D7.1
Design documents describing FAIR data services provided by selected RIs (V1)

Work Package	WP7
Lead partner	TIB
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	30-06-2020
Submission date	18-09-2020

Deliverable abstract

We present the ENVRI-FAIR approach to consultancy by ICT experts with research infrastructure managers and developers, aimed at technology convergence across the community of environmental research infrastructures (ENVRIs) in order to ensure the production of FAIR (meta)data across the community. We identify three critical areas—(meta)data Catalogue, Syntax and Semantics—in which technology convergence is recommended, and prioritize the FAIR principles for their relevance to the three critical areas. We then present the range of instruments for consultancy at our disposal and how we have used them so far to support ENVRIs in implementing the FAIR principles. First results can be seen in implementation of provenance.

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

DELIVERY SLIP

	Name	Partner Organization	Date
Main Author	Markus Stocker	TIB	
Contributing Authors	Barbara Magagna Keith Jeffery Markus Fiebig Peter Thijsse Daniele Bailo Xiaofeng Liao	EAA UKRI/BGS NILU MARIS INGV UvA	
Reviewer(s)	Margareta Hellström Christian Pichot Zhiming Zhao	ULUND INRAE UvA	
Approver	Andreas Petzold	FZJ	18-09-2020

DELIVERY LOG

Issue	Date	Comment	Author
V 1.0	13-07-2020	First Draft	Markus Stocker
V 2.0	31-07-2020	Second Draft	Markus Stocker Barbara Magagna Keith Jeffery
	17-08-2020	Comments from Reviewer 1	Margareta Hellström
	07-08-2020	Comments from Reviewer 2	Christian Pichot
V 3.0	31-08-2020	Final version	Markus Stocker

DOCUMENT AMENDMENT PROCEDURE

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

GLOSSARY

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

PROJECT SUMMARY

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

The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI Cluster to the EOSC.

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

TABLE OF CONTENTS

D7.1 - Design documents describing FAIR data services provided by selected RIs (V1)

1 Introduction

With the aim to increase the FAIR-ness of their (meta)data, the ENVRI-FAIR subdomains—Atmosphere, Marine, Biodiversity/Ecosystem, Solid Earth—and their RIs have each developed implementation plans (Fiebig et al., 2020; Thijsse et al., 2019; Papale, 2020; Solid Earth to be published) that present a roadmap towards FAIR (meta)data following the principles outlined in Wilkinson et al. (2016).

ENVRI-FAIR WP7 supports such implementations by closely interacting with RIs to provide targeted support for designing, recommending and implementing solutions to overcome gaps identified by WP5, and for testing and validating the development results beyond RIs at subdomain, cluster and EOSC levels. Task 7.1 focuses on the interaction with RIs and targeted consultation for selecting technologies, integrating solutions for currently missing functionality and establishing feasible development plans for FAIR (meta)data services.

As a key aim, WP7 works on technology convergence to ensure FAIR (meta)data at subdomain and ENVRI cluster levels. Towards this aim, this deliverable describes how ENVRI-FAIR plans to harmonize key aspects of RI and subdomain FAIR implementation plans, and presents the steps already taken to support these efforts. In a nutshell, this deliverable describes a strategy for a common implementation plan.

Gaining understanding of the plans at the RI and subdomain levels is valuable and essential for ENVRI-FAIR and WP7 activities, in particular. Indeed, these four plans are an important source of information for WP7 implementation support and consultancy. They are largely described from the point of view of individual RIs, and presented as collections by subdomain. This suggests an important open question: Will the individual implementations outlined in plans result in FAIR (meta)data within each subdomain, across ENVRI-FAIR RIs and more broadly RIs in the ENVRI community?

While FAIR (meta)data at the RI level is essential, it obviously is merely a first step towards a larger vision, which aims at FAIR (meta)data at the level of the ENVRI community (i.e., across ENVRIs). In the ENVRI-FAIR project, an intermediate milestone is FAIR (meta)data across domain specific RIs.

In practice this means that researchers and other stakeholders should be able to find and access (meta)data across ENVRIs and all (meta)data accessed are interoperable and reusable. To achieve this objective, we suggest three areas in which RIs should not just work on individual solutions but *common* solutions. The three areas are introduced as follows.

Table 1: The three critical areas and recommendations for common solutions.

Nr	Area	Short Name	Recommendation
1	Catalogue	Cat	ENVRIs must have the ability to present their metadata to a common (meta)data catalogue through which (meta)data published by each individual ENVRI can be found.
2	Syntax	Syn	(Meta)data published by ENVRIs must have a machine readable format (syntax) compliant with domain-relevant community standards.
3	Semantics	Sem	(Meta)data published by ENVRIs must have unambiguous interpretation (semantics).

The first area, Catalogue, is relatively trivial. In order to find (meta)data published by an RI, the (meta)data must be catalogued, most obviously in a catalogue operated by the RI itself. In order to find

(meta)data published by any RI, a common (meta) catalogue is needed. Such a common catalogue has to physically or virtually (through distributed querying) harvest individual RI catalogues.

The second area, Syntax, entails that it must be possible to process (meta)data published by RIs using approaches commonly used in the domain. For instance, if in a particular community it is common to publish observational data in NetCDF format then it is effective for RIs to publish such data in this format. Publishing data in another open format would be less effective and it would not be effective to use a proprietary format.

The third area, Semantics, implies that (meta)data that ought to be interpreted in a given way is indeed understood in this original meaning by everyone. For instance, if two RIs each publish an observation about water temperature then the values should be unambiguously interpreted as water temperature, by both humans and machines. This implies that the RIs must have common vocabularies or at least use concepts that are mapped to each other.

We suggest that in these three areas—(meta)data Catalogue, Syntax and Semantics—RIs should work on common solutions. Critical is that users can find data or services[1] providing access to data and that systems agree on an exchange protocol, but a (large) plurality of data access protocols can co-exist, assuming they are open, free, and universally implementable.

2 Critical Areas and FAIR Principles

In this section, we identify the FAIR (sub) principles that are key in light of the proposed critical areas—(meta)data Catalogue, Syntax and Semantics—in which RIs should work on common solutions. The provided recommendations will guide the presentation and discussion in this deliverable as well as the future work in WP7 activities.

