Contributor ID	0	0	0	0	3	0	0	1
	URI | Uniform Resource Identifier	0	0	0	0	3	0	0	1
	UUID | Universally Unique Identifier	3	0	0	3	0	0	3	3
F1-Data	DEIMS.ID	0	0	0	3	0	0	0	1
	DOI | Digital Object Identifier	3	2	0	3	3	3	3	5
	ePIC | Persistent Identifier Consortium for eResearch	0	0	0	0	3	0	0	1
	Handle System	0	0	3	3	3	0	0	3
	UUID | Universally Unique Identifier	0	0	0	0	0	0	3	1

Figure 1: Answers to the question regarding the FAIR principle F1. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

F1 subprinciple clearly states that to be satisfied, both data and the associated metadata are assigned a globally unique and persistent identifier. Despite the relevance of assigning a persistent identifier to metadata is still debated, the RIs of the Biodiversity and Ecosystem subdomain have implemented at least one FER in this regard. The most used persistent identifier for (meta)data is the Digital Object Identifier (DOI). DOI is long-term maintained by a solid and trusted infrastructure, which assures its stability and function over time. In addition, DOIs can be assigned to various types of digital objects, including research articles, datasets, books, images, and more. They are not limited to a specific content type or format, making them versatile for a wide range of scholarly and digital resources.

The Handle System is a comprehensive and extensible infrastructure for assigning, managing, and resolving persistent identifiers called handles. Handles are used to uniquely identify digital resources and enable their long-term accessibility. The Handle System provides a decentralised architecture, allowing multiple independent handle services to operate and collaborate while ensuring the integrity and persistence of the assigned identifiers. Those advantages make the handle the second FER to assign a persistent identifier to (meta)data. Regarding metadata, the Universally Unique Identifier (UUID) is also a common choice among the biodiversity and ecosystem subdomain RIs. UUIDs are commonly used to identify objects, entities, or resources in distributed systems and databases. Differently from DOI and Handle, they must be generated and maintained by the RI using specific algorithms that assure its uniqueness.

2.2 F2. Which metadata schemas do you use for findability?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
F2	CKAN | Comprehensive Knowledge Archive Network	3	0	0	0	0	0	0	1
	DataCite Metadata Scheme	0	0	0	0	3	0	0	1
	DCAT | Data Catalog Vocabulary Version 2	0	2	0	3	0	0	0	1
	DCAT-AP | Data Catalog Vocabulary Application Profile for Data Portals in Europe	2	0	0	0	3	0	0	1
	DIF | Directory Interchange Format	0	0	0	0	0	0	3	1
	EML | Ecological Metadata Language	2	0	0	0	0	0	0	0
	EML2.2.0 | Ecological Metadata Language 2.2.0	0	0	0	0	0	3	0	1
	ISO 19115 | Geographic information - Metadata	3	2	0	3	0	0	3	3
	ISO 19139 | Geographic information - Metadata - XML schema implementation	0	0	0	0	0	3	0	1
	MET Norway Metadata Format Specification	0	0	0	0	0	0	3	1
	NetCDF CF-1.7	0	2	0	0	0	0	0	0
	openDS | Open Digital Specimen	0	0	2	0	0	0	0	0
	Schema.org Dataset	0	0	0	0	3	0	0	1
	WMO Core Profile | World Meteorological Organization Core Metadata Profile	0	0	0	0	0	0	3	1

Figure 2: Answers to the question regarding the FAIR principle F2. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

The F2 sub-principle focuses on providing rich metadata and detailed descriptions to improve the findability of research data and related resources. It emphasises the importance of including accurate and comprehensive metadata that enable effective data discovery and retrieval. In particular, the FIP requires to list the metadata schema(s) implemented by the RIs. The RIs that deal with geospatial information implemented the use of the metadata schema ISO 19115, an international standard that provides a framework for describing geospatial data and services. The ISO 19115 schema covers various aspects of metadata, including identification, spatial and temporal extent, quality, lineage, distribution, and access information. It allows users to document key characteristics of geospatial data, such as its purpose, content, and usage constraints, facilitating effective discovery, understanding, and evaluation of the data. For the other RIs. what appears is that there is not a clear convergence toward a single implementation choice. Rather, each RI uses a different schema, and most of them more than one schema. This might be due to the fact that some RI manage more than one digital object, requiring thus different schemas to manage them. Implementing data interoperability among RIs that use different metadata schemas can pose several challenges and issues, including:

1. Semantic Heterogeneity: Different metadata schemas may use different vocabularies, concepts, and data models, leading to semantic heterogeneity. This makes it challenging to map and align the meaning and structure of metadata elements across different schemas, hindering effective interoperability.
2. Structural Differences: Metadata schemas can vary in their structure and organisation. Some may have different levels of granularity, different ways of representing relationships, or variations in the depth of metadata coverage. These structural differences make it difficult to align and integrate metadata seamlessly.
3. Mapping Complexity: Mapping metadata elements from one schema to another can be complex, especially when the schemas have divergent structures and semantics. The process of mapping involves identifying equivalent elements, resolving differences in data models, and handling mismatches in data types and formats. This mapping complexity adds overhead and requires careful attention to ensure accurate and meaningful interoperability.
4. Maintenance and Evolution: Metadata schemas evolve over time as new requirements and standards emerge. Managing interoperability between infrastructures with different metadata schemas requires ongoing maintenance to accommodate schema updates, additions, and changes. It requires continuous effort to keep the mappings up to date and ensure compatibility as the metadata schemas evolve.
5. Governance and Consistency: Interoperability requires governance mechanisms and agreements among infrastructure providers to establish common standards and best practices

for metadata representation. Ensuring consistency in metadata implementation and adherence to interoperability guidelines can be a challenge, particularly when multiple organisations or domains are involved.

Addressing these issues requires collaborative efforts, the use of standardised interoperability frameworks and protocols, development of common vocabularies and ontologies, and continuous coordination among infrastructure providers. It also calls for the adoption of data integration approaches and technologies that enable harmonisation and alignment of metadata across different schemas.

2.3 F3. What is the technology that links the persistent identifiers of your data to the metadata description?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
F3	DataCite | DataCite Ontology	3	0	0	0	0	3	0	2
	DOI | Digital Object Identifier	0	2	0	3	0	0	0	1
	FDO | Fair Digital Object	0	0	3	0	0	0	0	1
	LOD | Linked Open Data	0	0	0	0	3	0	0	1

Figure 3: Answers to the question regarding the FAIR principle F3. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

The F3 sub-principle emphasises the importance of ensuring that metadata and data are linked one to another. The association between a metadata file and the described data should be made explicit by mentioning a dataset's globally unique and persistent identifier in the metadata. Three out of five subdomain RIs use a DOI-related technology to link the metadata to the data. This was expected, giving that the DOI is the most used persistent identifier in the subdomain. The FDO and the LOD are more frameworks than technological choices, nonetheless fundamental for the development of a FAIR environment of digital resources.

2.4 F4. In which search engines are your (meta)data indexed?

<table>
<tr><th>FIP question</th><th>FAIR Enabling Resources</th><th>AnaEE-ERIC</th><th>Danubius-RI</th><th>DiSSCo</th><th>eLTER-RI</th><th>ICOS-ERIC</th><th>LifeWatch ERIC</th><th>SIOS</th><th>Total</th></tr>
<tr><td rowspan="19">F4-Metadata</td><td>B2FIND</td><td>0</td><td>0</td><td>0</td><td>3</td><td>3</td><td>0</td><td>0</td><td>2</td></tr>
<tr><td>DataCite | DataCite Ontology</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>3</td><td>0</td><td>2</td></tr>
<tr><td>DEIMS-SDR | Dynamic Ecological Information Management System - Si</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>Dimensions</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>ECOI | DiSSCo European Collection Objects Index</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>EcoPortal</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>1</td></tr>
<tr><td>eLTER Digital Asset Registry</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>GBIF search engine | Global Biodiversity Information Facility Search En</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>0</td><td>3</td><td>2</td></tr>
<tr><td>GeoNetwork</td><td>3</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>GEOSS Portal | Global Earth Observation System of Systems Portal</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td></tr>
<tr><td>Google Dataset Search</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>Google Search</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>1</td></tr>
<tr><td>ICOS Carbon Portal | Integrated Carbon Observation System Data Porta</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>LifeWatch ERIC Metadata Catalogue</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>1</td></tr>
<tr><td>OpenAIRE Research Graph</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>Sios Metadata Search</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>1</td></tr>
<tr><td>Svalbard Integrated Arctic Earth Observing System CSW Catalogue</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>1</td></tr>
<tr><td>WIGOS | WMO Integrated Global Observing System</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>WMO Search | World Meteorological Organization Search</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td></tr>
<tr><td rowspan="18">F4-Data</td><td>Danubius Demonstrator Portal</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>DataCite | DataCite Ontology</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>1</td></tr>
<tr><td>Dimensions</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>ECOI | DiSSCo European Collection Objects Index</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>eLTER Digital Asset Registry</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>eLTER DIP | eLTER Data Integration</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>eLTER-CDN | eLTER Central Data Node</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>ERDDAP</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>GEOSS Portal | Global Earth Observation System of Systems Portal</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td></tr>
<tr><td>Google Dataset Search</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>Google Search</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>1</td></tr>
<tr><td>ICOS Carbon Portal | Integrated Carbon Observation System Data Porta</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>LifeWatch ERIC Metadata Catalogue</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>1</td></tr>
<tr><td>re3data | Registry of Research Data Repositories</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>Sios Metadata Search</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>1</td></tr>
<tr><td>Svalbard Integrated Arctic Earth Observing System CSW Catalogue</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>1</td></tr>
<tr><td>WIGOS | WMO Integrated Global Observing System</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>WMO Search | World Meteorological Organization Search</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>2</td><td>1</td></tr>
</table>

Figure 4: Answers to the question regarding the FAIR principle F4. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

The FER listed for this question are 19 for metadata and 18 for data. This is mainly due to the fact that each RI has developed their own data portal and metadata catalogues. Further, single RIs resources are also indexed in additional generalist indexing engines, for instance Google. There are several reasons why RIs did not converge towards a single or few portals. These include:

Domain-specific Needs: RIs often serve specific scientific domains or communities with unique data requirements. Building dedicated portals allows them to design and customise features, functionalities, and interfaces that align with the specific needs and workflows of their user communities. This level of specialisation ensures that researchers can access and interact with data in ways that are most relevant and useful to their particular field of study.

Governance and Control: RIs often operate under their own governance structures, which include policies, regulations, and data management guidelines specific to their organisation or consortium. Building their own portals allows them to enforce these governance rules, ensure compliance with legal and ethical standards, and maintain control over access, usage, and data sharing policies. It provides them with the flexibility to define their own data management practices and make decisions based on their specific objectives and priorities.

Community Building and Collaboration: RIs aim to foster collaboration and community engagement among researchers and data users. By developing their own data portals and metadata catalogues, they can create dedicated spaces for community interaction, data sharing, and knowledge exchange. These platforms can facilitate networking, discussions, and collaborations within the research community, promoting a sense of ownership, community building, and fostering closer ties between data producers and data users.