F1. (Meta)data are assigned a globally unique and persistent identifier. This principle is important for two of our critical areas: Catalogue and Semantics. RIs assets, i.e. metadata, data, instruments, software, etc., should be assigned globally unique and resolvable persistent identifiers. Identifiers should be governed by a registry that ensures uniqueness and persistence or use a machine-generated unique ID (to avoid the management overhead of the registry). Identifiers are essential for cataloguing, finding and accessing RI assets and are thus essential for an ENVRI common catalogue that ensures community-wide findability of assets. Identifiers are also essential for unambiguous semantics, thus ensuring that other people and machines understand the intended meaning of data. Hence, the string "Andreas Petzold" is ambiguous while the globally unique and persistent identifier https://orcid.org/0000-0002-2504-1680 isn't. Similarly, the string "ACE2" is ambiguous while the identifier https://www.uniprot.org/uniprot/Q9BYF1 isn't. Increasingly, it is becoming standard for the IDs to be resolvable (via addressing) to a digital object. This links with A1.

Recommendation. A plurality of globally unique and persistent identifiers can co-exist in ENVRI, but ENVRI-FAIR should ensure that the same identifiers are used across the RIs when referring to same entities, or that if different identifiers are used that they are related as being equivalent. This applies not only to the quantities and objects related to the acquired data, but also to the sites, facilities, infrastructures and actors.

F2. Data are described with rich metadata. This principle is important for all three of our critical areas. Rich metadata[2] is essential for findability, e.g. if the identifier is not known. Metadata harvested into an ENVRI common catalogue will support findability to the extent that metadata richly describe assets, e.g. datasets or services. Moreover, ENVRI benefits from agreement on metadata elements,

[1] While it is critical to be able to find and access services and that services should be interoperable and reusable, we note that the FAIR data principles were not designed for services, but (meta)data. For services, the set of FAIR principles would be a different one, even though the overarching principles of findability, accessibility, interoperability and reusability apply to services, too. Koers et al. (2020) discuss recommendations for data and infrastructure service providers to support FAIR research data.

[2] See CODATA-ICSTI Task Group on Data Citation Standards and Practices (2013) for a brief discussion about the different metadata categories, e.g. descriptive, administrative, and structural metadata

whether they are mandatory or optional, the accepted value space (controlled vocabularies) since such agreement will increase the quality of the metadata harvested across ENVRIs, increase the value of a common catalogue, and support findability across the community. The principle is also important for the other two critical areas: Syntax and Semantics. Quality controlled rich metadata that follow domain-relevant community standards will result in shared syntax and semantics of metadata published by ENVRIs and thus improve its reuse.

Recommendation. ENVRI-FAIR should ensure that metadata published by ENVRIs converge on shared schemas and controlled vocabularies.

F3. Metadata clearly and explicitly include the identifier of the data they describe. This principle is important for the Catalogue critical area. In order to support the linking to data in an ENVRI common catalogue, it is necessary that the harvested metadata includes the identifier of the data they describe and that this link is qualified with formal syntax and rich declared semantics. In order to support automated workflows, it is useful if the metadata attribute for the identifier is the same across ENVRI metadata. This principle is linked with I3: the relationship between one (meta)data object and another should be qualified—implying more than a simple attribute value and encoding of the role of the relationship (optionally with other attributes).

Recommendation. ENVRI-FAIR should ensure ENVRIs-wide agreement for the metadata attribute name used to include the identifier.

F4. (Meta)data are registered or indexed in a searchable resource. This principle is important for the Catalogue critical area. An ENVRI common catalogue can act as such a resource. However, the metadata should be harvested from RI-specific catalogues that register and index metadata. For some identifier types, e.g. DOI and ORCID, the service providers also operate catalogues, albeit specifically for the metadata they collect upon identifier minting. The metadata registered and indexed by such service providers may thus be different from the metadata registered and indexed by RI-specific catalogues. The indexing of data to make the data themselves searchable (as in searching inside of the data rather than searching for data in a catalogue) should, in general, not be lifted above the RI level.

Recommendation. ENVRI-FAIR should ensure metadata are harvested and indexed in an ENVRI common catalogue as its searchable resource. In contrast, the indexing of data should remain at the level of the individual RI.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol. As we suggested above, a plurality of communication protocols can co-exist in ENVRI, as long as the protocols are (1) open, free, and universally implementable and (2) allow for authentication and authorisation procedures. Relevant for our critical area Catalogue is that data and metadata can be retrieved by resolving their identifiers.

Recommendation. ENVRI-FAIR should ensure that (meta)data can be retrieved by resolving their identifiers registered with an ENVRI common catalogue.

A2. Metadata are accessible, even when the data are no longer available. This principle is important for the Catalogue critical area and can be implemented by an ENVRI common catalogue.

Recommendation. ENVRI-FAIR should ensure that metadata registered and indexed by an ENVRI common catalogue are accessible even if the underlying data are no longer available through the RI.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. This principle is relevant to all three critical areas. Most importantly, it is key to Semantics and ensuring that (meta)data published by ENVRIs have unambiguous interpretation, for machines and people. It is also relevant to (meta)data Syntax because formal languages for knowledge representation structure data in machine readable manner, indeed the structure specification itself is machine readable. Notably, the considerations are not just valid for metadata, but for data as well. In order to be reusable, datasets should be represented following a formal (i.e., machine readable) language that is shared in a community. A concrete example is NetCDF CF Metadata Conventions, an approach to include CF attributed in NetCDF encoded data. In this case, there is community agreement on the metadata as CF attributes and agreement on encoding data as NetCDF. The languages used are

formal, accessible, shared and broadly applicable. Ultimately, this principle benefits Cataloguing because the harvested metadata are readable by standard tools, since metadata syntax and semantics are encoded using a formal language that is common and for which tools have been developed.

Recommendation. ENVRI-FAIR should ensure convergence on languages for knowledge representation that are formal, accessible, shared, and broadly applicable throughout the community.

I2. (Meta)data use vocabularies that follow FAIR principles. This principle is relevant to all three critical areas. Vocabularies and their terms should be used to structure (meta)data, but can also be used as data elements. Clearly, if terms used are not defined in FAIR vocabularies, then their use in (meta)data is not optimal. For instance, if a dataset uses the identifier https://www.uniprot.org/uniprot/Q9BYF1 for a protein but this identifier cannot be resolved to retrieve data about the protein then the value of having the identifier in the dataset is minimal. Indeed, it may then be better to include the label "ACE2" or to provide both.

Recommendation. ENVRI-FAIR should ensure convergence on vocabularies that follow the FAIR principles.

I3. (Meta)data include qualified references to other (meta)data. This principle is relevant to the Catalogue and Semantics critical areas. References can be more or less qualified and more qualified references are preferred. For example, "dataset B is a new version of dataset A" is a more qualified reference than "dataset B is related to dataset A". Qualified references increase the quality and capabilities of the catalogue and decrease ambiguity by accurately relating to contextual information, and thus enable more precise encoding of intended meaning. However, we suggest that this principle is largely covered by F2 and R1, which together suggest that (meta)data are richly described. We suggest that for (meta)data to be rich it must include qualified references to other (meta)data.

Recommendation. See F2 and R1.

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes. This principle is relevant to all three critical areas. For cataloguing, rich description improves findability and, according to GO-FAIR[3] , R1 should ensure that users can find data and determine if they are useful in a particular context of reuse. Data should thus describe the context in which they were generated. The principle is also relevant to (meta)data Syntax and Semantics, since decisions on which attributes are relevant affect (meta)data structure and interpretation.

The RDA Metadata Interest Group[4]has been trying to achieve convergence throughout the community on a set of metadata elements of general applicability.

Recommendation. ENVRI-FAIR should ensure that the sub-domains converge on which essential attributes are required for users to decide whether their data can be reused in contexts that differ from the one in which the data were generated.

R1.1. (Meta)data are released with a clear and accessible data usage license. This principle is relevant to legal interoperability. On a technical dimension, it is useful to agree on a common controlled vocabulary that ensures unambiguous interpretation of license information. Hence, this principle is relevant to the Syntax and Semantics critical areas. Work is ongoing to translate licences into a set of logic statements that could then be machine actionable for managing authorisation.

Recommendation. ENVRI-FAIR should ensure convergence on a controlled vocabulary for license information.

R1.2. (Meta)data are associated with detailed provenance. Provenance information is essential for reuse and this principle is relevant to all three critical areas. Provenance captures important contextual information invaluable for ensuring unambiguous interpretation of intended meaning and is thus

[3] https://www.go-fair.org/fair-principles/r1-metadata-richly-described-plurality-accurate-relevant-attributes/
[4] https://www.rd-alliance.org/groups/metadata-ig.html

relevant to the Semantics critical area. It is also relevant to the Catalogue critical area because provenance metadata may be useful to find or contextualise data. Finally, to ensure interoperability of provenance metadata across ENVRIs, it is important to converge on a syntax for the encoding of provenance metadata.

Recommendation. ENVRI-FAIR should ensure convergence on how provenance information is represented across ENVRIs and related to (meta)data catalogues.

R1.3. (Meta)data meet domain-relevant community standards. This principle is relevant to all three critical areas. The adoption of domain-relevant community standards supports cataloguing and effectively finding (meta)data in catalogues. They also ensure that (meta)data follow well-known and tool-supported syntax and that the intended meaning of (meta)data contents are interpreted unambiguously.

Recommendation. ENVRI-FAIR should ensure convergence on which domain-relevant standards to use to encode (meta)data.

Table 2: Recommendations by FAIR principle, prioritized by number of relevant critical areas.

FAIR Principle	Recommendation	Critical areas
F2. Data are described with rich metadata.	ENVRI-FAIR should ensure that metadata published by ENVRIs converge on shared schemas and controlled vocabularies.	Cat, Syn, Sem
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	Ensure convergence on languages for knowledge representation that are formal, accessible, shared, and broadly applicable throughout the community.	Cat, Syn, Sem
I2. (Meta)data use vocabularies that follow FAIR principles.	Ensure convergence on vocabularies that follow the FAIR principles.	Cat, Syn, Sem
R1. Meta(data) are richly described with a plurality of accurate and relevant attributes.	ENVRI-FAIR should ensure that the sub-domains converge on which essential attributes are required for users to decide whether their data can be reused in contexts that differ from the one in which the data were generated.	Cat, Syn, Sem
R1.2. (Meta)data are associated with detailed provenance.	ENVRI-FAIR should ensure convergence on how provenance information is represented across ENVRIs and related to (meta)data catalogues.	Cat, Syn, Sem
R1.3. (Meta)data meet domain-relevant community standards.	ENVRI-FAIR should ensure convergence on which domain-relevant standards to use to encode (meta)data.	Cat, Syn, Sem
F1. (Meta)data are assigned a globally unique and persistent identifier.	Ensure that the same identifiers are used across the RIs when referring to the same entities, or that if different identifiers are used that they are related as being equivalent. This applies not only to the quantities and objects related to the acquired data, but also to the sites, facilities, infrastructures and actors.	Cat, Sem
I3. (Meta)data include qualified references to other (meta)data.	See F2 and R1.	Cat, Sem
R1.1. (Meta)data are released with a clear and accessible data	Ensure convergence on a controlled vocabulary for license information.	Syn, Sem

FAIR Principle	Recommendation	Critical areas
usage license.		
F3. Metadata clearly and explicitly include the identifier of the data they describe.	Ensure ENVRIs-wide agreement for the metadata attribute name used to include the identifier.	Cat
F4. (Meta)data are registered or indexed in a searchable resource.	Ensure metadata are harvested and indexed in an ENVRI common catalogue as its searchable resource. In contrast, the indexing of data should remain at the level of the individual RI.	Cat
A1. (Meta)data are retrievable by their identifier using a standardised communications protocol.	Ensure that (meta)data can be retrieved by resolving their identifiers registered with an ENVRI common catalogue.	Cat
A2. Metadata are accessible, even when the data are no longer available.	Ensure that metadata registered and indexed by an ENVRI common catalogue are accessible even if the underlying data are no longer available through the RI.	Cat

3 Analysis of Implementations

This section presents a preliminary analysis of the implementations planned or developed by RIs in the different ENVRI-FAIR subdomains. With this analysis we aim to understand the plans RIs have to implement the FAIR principles, the heterogeneity of solutions, and ultimately where we should ensure convergence on solutions. The analysis is guided by the results presented in Section 2 in that we prioritize those FAIR principles that affect all three critical areas (Table 3). A comprehensive analysis of all relevant FAIR principles will be conducted during ENVRI-FAIR and the final results will be presented in a later deliverable (D7.2).

According to Table 3, the FAIR principles that affect all three critical areas in which ENVRI should develop common solutions are: F2, I1, I2, R1, R1.2 and R1.3. The goal in this section is to look at these principles, provide an overview of the solutions, and analyse to what extent the implementation plans meet the proposed recommendation. Table 3 below provides an overview of the solutions that RIs are aiming to implement during the ENVRI-FAIR project duration.

There is indeed high heterogeneity in the solutions different RIs implement to address the six principles that affect all three critical areas. Most convergence can be found for principle I1 on RDF, related serializations (e.g., JSON-LD) and OWL as a formal language for knowledge representation that builds on the RDF data model. Some convergence transpires also for principle R1.2 on W3C Provenance technology. In the Marine subdomain, there is some convergence on using the NERC Vocabulary Server.