While it may seem advantageous to have a unified data portal or metadata catalogue, the diverse needs, complexities, governance considerations, and community-building aspects make it more practical and effective for RIs to develop their own data portals and metadata catalogues. These dedicated platforms ensure tailored support for domain-specific research, enable specialised data management functionalities, maintain control over governance and policies, and foster collaboration within their respective communities.

3 Accessibility

The Accessibility principle states that data and metadata should be easily and freely accessible through appropriate and widely adopted protocols and standards. It emphasises the importance of removing technical, legal, and financial barriers that could hinder access to (meta)data. To achieve accessibility, data should be made available in a format that is machine-readable and easily understandable by both humans and computers. It recognises that access to data is essential for advancing scientific knowledge, enabling data-driven decision-making, and maximising the value and impact of research outputs.

3.1 A1.1. Which standardised communication protocol do you use for (meta)data?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
A1.1-Metadata	CSW | Catalog Service for the Web	3	0	0	3	3	0	3	4
	DEIMS-SDR-REST-API	0	0	0	3	0	0	0	1
	DOIP | Digital Object Interface Protocol	0	0	3	0	0	0	0	1
	ERDDAP	0	0	0	0	3	0	0	1
	HTTPS | Hypertext Transfer Protocol Secure	3	0	0	0	3	3	0	3
	OAI-PMH Schema | Open Archives Initiative Protocol for Metadata Harvesting Sche	0	0	0	3	2	0	3	2
	OPeNDAP | Open-source Project for a Network Data Access Protocol	0	2	0	0	3	0	0	1
	OpenSearch	0	0	0	0	0	0	3	1
	Python programming library ICOSCP	0	0	0	0	3	0	0	1
	REST | Representational state transfer	0	0	3	3	3	0	0	3
	SPARQL (open) endpoint	0	0	0	0	3	0	0	1
	WFS | Web Feature Service	0	0	0	3	0	0	0	1
	WMS | Web Map Service	0	0	0	3	3	0	0	2
A1.1-Data	CSW | Catalog Service for the Web	0	0	0	0	3	0	0	1
	DOIP | Digital Object Interface Protocol	0	0	3	0	0	0	0	1
	ERDDAP	0	0	0	0	3	0	0	1
	FTP | File Transfer Protocol	0	0	0	0	0	0	3	1
	HTTPS | Hypertext Transfer Protocol Secure	3	0	0	3	3	3	3	5
	OGC CS | Open Geospatial Consortium Catalogue Services 3.0	0	2	0	0	0	0	0	0
	OPeNDAP | Open-source Project for a Network Data Access Protocol	0	0	0	0	3	0	3	2
	REST | Representational state transfer	0	0	3	0	0	0	0	1
	SOS | Sensor Observation Service	0	0	0	3	0	0	0	1
	WCS | Web Coverage Service	0	0	0	0	3	0	0	1
	WFS | Web Feature Service	0	0	0	2	0	0	0	0
	WMS | Web Map Service	0	0	0	3	0	0	3	2

Figure 5: Answers to the question regarding the FAIR principle A1.1. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

Clearly, one of the most utilised communication protocol for (meta)data is HTTPS (Hypertext Transfer Protocol Secure). HTTPS is a secure communication protocol widely used for transferring data over the internet, constituting the world standard for web transfer protocols. It is an extension of the standard HTTP protocol that adds encryption and authentication mechanisms to ensure secure communication between a web browser and a web server. The capabilities of HTTPS include data encryption, which prevents eavesdropping and tampering, and server authentication, which verifies the identity of the server, ensuring that users are connecting to the intended website. The advantages of using HTTPS include data privacy and protection, and it also helps to establish trust with website visitors, as the presence of a valid SSL/TLS certificate indicates a secure and reliable connection.

The RIs in the subdomain which mainly deal with geospatial data converge toward the use of the Catalogue Service for the Web (CSW). This is a system that enables the discovery, access, and retrieval of geospatial metadata and data resources. CSW supports the use of standardised metadata schemas, such as ISO 19115, that is indeed one of the convergent choices of the same involved RIs (AnaEE-ERIC, eLTER-RI and SIOS - see Paragraph F2 above).

3.2 A1.2 Which authentication & authorisation technique do you use for (meta)data?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
A1.2-Metadata	DAAI | DiSSCo Federated Authentication and Authorization Infrastructure	0	0	1	0	0	0	0	0
	eduGAIN Interfederation Service	0	2	0	0	3	0	0	1
	GeoNetwork account in the internal catalog database	3	0	0	0	0	0	0	1
	ICOS local account	0	0	0	0	3	0	0	1
	Keycloak Single-Sign On	0	0	0	0	0	2	0	0
	LDAP | Lightweight Directory Access Protocol	2	0	0	0	0	0	0	0
	OAuth | Open Authorization	0	0	0	0	3	0	0	1
	Open Data	0	0	0	3	3	0	3	3
	ORCID | Open Researcher and Contributor ID	0	0	0	0	3	0	0	1
	SAML2 | Security Assertion Markup Language 2.0	0	0	0	0	3	0	0	1
A1.2-Data	DAAI | DiSSCo Federated Authentication and Authorization Infrastructure	0	0	1	0	0	0	0	0
	eduGAIN Interfederation Service	0	2	0	0	3	0	0	1
	ICOS local account	0	0	0	0	3	0	0	1
	Keycloak Single-Sign On	0	0	0	0	0	2	0	0
	LDAP | Lightweight Directory Access Protocol	3	0	0	0	0	0	0	1
	Open Data	0	0	3	3	3	0	3	4
	ORCID | Open Researcher and Contributor ID	0	0	0	0	3	0	0	1
	SAML2 | Security Assertion Markup Language 2.0	0	0	0	0	3	0	0	1

Figure 6: Answers to the question regarding the FAIR principle A1.2. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

Even though "Accessibility" does not imply openness of data, from the FIPs of the subdomain it emerges that both data and the associated metadata are often free to be accessed under an Open Data policy, i.e., no authentication & authorisation protocol is required. This might be the result of the "as open as possible, as closed as necessary" policy conducted by the European Commission. In this regard, all the data obtained with public funding are provided as open, unless there is a specific reason to not do so. Overall, offering data in an open manner enhances scientific integrity, fosters collaboration, and supports societal benefit, driving scientific progress and societal advancement. However, in order to actively upload data and make use of services to analyse them, most of the RIs require a registration by means of an access protocol often based on standard existing ones (i.e. ORCID, eduGAIN).