However, on other equally important principles the solutions are presently very diverse. For instance, F2 has no overlap in solutions used to ensure that data are described with rich metadata. Moreover, some infrastructures use their own developments (e.g., DANUBIUS, DISSCO). This is not inherently negative, but it complicates convergence. More importantly, the provided solutions—which reflect the information extracted from collected materials (e.g., implementation plans)—cannot be assessed for how good a solution they truly are, i.e. how well they support describing data with rich metadata.

For principles I2 and R1.3, convergence is inherently difficult because of community specificity. While convergence on the use of vocabularies that are FAIR (I2) is practicable, convergence on the vocabularies themselves, especially beyond the subdomain boundaries, is for obvious reasons not straightforward, since vocabularies are often domain-specific. The problem is even more obvious for

R1.3 as the principle sets an explicit emphasis on domain-relevant community standards. A high degree of diversity and consequent difficult or outright impossible convergence is arguably unavoidable here. Viable possibilities for harmonizations are probably only (1) convergence on abstract concepts (upper ontologies) and (2) where possible alignments of concrete concepts.

Finally, given the analyzed materials, we were unable to extract useful information on R1, perhaps with the exception of EPOS. In contrast to F2, R1 focuses on how well rich metadata supports reuse. Hence, data not only need to be described with rich metadata, but the description richness must be so that it supports deciding whether or not data can be reused in a different context. In general, the available material does not provide satisfactory answers to this question. This is even though we classify R1 as a principle that affects all three critical areas, Catalogue, Syntax, and Semantics and is thus a principle that requires attention.

Table 3: Solutions for the FAIR principles that affect all three critical areas.

F2. Data are described with rich metadata. *REC. Ensure that metadata published by ENVRIs converge on shared schemas.*	
ICOS-Marine	• Keywords following Dewey scheme[5]
EISCAT_3D	• DataCite Schema • ESPAS ontology[6]
IAGOS	• WIGOS metadata profile[7]
EPOS	• CERIF[8]
ANAEE	• ISO 19115[9] / ISO 19139
DANUBIUS	• DANUBIUS-Commons
DISSCO	• DiSSCo Data Management Plan[10]
eLTER	• ISO 19119 / ISO 19139 • INSPIRE EF[11]
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. *REC. Ensure convergence on languages for knowledge representation that are formal, accessible, shared, and broadly applicable throughout the community.*	
LifeWatch	• JSON-LD
SeaDataNet	• RDF
ICOS-Atmospheric	• NetCDF-CF
EPOS	• CERIF, RDBMS to RDF

[5] https://en.wikipedia.org/wiki/Dewey_Decimal_Classification
[6] http://ontology.espas-fp7.eu/
[7] https://library.wmo.int/index.php?lvl=notice_display&id=19925
[8] https://www.eurocris.org/cerif/main-features-cerif
[9] https://www.iso.org
[10] https://www.dissco.eu/data-management-plan-for-dissco/
[11] https://inspire.ec.europa.eu/theme/ef

ANAEE	• XML, RDF
DISSCO	• JSON • XML • RDF
ICOS-Ecosystem	• OWL
I2. (Meta)data use vocabularies that follow FAIR principles. *REC. Ensure convergence on vocabularies that follow the FAIR principles.*	
ICOS-Marine	• Register ICOS schemas with schema.org • Develop controlled vocabularies applying the relevant standards • Publish controlled vocabularies in the relevant registry • NERC, CF, EDMO, ORCID vocabularies
LifeWatch	• Register with FAIRsharing.org
Euro-Argo	• Linked data vocabulary service
SeaDataNet	• Incorporate vocabularies, e.g. NERC
EISCAT	• ESPAS ontology
ICOS-Atmospheric	• Publish ICOS ontology as controlled vocabulary
EPOS	• DCAT for ingestion • CERIF controlled ontology
AnaEE	• AnaeeThes, OBOE based ontology
eLTER	• EnvThes • UNECE ICP Integrated Monitoring programme common reporting format
R1. Meta(data) are richly described with a plurality of accurate and relevant attributes. *REC. Ensure that the sub domains converge on which essential attributes are required for users to decide whether their data can be reused in contexts other than the one in which the data were generated.*	
EPOS	• CERIF • DCAT-AP • ISO 19115
R1.2. (Meta)data are associated with detailed provenance. *REC. Ensure convergence on how provenance information is represented across ENVRIs.*	
ICOS-Marine	• Provenance Templates
SeaDataNet	• Structured provenance metadata, using linked data principles and SeaDataNet directories and vocabularies

ICOS-Atmospheric	• Provenance metadata in RDF
EPOS	• CERIF • PROV
R1.3. (Meta)data meet domain-relevant community standards. *REC. Ensure convergence on which standards to use to encode (meta)data.*	
ICOS-Marine	• NERC, CF, EDMO vocabularies
LifeWatch	• EMODnet, EurOBIS formats • Worms, Marine regions, NERC vocabularies
ACTRIS	• ISO 19115 metadata with WMO profile
EISCAT	• DataCite schema • ESPAS ontology
IAGOS	• WIGOS metadata profile
ICOS-Atmospheric	• ISO 19135 XML format landing page
EPOS	• ISO 19115 • CERIF • Many domain standards for metadata and data

4 Instruments for Consultancy

In this section, we present the instruments used for consultancy we identified as viable in ENVRI-FAIR. Based on the analysis in previous sections, WP7 employs these instruments to deliver consultancy to ENVRIs in the ENVRI-FAIR project with the aim to support technological convergence. As we suggested, it is sensible to prioritize this work and start with those principles that affect all critical areas first. The instruments are: Technology Demonstrators, ENVRI Knowledge Base (ENVRI KB), ENVRI-FAIR Task Forces, Subdomain Working Groups, Subdomain Knowledge Transfer,Webinars and Face-to-Face Training Events. We present these instruments next.

4.1 Technology Demonstrators

Technology Demonstrators are online digital resources that provide a possible technology solution to address a FAIR principle or possibly an aspect of a principle. They are perhaps the most important, tangible and scalable instrument for consultancy.

While Technology Demonstrators can take on various forms, including textual documents, ideally they allow for execution and hands-on experimentation with the technology. Demonstrators should be intuitive, self-explaining, and easy to follow for anyone interested in the ENVRI community, and should thus also demonstrate the solution in one or multiple context of applications related to ENVRI (e.g., rather than demonstrating how to use an ontology editor by creating an ontology about pizzas, demonstrate how to do so using a concept from one of the ENVRI domains, such as “lightning strike”).

We suggest that an ideal approach to deliver Technology Demonstrators involves Jupyter Notebooks, or similar executable resources, their versioning in a Git repository and online execution with services such as mybinder.org. This approach comes with extremely low entry barriers for executable

Technology Demonstrators, since there are no requirements to install software, update libraries, etc. on the client side. Users can thus fully concentrate on assimilating the delivered content.

While we do not exclude other forms of Technology Demonstrator delivery, we thus strongly suggest using this approach, not just for demonstrators developed within WP7 but also for those contributed in ENVRI-FAIR more widely.

4.2 ENVRI Knowledge Base

The ENVRI community creates, maintains and publishes a large amount of information, in particular also about their RIs, sites, instruments and produced data; infrastructure R&D and best practices; and the state of FAIR principles implementation, which is assessed periodically within the ENVRI-FAIR project. This information is buried in all kinds of resources (websites, deliverables, etc.) that are not easy to find, not interoperable, difficult to reuse and thus un-FAIR.

The OntoWiki[12]-based (Auer et al., 2006) ENVRI Knowledge Base[13] (KB) tries to capture critical information about the ENVRI community in machine readable form with the aim to enable machine processing of this information.

Specifically, the ENVRI KB includes information of the following types:

- ENVRI Service Portfolio, including a catalogue of ENVRI services developed in the ENVRIplus project. The new development of FAIR data services will also be included.
- Research infrastructures of the ENVRI community, including their label, acronym, description, homepage, and research domain.
- ENVRI research domains, namely Atmosphere, Marine, Ecosystem and Solid Earth.
- FAIR Principles, for which we reuse the vocabulary[14] developed by Tobias Kuhn and Michel Dumontier.
- FAIR-ness Assessments conducted within ENVRI-FAIR (Magagna et al., 2020).
- Technology Demonstrators, in particular their location on the Web, name, author and creation date.

Figure 1 provides a schematic overview of the Information Types currently included in the ENVRI KB. Naturally, the KB can be extended with additional types and the information currently available for any existing type will grow as ENVRI-FAIR develops new Technology Demonstrators, performs another iteration of FAIR-ness Assessment, etc. It can be expected that additional information such as the description of sites and experimental facilities or the list of acquired variables will come from RIs.

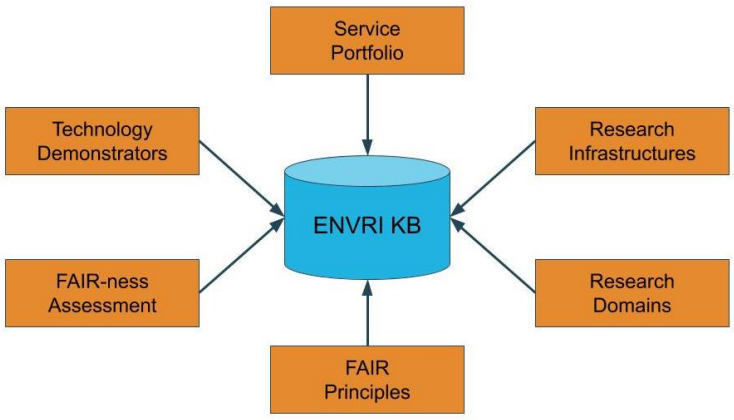

Figure 1: The Information Types currently included in the ENVRI KB.

[12] http://ontowiki.net/
[13] https://envrifair1.test.fedcloud.eu/OntoWiki/
[14] https://w3id.org/fair/principles/terms/FAIR-Vocabulary

Figure 2 depicts the relationships between three Information Types, namely the FAIR Principles, FAIR-ness Assessments and Technology Demonstrators. The reason why these are particularly important is because FAIR-ness Assessments provide information about past and current development in FAIR principles implementation by ENVRIs, and thus allowing to follow and reconstruct the state over time as well as supporting the identification of gaps. Technology Demonstrators provide information about possible technology solutions for the implementation of FAIR principles. Finally, FAIR Principles ground important vocabulary needed by the other Information Types to ensure that data are linked and interoperable. By linking these three Information Types it is possible to support users in discovering technology solutions for identified gaps.

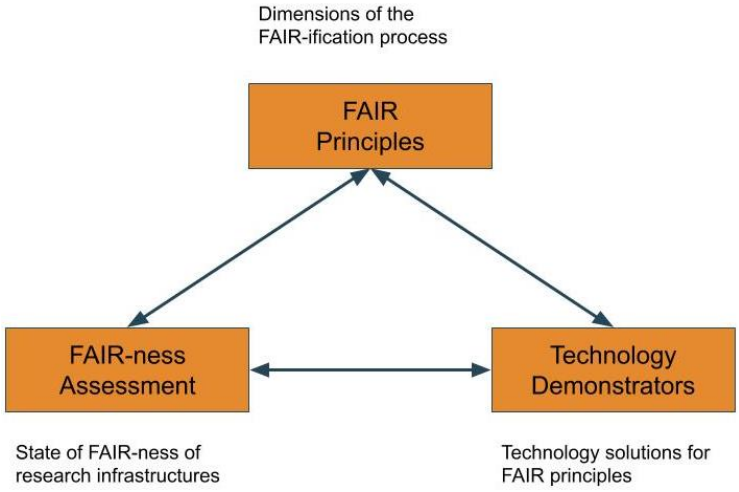

Figure 2: The relationship between three Information Types.

4.3 ENVRI-FAIR Task Forces

In 2019, ENVRI-FAIR and specifically WP5 decided to initiate Task Forces to advance shared understanding and development around the following six themes:

1. ENVRI Catalogue of Services
2. ENVRI (VO) AAI implementation
3. PIDs, identification types and registries
4. Triple stores and data storage certification
5. Licenses citation and usage tracking
6. User oriented cross-domain demonstration cases

All RIs were requested to have representation in all six themes. The aim was to transfer knowledge and ultimately converge on technology solutions in these thematic areas across the ENVRI-FAIR project partners. The Task Forces were initially given a 12-month lifetime, and asked to provide their outputs and recommendations by the end of 2020.

During Q1 2020, Task Force Leads developed Charters detailing the work plan of individual Task Forces and subsequently started the work. Given COVID-19, discussion occurred exclusively virtually and primarily in conference calls.

In general, from the point of view of this WP, these Task Forces are understood as instruments for consultancy, platforms for knowledge transfer and for identifying viable technology solutions, and thus support convergence.