3.3 A2 Which metadata longevity plan do you use?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
A2	AnaEE Metadata Preservation Policy	2	0	0	0	0	0	0	0
	DANUBIUS Preservation Policy	0	2	0	0	0	0	0	0
	DiSSCo Data Management Plan	0	0	2	0	0	0	0	0
	eLTER Data Management Plan Draft	0	0	0	3	0	0	0	1
	SIOS Data Management Plan	0	0	0	0	0	0	2	0
	The ICOS Preservation Policy	0	0	0	0	3	0	0	1

Figure 7: Answers to the question regarding the FAIR principle A2. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

What appears from the responses obtained by question A2 is that the longevity plan for metadata preservation in RIs of Biodiversity and Ecosystem subdomain is mostly to be implemented in the near future. Often, the longevity plan goes hands in hands with the Data Management Plan, that is a document very specific to each organisation. Even though a convergence in this case cannot be achieved, a convergence in the standardised practices and guidelines for metadata management and preservation across different RIs would be desirable. This effort may require additional funding, taking in further consideration that the dynamic nature of RIs, technological advancements, and evolving metadata standards, pose an additional challenge in timely updating such policies. To address these issues, there is a need for coordinated efforts, collaboration, and dedicated resources to prioritise and implement robust metadata preservation strategies among the ENVRI RIs in general.

4 Interoperability

Interoperability emphasises the importance of enabling seamless (meta)data integration and exchange across different systems and platforms. Interoperability ensures that (meta)data can be easily combined, shared, and reused by both humans and machines. It involves using common standards, formats, and protocols for representing and organising data, as well as establishing clear and consistent metadata practices. By adhering to interoperability, RIs enable efficient and effective (meta)data discovery, integration, and analysis. It allows researchers to access and utilise data from multiple sources, facilitating cross-disciplinary collaborations and enhancing the value and impact of research outcomes.

4.1 I1. Which knowledge representation languages (allowing machine interoperation) do you use for (meta)data?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
I1-Metadata	DwC-A | Darwin Core Archive	0	0	3	0	0	0	0	1
	JSON | JavaScript Object Notation	0	0	0	0	0	3	0	1
	JSON Schema | JavaScript Object Notation Schema	0	0	0	3	0	0	0	1
	JSON-LD | JavaScript Object Notation for Linking Data	0	0	0	0	3	0	0	1
	OWL | Web Ontology Language	0	0	0	0	3	0	0	1
	RDFS | Resource Description Framework Schema	0	0	0	0	3	3	0	2
	XMLS | eXtensible Markup Language Schema	3	0	0	3	3	3	3	5
I1-Data	DwC-A | Darwin Core Archive	0	0	0	0	0	3	3	2
	JSON Schema | JavaScript Object Notation Schema	0	0	3	0	0	0	0	1
	JSON-LD | JavaScript Object Notation for Linking Data	0	0	2	0	0	0	0	0
	NetCDF | Network Common Data Form	3	2	0	0	3	3	3	4
	RDFS | Resource Description Framework Schema	0	0	0	0	3	0	0	1
	XMLS | eXtensible Markup Language Schema	0	0	0	3	0	3	0	2

Figure 8: Answers to the question regarding the FAIR principle I1. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

Regarding metadata, different implementation choices are made by the subdomain RIs. Clearly, the eXtensible Markup Language (XML) schema is the most implemented choice in WP11. This is a specification that defines the structure, data types, and constraints of XML documents. Overall, the XML schema is a widely used and established technology for defining and validating the structure of XML documents, enabling interoperability and data consistency across different systems.

For what regard data, NetCDF (Network Common Data Form) is the most implemented choice. This is a data storage and access format specifically designed for scientific data that provides a self-describing and machine-independent way to store multidimensional scientific data, such as climate, atmospheric, and oceanographic data. NetCDF is widely used in scientific disciplines, such as atmospheric science, climate modelling, oceanography, and geophysics, as it enables the storage, sharing, and analysis of complex multidimensional datasets. This makes a very common data format for many disciplines captured by the umbrella of Biodiversity and Ecosystem subdomain.