Of particular interest here, is the work and development within TF1 and TF4. The ENVRI Catalogue of Services is an important component toward cluster-wide FAIR-ness and relies on convergence for metadata about services, i.e. cluster-wide agreement on what metadata about services should be captured and how these metadata are represented and exchanged. TF4 is particularly relevant to convergence on metadata representation and exchange protocols, and the technology solutions on triple

stores and more broadly technologies of the W3C Semantic Web Activity (RDF Schema and the Web Ontology Language) are key in particular to interoperability, since they are de facto standard formal languages for knowledge representation.

4.4 Subdomain Working Groups

Individual subdomains, e.g. the atmospheric subdomain, have initiated their own Working Groups to advance shared understanding and development around themes of particular relevance to the subdomain. Such WGs are an additional instrument for consultancy to catalyse convergence on technology solutions, especially where convergence matters most.

The advantage of subdomain WGs is that they can address issues of particular interest and urgency to the subdomain and do so at the necessary level of technical and domain-specialized detail, which is typically not possible at the ENVRI cluster level, where domain specifics often need to be abstracted in order to reduce complexity. Hence, this instrument is particularly important also to develop solutions for those FAIR principles that are domain specific (e.g., R1.3).

4.5 Inter-Subdomain Knowledge Transfer

With this instrument, experts in a subdomain, with expertise in a technology, transfer their knowledge to experts of another subdomain. This productive form of consultancy is related to the ENVRI-FAIR Task Forces since it occurs to some extent in Task Forces. However, subdomain knowledge transfer can also occur ad hoc, outside the organized realm of ENVRI-FAIR simply by two or more people knowing about the expertise of another person and for them to join forces by sharing knowledge about technology, organizing Webinars, and even sharing code. The disadvantage of this instrument is little scalability, especially in contrast to Technology Demonstrators, since the instrument relies on peer interaction.

4.6 Webinars and Face-to-Face Training Events

Webinars are another important instrument for consultancy. Being virtual meetings, they are inexpensive, and can be organized often and relatively quickly. They are good instruments to form ad hoc groups to discuss or present technology solutions for a particular problem. Similarly to Subdomain Knowledge Transfer, their downside is little scalability. However, the issue can be significantly mitigated by recording Webinars and making all related materials available for later use by project partners, and (if desired) by a wider community.

This instrument is closely linked to WP6 training activities and proposals for Webinars can emerge from discussions, requirements or results developed in Task Forces or subdomain Working Groups. In this sense, Webinars are an instrument to discuss or demonstrate technology solutions broadly across ENVRIs, including RIs not represented in ENVRI-FAIR, with the potential to build shared understanding and convergence beyond subdomains at the cluster level.

Finally, face to face meetings are an additional important instrument for consultancy used in ENVRI-FAIR to catalyse convergence on technology solutions where it matters most. Similarly to other instruments, such physical meetings also suffer from little scalability.

Of particular interest are training events during F2F meetings, e.g. ENVRIweek. Similar to Webinars, this instrument is thus closely linked to WP6 training activities, which are generally organized by partners involved in WP6. However, the content for such training activities and thus the priorities emerge from the ENVRI community and needs identified in Task Forces or subdomain Working Groups.

In contrast to Webinars, F2F Meetings are naturally more resource intense in organization. Moreover, following the 2020 ENVRIweek, COVID-19 has effectively made this instrument impracticable. Time will tell for how long.

5 Instruments in Action

The instruments presented in Section 4 have largely been used in practice to deliver consultancy on technology solutions. In this section, we present how the instruments have been used and their results so far.

5.1 Technology Demonstrators

So far, we have developed two Technology Demonstrators as a proof-of-concept for how this instrument can be implemented in practice. As we explained above, these demonstrators are discoverable via the ENVRI KB. As we suggested, both demonstrators adopt Jupyter Notebook and mybinder.org for demonstrator implementation and delivery and showcase the technology solution in the context of ENVRI research domains.

The Provenance Demonstrator was developed within WP7 and showcases the implementation of machine readable provenance using PROV-O (Lebo et al., 2013) for three example datasets by NILU and following a published provenance model[15] (the automatic Quality Control path for "Online Observations"). The demonstrator can be executed by following the link https://mybinder.org/v2/gh/envri-fair/provenance-demonstrator.git/master and is published at https://github.com/envri-fair/provenance-demonstrator. The demonstrator was prepared for the WP8 Paris meeting on December 11-12, 2019, which was held as a virtual meeting. The Provenance Demonstrator was presented on January 27, 2020 through a Webinar in the context of a WP8 conference call. The feedback by RI representatives was very positive, one representative suggested that the demonstrator made the use of PROV-O in practice very tangible. This feedback was early encouragement and support for the form chosen to deliver Technology Demonstrators as instruments for consultancy in ENVRI-FAIR.

The Vocabularies Demonstrator was contributed by BODC and showcases how to use SPARQL to query vocabulary published by the NERC Vocabulary Server. It is a practical, hands-on and executable example for how both human and machine agents can formulate SPARQL queries to (programmatically) extract information about vocabulary published using a SPARQL endpoint. The demonstrator can be executed by following the link https://mybinder.org/v2/gh/envri-fair/vocabularies-demonstrator.git/master and is published at https://github.com/envri-fair/vocabularies-demonstrator.

5.2 ENVRI Knowledge Base

The ENVRI KB not only supports the discovery of information about Technology Demonstrators, but links such information with FAIR-ness Assessments and FAIR Principles (Zhao et al., 2019). Building on the ENVRI KB, WP7 developed a first prototype for a specialized Web-based user interface that supports the discovery of gaps in FAIR principle implementation at the granularity of RI repositories and the discovery of possible technology solutions to address such gaps. The user interface is available at the following address https://envri-fair.github.io/knowledge-base-ui/ and Figure 3 is a screenshot of the application.