4.2 I2. Which structured vocabularies do you use to annotate your metadata and encode your data?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
I2-Metadata	ABCD | Access to Biological Collection Data	0	0	3	0	0	0	0	1
	ABCDEFG | Access to Biological Collection Databases Extended for Geosciences	0	0	3	0	0	0	0	1
	ALIENSPECIES | Alien Species Thesaurus	0	0	0	0	0	2	0	0
	anaeeThes | AnaEE Thesaurus	3	0	0	0	0	0	0	1
	DCAT | Data Catalog Vocabulary Version 2	0	0	0	0	0	2	0	0
	eLTER Controlled Lists	0	0	0	2	0	0	0	0
	EML2.2.0 | Ecological Metadata Language 2.2.0	0	0	0	0	0	3	0	1
	EnvThes | Environmental Thesaurus	0	0	0	3	0	0	0	1
	EUNIS Habitat Classification	0	0	0	0	0	2	0	0
	FISHTRAITS | Fish Traits Thesaurus	0	0	0	0	0	2	0	0
	GCMD Keywords | Global Change Master Directory Keywords	0	0	0	0	0	0	3	1
	GEMET | General Multilingual Environmental Thesaurus	0	0	0	0	0	2	0	0
	GeoDCAT-AP	0	0	0	0	0	2	0	0
	ICOS Ontology | Integrated Carbon Observation System Ontology	0	0	0	0	3	0	0	1
	ISO 3166-1:2013 Codes for the representation of names of countries and their subc	0	0	0	0	0	3	0	1
	MACROALGAETRAITS | Macroalgae Traits Thesaurus	0	0	0	0	0	2	0	0
	MMD Vocabulary	0	0	0	0	0	0	3	1
	NVS | NERC Vocabulary Service	0	2	0	0	0	2	0	0
	PHYTOTRAITS | The Phytoplankton Traits	0	0	0	0	0	2	0	0
I2-Data	ABCDEFG | Access to Biological Collection Databases Extended for Geosciences	0	0	3	0	0	0	0	1
	ALIENSPECIES | Alien Species Thesaurus	0	0	0	0	0	3	0	1
	CF SN | Climate and Forecast Standard Names Parameter Vocabulary	0	0	0	0	0	2	3	1
	DwC-A | Darwin Core Archive	0	0	0	0	0	3	0	1
	EUNIS Habitat Classification	0	0	0	0	0	3	0	1
	FISHTRAITS | Fish Traits Thesaurus	0	0	0	0	0	3	0	1
	ICOS Ontology | Integrated Carbon Observation System Ontology	0	0	0	0	3	0	0	1
	ISO 3166-1:2013 Codes for the representation of names of countries and their subc	0	0	0	0	0	3	0	1
	IUCN | International Union for Conservation of Nature - Red List of Threatened Spe	0	0	0	0	0	3	0	1
	MACROALGAETRAITS | Macroalgae Traits Thesaurus	0	0	0	0	0	3	0	1
	Marine Regions georeferences	0	0	0	0	0	3	0	1
	NVS | NERC Vocabulary Service	0	2	0	0	0	3	0	1
	OBOE-AnaEE ontology	2	0	0	0	0	0	0	0
	PHYTOTRAITS | The Phytoplankton Traits	0	0	0	0	0	3	0	1
	WoRMS | World Register of Marine Species	0	0	0	0	0	3	0	1

Figure 9: Answers to the question regarding the FAIR principle I2. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

The environmental domain encompasses a vast array of disciplines, ranging from ecology and biodiversity to climate science and natural resource management. Each of these disciplines has its unique terminology, concepts, and data structures. As a result, there is a proliferation of controlled vocabularies and semantic artifacts. Different researchers, institutions, and projects may use different terms and concepts to describe similar phenomena or variables. This heterogeneity poses challenges when attempting to combine and compare data from different sources.
While converging towards a single common vocabulary may seem desirable, it is often impractical due to the diverse terminology and concepts used across different disciplines and domains. Each field has its unique context-specific terms and nuances that cannot be easily unified. Instead, providing efficient mapping mechanisms allows for the integration and interoperability of diverse vocabularies and semantic artifacts. It recognises that different communities and disciplines may have their own well-established vocabularies and concepts that are essential for their specific research and understanding. By enabling mappings between these vocabularies, researchers can bridge the gaps and establish connections, facilitating effective communication and knowledge exchange.

In this context, the use of ontologies plays a significant role in linking semantic artifacts. Ontologies provide a formal representation of knowledge, capturing the relationships and hierarchies between concepts. I-ADOPT, as an example of an ontology-based approach, is designed to link and harmonise diverse semantic artifacts in the environmental domain. It provides a framework for connecting existing ontologies, controlled vocabularies, and other semantic resources. By leveraging I-ADOPT or similar approaches, RIs can benefit from the interconnectedness and interoperability of semantic artifacts, enabling advanced data analysis, discovery, and decision support.

4.3 I3. Which models, schema(s) do you use for your (meta)data?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
I3-Metadata	DC | Dublin Core	0	2	0	0	0	0	0	0
	DEIMS-SDR_Site_Metadata_Model	0	0	0	3	0	0	0	1
	DIF | Directory Interchange Format	0	0	0	0	0	0	3	1
	EML2.2.0 | Ecological Metadata Language 2.2.0	0	0	0	0	0	3	0	1
	I-ADOPT Framework	0	0	0	2	0	0	0	0
	ICOS Ontology | Integrated Carbon Observation System Ontology	0	0	0	0	3	0	0	1
	INSPIRE EMF | Infrastructure for Spatial Information in the European C	0	0	0	3	0	0	0	1
	ISO 19115 | Geographic information - Metadata	3	0	0	3	0	0	3	3
	ISO 19139 | Geographic information - Metadata - XML schema impleme	0	0	0	3	0	3	0	2
	LifeWatch ERIC Metadata Catalogue	0	0	0	0	0	3	0	1
	LUPO | LifeWatch ERIC Upper Ontology	0	0	0	0	0	3	0	1
	openDS | Open Digital Specimen	0	0	1	0	0	0	0	0
	WMO Core Profile | World Meteorological Organization Core Metadat	0	0	0	0	0	0	3	1
I3-Data	DCAT-AP | Data Catalog Vocabulary Application Profile for Data Portals	0	2	0	0	0	0	0	0
	DwC-A | Darwin Core Archive	0	0	0	0	0	3	3	2
	eLTER_Data_Specification_Draft	0	0	0	3	0	0	0	1
	I-ADOPT Framework	0	0	0	2	0	0	0	0
	ICOS Ontology | Integrated Carbon Observation System Ontology	0	0	0	0	3	0	0	1
	NetCDF CF-1.7	0	2	0	0	3	0	3	2
	NetCDF_CF_SDN | NETCDF CF format SeaDataNet Profile	0	0	0	0	0	3	0	1
	OBOE | Extensible Observation Ontology	3	0	0	0	0	0	0	1
	OGC SensorML | Open Geospatial Consortium Sensor Model Language	0	0	0	3	0	0	0	1
	openDS | Open Digital Specimen	0	0	1	0	0	0	0	0
	WMO Core Profile | World Meteorological Organization Core Metadat	0	0	0	0	3	0	0	1

Figure 10: Answers to the question regarding the FAIR principle I3. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

At first glance this question seems overlapping with F2. The help text provided along with question I3 in the FIP Wizard (https://fip-wizard.ds-wizard.org/dashboard) state that *"this question requests a FAIR Enabling Resource of type "semantic model" which is a specification that defines qualified relations between entities describing data or other digital objects using structured vocabularies. A semantic model can be a conceptual model expressed as an ontology or as a metadata scheme that reuses terms from FAIR vocabularies."* The use of semantic models is crucial for the FAIRness of the resources, in particular for the Interoperability, which is known to be the most difficult to be addressed. In connection with I2, semantic models are fundamental for enhancing the interoperability and interconnectedness of semantic artifacts, supporting advanced knowledge representation and analysis in the environmental field. From the analyses of FIP in the Biodiversity and Ecosystem subdomain it looks like there is not a real convergence toward a few FERs.

5 Reusability

The Reusability of (meta)data emphasises the importance of ensuring that they are easily understandable and usable by both humans and machines. It involves providing rich metadata and provenance, a clear documentation, as far as a clear license to reuse them and relevant attributes that facilitate the understanding of the (meta)data and their reuse in various contexts, even by researchers from other domains.

5.1 R1.1. Which usage license do you use for your (meta)data records?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
R1.1-Metadata	CC BY 4.0 | Attribution 4.0 International	3	2	2	0	0	0	0	1
	CC0 1.0 | CC0 1.0 Universal Public Domain Dedication	0	0	2	3	3	3	0	3
R1.1-Data	CC BY 4.0 | Attribution 4.0 International	3	2	2	0	3	3	3	4
	CC BY SA 4.0 | Attribution-ShareAlike 4.0 International	0	0	0	0	0	0	3	1
	CC BY-NC 4.0 | Attribution-NonCommercial 4.0 International	0	0	0	3	0	3	3	3
	CC BY-NC-SA 4.0 | Attribution-NonCommercial-ShareAlike 4.0 International	0	0	0	0	0	0	3	1
	CC0 1.0 | CC0 1.0 Universal Public Domain Dedication	0	0	0	0	3	3	0	2

Figure 11: Answers to the question regarding the FAIR principle R1.1. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

The Creative Commons Zero (CC0) license is a permissive public domain dedication that allows for the unrestricted use, distribution, and modification of a work without requiring any attribution or granting any warranties. For European projects, metadata should always be openly available and licenced under a public domain dedication CC0. According to this principle, the RIs of the subdomain published their metadata mostly as CC0.

Regarding data, CC BY licenses, or Creative Commons Attribution licenses, are more appropriate, because they assure that data are cited. CC BY licenses are a set of open licenses that allow creators to share their work with others while maintaining certain rights. Those licenses require attribution, meaning that anyone using the licensed material must give appropriate credit to the original creator. This is pivotal in research, because it gives proper credit and recognition to the original creators or contributors of the dataset by acknowledging their intellectual and scholarly efforts.

5.2 R1.2. Which metadata schemas do you use for describing the provenance of your (meta)data?

FIP question	FAIR Enabling Resources	AnaEE-ERIC	Danubius-RI	DiSSCo	eLTER-RI	ICOS-ERIC	LifeWatch ERIC	SIOS	Total
R1.2-Metadata	ISO 19115 | Geographic information - Metadata	0	0	0	3	0	0	0	1
	PROV-O | W3C PROV Ontology	2	0	2	2	3	0	0	1
R1.2-Data	DCAT-AP | Data Catalog Vocabulary Application Profile for Data Portals in Europe	0	2	0	0	0	0	0	0
	PROV-O | W3C PROV Ontology	2	0	2	2	3	0	0	1

Figure 12: Answers to the question regarding the FAIR principle R1.2. Legend: 0 - Resource not declared by community; 1 - Resource in development, future use; 2 - existing Resource, future use; 3 - existing Resource, current use.

Provenance is highly important in understanding digital artifacts as it provides crucial information about their origin, history, and transformations. It allows researchers and users to trace the lineage of data, understand the processes and steps involved in its creation, and assess its reliability and trustworthiness. How are metadata different from provenance though? One can think of metadata as data descriptions that assign meaning to the data, and data provenance as the information about how data was derived (CODATA definition). In this regard, only eLTER-RI and ICOS-ERIC did implement a specific schema for provenance. In general, the discussion on the general guidelines for documenting provenance is relatively new and still in progress. This reflects the situation in general for the implementation choices made by the ENVRI-RIs, with PROV-O being one of the choices that most of the RIs are implementing to now. While PROV-O, also known as the W3C provenance ontology, offers a general framework for representing provenance information, there is a need for more specific and detailed implementations. One such implementation is DDI-CDI (https://codata.org/initiatives/ decadal-programme2/ddi-cross-domain-integration/), which already incorporates some of these concepts but requires further refinement to meet specific requirements and use cases.