[15] https://bit.ly/2FA9Ve4

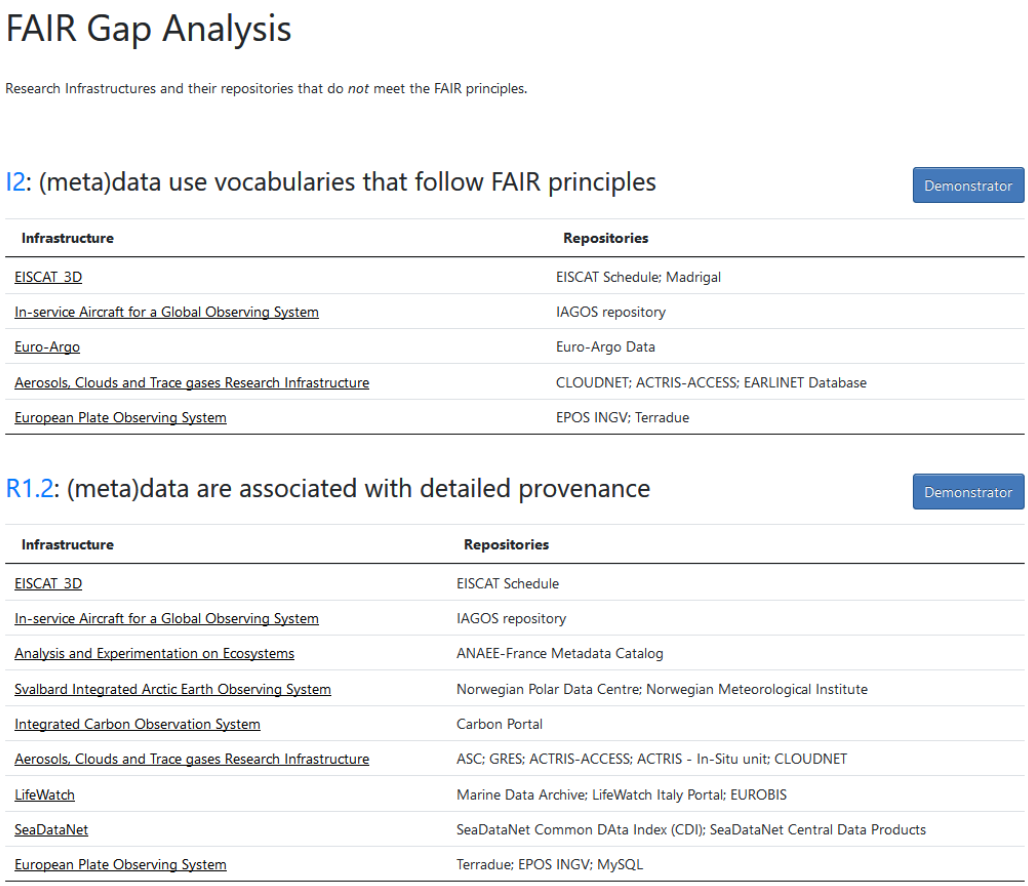

Figure 3: Screenshot of the Web-based user interface for discovery of FAIR-ness gaps at the granularity of RI repositories and corresponding Technology Demonstrators.

This user interface is dynamically created by querying the ENVRI KB using the OntoWiki SPARQL endpoint. Indeed, by modifying the FAIR-ness Assessment of a repository of a particular RI—which is functionality natively supported by OntoWiki—for instance the information on whether or not the repository has machine readable provenance information, the interface automatically adapts to either include or exclude the corresponding repository under the relevant FAIR principle (R1.2 in case of provenance). By selecting an RI, the user interface presents a summary view for the RI.

Given the ENVRI KB, it is relatively straightforward to build such specialized user interfaces that merely present KB data in a manner suitable to satisfy human information needs. We argue that such high level applications are powerful instruments for consultancy, in particular because they easily scale to cluster level demand, since RI representatives can themselves discover gaps in FAIR principle implementations and, importantly, discover and learn about possible technology solutions for the discovered gaps. Such a tool can thus reduce the need for direct human interaction with technology experts in ENVRI-FAIR, enable independent learning and have thus the potential to streamline convergence on technology solutions.

Expert users have the possibility to navigate the ENVRI KB directly at the level of OntoWiki. OntoWiki supports the browsing as well as editing of content. Hence, through OntoWiki it is in principle possible to modify and possibly fix errors in ENVRI KB content, e.g. errors in conducted FAIR-ness Assessments. It is also possible to use OntoWiki to browse the FAIR-ness Assessment information on specific repositories of an RI and discover how the FAIR-ness Assessment relates to FAIR principles and thus to Technology Demonstrators. While technically possible to browse information in this manner, using specialized user interfaces that present the same information in digested manner is arguably more efficient, for both casual and expert users who have information needs that the specialized user interface is designed to satisfy.

The ENVRI KB, the underlying technologies and applications it enables, such as the described specialized Web-based user interface, were presented to the ENVRI community in a Webinar

organized by WP6 on June 3, 2020 with a talk entitled "FAIR the smart way: Introducing the ENVRI Knowledge Base"[16].

5.3 ENVRI-FAIR Task Forces

Since its inception, ENVRI-FAIR Task Force 1 (TF1) has taken as its starting point the ENVRIplus recommendation of using CKAN as used in EUDAT and CERIF as data models for the common catalog. TF1 agreed quickly that CKAN was a subset of CERIF so the work has concerned how to map from various metadata formats in the subdomains (for example in solid earth subdomain 17 metadata formats are converted to/from CERIF). CERIF is a rich superset of the other considered metadata formats, has formal syntax and declared semantics with an in-built ontology including capability for multilinguality and crosswalks among terminology bases. CERIF uses n-tuples to represent assertions of FOL (First Order Logic) and furthermore divides the world of interest into base entities (things in the real world) and linking entities (relationships between them). The relationships not only record the qualification of the relationship in role, but also in temporal duration. This provides built-in provenance and curation management. The other key aspect of the work has been concerned with defining APIs to manage the catalog content and relate the catalog content to scientific data processing. Inevitably, considerations of AAAI—and especially authorisation—have been discussed. The Research Data Alliance (RDA) Metadata Interest Group is relevant with their work on defining a metadata element set for interoperation where each element has an internal structure and declared semantics (Bailo, 2020).

For both triple stores and data storage certification schemes, most importantly Core Trust Seal[17], the ENVRI-FAIR Task Force 4 (TF4) has had a focus on collecting experience reports about these technologies from RI representatives who have already used them in production. On triple stores, BODC, ICOS and ANAEE have demonstrated experiences and plans to (or already) deliver(ed) short experience reports at the time of writing this deliverable. On certification, IFREMER and BODC have compiled experience reports while EMSO, ICOS and IAGOS planned to do so during 2020/21. These experience reports are useful documentation for those RI representatives that are only planning to employ the corresponding technologies and can leverage the experiences of others in their decision making, possibly learning and ultimately deciding to adopt similar approaches and thus converge on common technology solutions. At the time of writing this deliverable, the following RIs have plans to employ triple stores: IFREMER for vocabulary management, strongly aligned to approaches developed by BODC; EMSO, which plans to carry out a cost-benefit analysis of triple stores before making further decisions; IAGOS; SIOS, possibly for vocabulary management; LifeWatch; ATRICS, possibly for provenance management. The following RIs have plans to certify their data repositories: EMSO, ICOS, IAGOS, LifeWatch, ANAEE and EISCAT_3D, whereby ANAEE focuses on Core Trust Seal, specifically.