6 Discussion

Achieving FAIRness in RIs service provision is of utmost importance for several reasons. First and foremost, it promotes openness and transparency, enabling seamless access to data, tools, and services for researchers and other stakeholders. By adhering to the FAIR principles, RIs can ensure that their services are Findable, Accessible, Interoperable, and Reusable, thus fostering collaboration and maximising the potential impact of research outcomes. Furthermore, FAIRness enhances data sharing and data-driven research, allowing for efficient and effective knowledge discovery. By implementing standardised metadata, persistent identifiers, and interoperable data formats, RIs can facilitate the discovery and reuse of valuable data assets. This not only accelerates scientific progress but also enables reproducibility and validation of results.

On top of this, the FAIR principles align with broader initiatives and policies at the European level, such as the European Open Science Cloud (EOSC) and the Digital Single Market strategy. By adopting FAIR principles, research infrastructures contribute to the harmonisation and integration of European research efforts, promoting cross-disciplinary collaborations and unlocking the full potential of data-driven research.

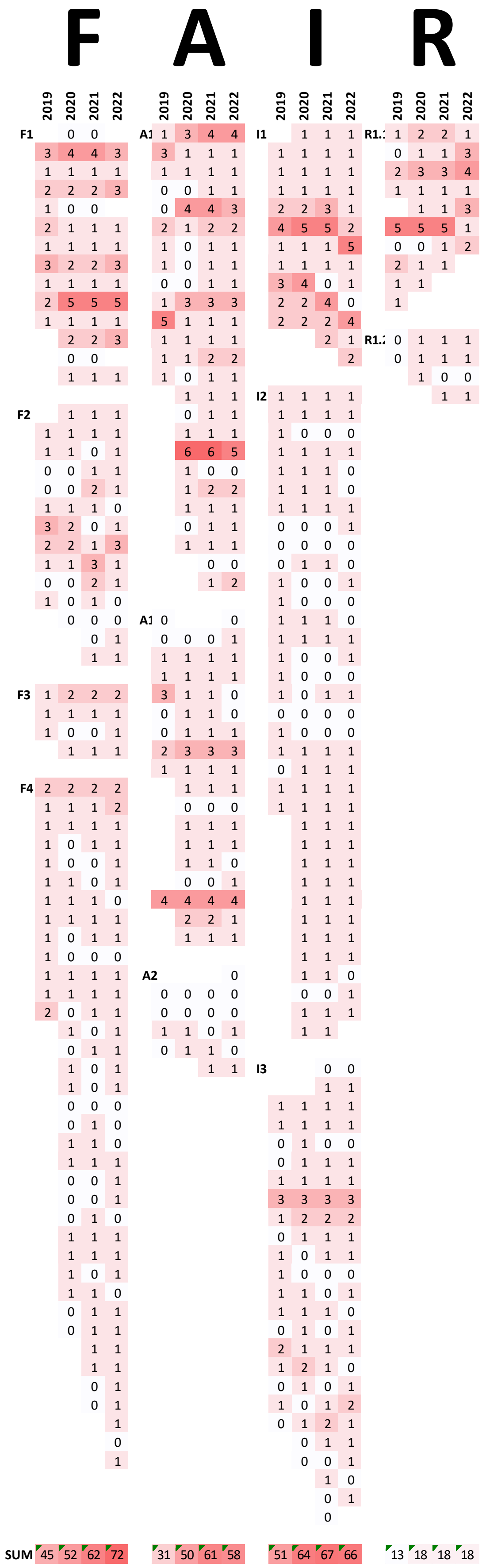

Figure 13: Overview on the development in FAIRness

If we look at the overall picture presented in the Figure 13 above, it is clear that, for what regards the RIs that belong to the Biodiversity and Ecosystem domain the number of FERs is high.
The total number of FERs reported by the RIs is generally quite high for each of the principles. This is partly due to the fact that for some technical implementations, such as the use of controlled vocabularies or data and metadata schemas, they are specific to the scientific community to which each RI refers. This is particularly true for RIs belonging to the biodiversity and ecosystems domain, as it represents a highly heterogeneous and interdisciplinary field. Within this domain, RIs range from those dealing with museum specimens (such as DiSScO), those primarily focused on monitoring (such as eLTER, ICOS), to those more oriented towards biodiversity observation (such as LifeWatch). Nevertheless, it is still evident that there is a certain level of convergence in some choices, especially towards technologies perceived as highly robust, such as the use of DOI as a Persistent Identifier (sub-principle F1), the use of XML as the metadata exchange language (sub-principle I1), or the adoption of CC-BY licenses (sub-principle R1).

If we look at the evolution of the total number of solutions adopted by the RIs in the subdomain, it is clear that there is a noticeable increase over the four years of the project. The increase in the number of FERs has affected all RIs, regardless of their level of maturity. This increase can be partly explained by

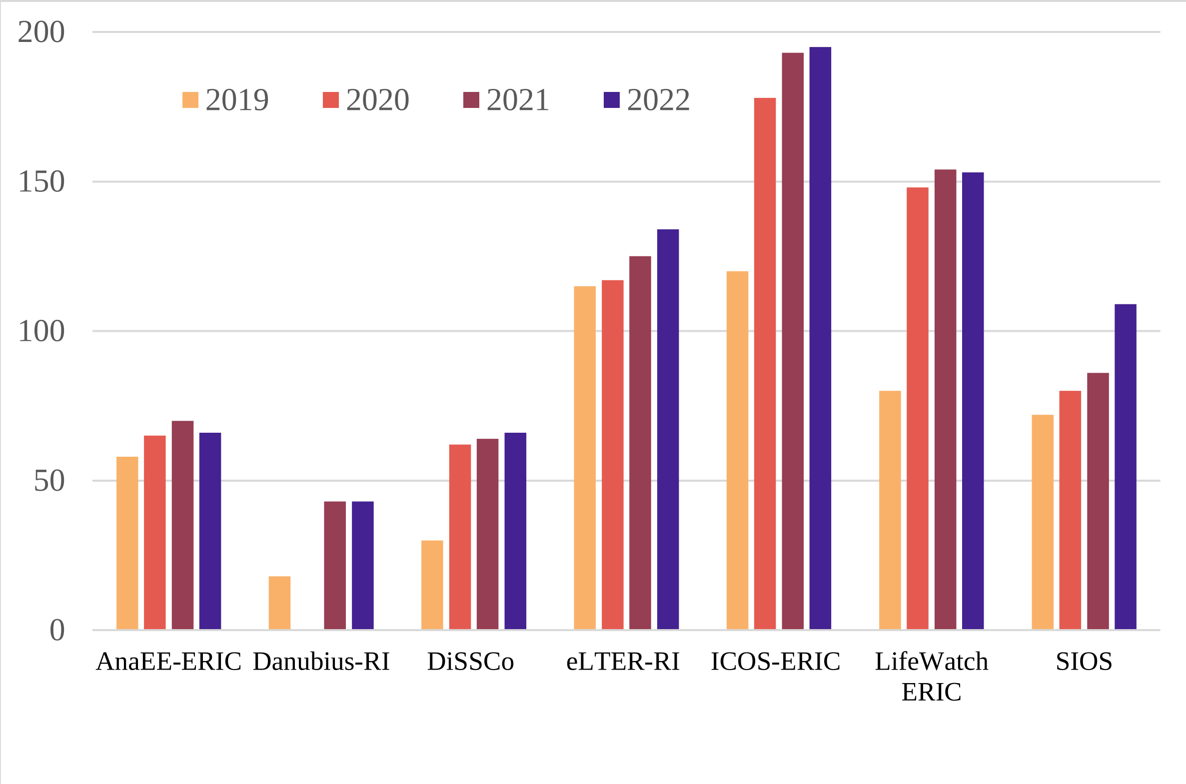

Figure 14: Evolution of the total number of FERs

the fact that the expanded network created by the ENVRI cluster has enabled the personnel of the involved RIs to exchange more information and enhance their understanding of FERs, leading them to decide to adopt or plan their usage in the near future.
In conclusion, the ENVRI-FAIR project has successfully established the technological baseline and facilitated the network required for the implementation of FAIR principles in the RIs of WP11 and the broader ENVRI cluster. The project has laid a solid foundation for promoting the FAIRness of data and services within the research infrastructures. However, the future implementation and sustainability of these efforts will depend on the availability of budgets from new projects and the governance choices made by the individual RIs. Continued support and funding will be crucial in ensuring the long-term success and impact of the FAIR initiatives across the European research landscape. It is imperative that stakeholders continue to prioritise the adoption of FAIR principles to enable greater accessibility,

interoperability, and reusability of research data, ultimately fostering scientific advancements and collaboration in the environmental domain.

7 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. Special thanks to Barbara Magagna who provided us the full picture of the FIPs from which the tables where realised.

8 References

1. Schultes, Erik, et al. "Reusable FAIR implementation profiles as accelerators of FAIR convergence." Advances in Conceptual Modeling: ER 2020 Workshops CMAI, CMLS, CMOMM4FAIR, CoMoNoS, EmpER, Vienna, Austria, November 3–6, 2020, Proceedings. Cham: Springer International Publishing, 2020.
2. Magagna, Barbara, et al. "The I-ADOPT Interoperability Framework for FAIRer data descriptions of biodiversity." arXiv preprint arXiv:2107.06547 (2021). Retrieved from https://doi.org/10.48550/arXiv.2107.06547
3. European Commission. (2023). EU Annotated Grant Agreement [V1.0 DRAFT– 01.04.2023]. Retrieved from https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf

9 Appendices

9.1 Appendix A: Glossary and terminology

The following is the list of acronyms and terms used in this deliverable[1]:

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
DANUBIUS-RI	International Centre for Advanced Studies on River-Sea Systems
DataCite	A leading global non-profit organisation that provides persistent identifiers (DOIs) for research data and other research outputs
DCAT	Data Catalogue Vocabulary
DDI-CDI	Data Documentation Initiative - Cross-Domain Integration
DEIMS-SDR	Dynamic Ecological Information Management System - Site and dataset registry
DiSSCo	Distributed System of Scientific Collections
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
EML	Ecological Metadata Language
ENVRI	Environment research infrastructures
ENVRI-hub	A federated machine-to-machine interface to access environmental data and services provided by the contributing ENVRIs
EOSC	European Open Science Cloud
ERIC	European Research Infrastructure Consortium
FAIR	Findable Accessible Interoperable Reusable
JSON	JavaScript Object Notation

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

I-ADOPT	InteroperAble Descriptions of Observable Property Terminology
ICOS	Integrated Carbon Observation System
ISO	International Organisation for Standardisation
LifeWatch	LifeWatch European Research Infrastructure Consortium
LOD	Linked Open Data
NERC	Natural Environment Research Council
NetCDF	Network Common Data Format.
OpenID	Open standard authentication protocol (it allows for signing into multiple websites with a unique account)
ORCID	Open Researcher and Contributor ID
OWL	Web Ontology Language
PID	Persistent Identifier
PROV-O	Provenance Ontology
RDF	Resource Description Framework
RI	Research Infrastructure
SIOS	Svalbard Integrated Arctic Earth Observing System
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
WP	Work Package
XML	eXtensible Markup Language