5.4 Subdomain Working Groups

The atmospheric subdomain has organized some of its work in Working Groups. The WG on Semantic Search is of particular interest here. This group aimed at reviewing the state of the art in semantic search interfaces and implementing a prototype for semantic search motivated by a use case based on real needs in the atmospheric domain (ground based in situ observations of aerosol_absorption_coefficient using filter_absorption_photometer). To enable powerful semantic search, structured data has to be published in a way that allows for interlinking of resources. For this purpose, the NetCDF dataset was transformed to RDF using an *ad hoc* script in Jupyter Notebook and then enriched with a semantic description based on different semantic artefacts (WIGOS, CF Standard Name and ACTRIS vocabulary), aligned to each other for this use case.

5.5 Webinars and Face-to-Face Training Events

While COVID-19 has brought F2F Meetings to a complete halt following national lockdowns mid March, with the training on terminologies[18] delivered at ENVRIweek 2020 (Hellström et al., 2020) we

[16] Materials via https://envri.eu/online-webinar-fair-the-smart-way-introducing-the-envri-knowledge-base/
[17] https://www.coretrustseal.org/
[18] https://envri.eu/event/envri-fair-first-training-event/

also had an opportunity to prototype this form of instrument for consultancy. Similar to Webinars, the preliminary results are encouraging as they suggest that this form, too, is effective at sharing knowledge about technology solutions in the ENVRI community and thus support convergence on critical FAIR principle implementations.

6 Discussion

WP7 is primarily concerned with supporting the implementation of common technology solutions with the aim to ensure that (meta)data are FAIR at the level of the ENVRI cluster.

With this aim in mind, we started out suggesting that there are three critical areas in which ENVRIs should implement common technology solutions, namely: (meta)data Catalogue, Syntax and Semantics. We then suggested that the FAIR principles differ in their relevance to these three critical areas. Hence, ENVRI-FAIR work aimed at ensuring compliance with the FAIR principles can be prioritized accordingly. The resulting priority list serves as a guide which principles should be tackled first.

We support the implementation of common technology solutions through consultancy. To deliver consultancy, it is critical to gain understanding for the implementations ENVRIs plan or have already completed. Moreover, we rely on a number of instruments used to deliver consultancy, which we presented. As presented in Chapter 5, we have used these instruments in practice to deliver consultancy.

The approach proposed here for ENVRI-FAIR consultancy is bottom-up and pragmatic. It reflects three observations. First, not all FAIR principles need common technology solutions across ENVRIs. Indeed, there are principles that can be implemented using a plurality of technology solutions, and they can coexist without negatively affecting FAIR-ness in the ENVRI cluster and potentially beyond. Hence, we developed the model of critical areas in which ENVRIs should implement common technology solutions. Second, implementing the FAIR principles is not trivial and converging on technology solutions across ENVRIs is even harder. Hence, we analysed the FAIR principles for their relevance to critical areas and prioritized the FAIR principles to guide consultancy on those principles first that are most relevant to the critical areas for technology solution convergence. Third, effective consultancy relies on strong interaction with ENVRIs to understand their needs and requirements. Hence, we developed a range of instruments designed to deliver consultancy directly in physical or virtual meetings or indirectly through Web-based tools.

So far we have tested all instruments for their viability in practice. However, it is too early to evaluate their effectiveness in meeting the primary objective of technology solution convergence.

The most clear results are for principle *R1.2. (Meta)data are associated with detailed provenance*, for which we have delivered a technical demonstrator and webinars, in particular in the Atmospheric subdomain. Feedback from both ACTRIS and NILU suggest that these activities had an impact on their decision to evaluate the use of triple stores to manage provenance information, and represent provenance information using PROV-O. This sort of feedback is encouraging and may suggest that the approach developed and pursued here could lead to successful convergence across RIs and subdomains in ENVRI-FAIR.

However, reaching technology solution convergence across ENVRIs is an enormous task, for which ENVRI-FAIR only has the resources to develop a viable path. This is what WP7 has embarked on and is reporting in this deliverable. We will build on this work to further catalyse technology convergence, by developing additional Technology Demonstrators and delivering consultancy by means of the other instruments. However, ultimately it is the RIs that need to walk the proposed path, and do so well beyond the ENVRI-FAIR project. Even during ENVRI-FAIR, WP7 heavily relies on contributions by RIs, especially on Technology Demonstrators and Webinars. Hence, it is our aim to identify ENVRIs and their solutions that could be transferred into Technology Demonstrators or communicated to other ENVRIs in Webinars. We use especially Task Forces to identify such opportunities and motivate ENVRIs to contribute.

To support convergence on FAIR enabling resources, WP5/7 will reuse the concept of FAIR Implementation Profiles (FIPs) as proposed by GO-FAIR[19] (Schultes, 2020). The GO-FAIR development team, in which ENVRI-FAIR members take part, is creating an online platform (FAIR Convergence Matrix) that uses a core ontology (Convergence Matrix Ontology[20] and GO FAIR Core Ontology[21]) to compile, for any Community of Practice (i.e., the columns in the matrix) and for each FAIR principle, an inventory of their FAIR implementation Choices and Challenges selected from a list of existing or proposed digital resources (i.e., the rows in the Matrix). Each column of the Matrix comprises a profile (FIP) characterizing how the corresponding community has chosen to implement FAIR and as such it is a unique signature representing the community. FIPs can be used as a powerful accelerator of convergence on FAIR standards and technologies.

7 Conclusions

We presented the ENVRI-FAIR approach to consultancy by ICT experts with research infrastructure managers and developers, aimed at technology convergence across the cluster of environmental research infrastructures (ENVRI) in order to ensure FAIR (meta)data across the cluster.

We identified three critical areas—(meta)data Catalogue, Syntax and Semantics—in which technology convergence is recommended and prioritized the FAIR principles for their relevance to the three critical areas. For the principles relevant to all three critical areas, we analysed the landscape of implementations across ENVRIs to better understand the heterogeneity of solutions pursued and the possible difficulties the cluster may face to ensure FAIR (meta)data at the subdomain and cluster levels. We then presented the range of instruments for consultancy at our disposal and how we have used them so far to support ENVRIs and their FAIR principle implementations. First results can be seen in implementation of provenance, where we have had encouraging signals of multiple RIs aiming to evaluate the use of the same technologies for provenance information representation, curation and publication.

We will now build on the developed framework presented here to organize the consultancy work and try to catalyse convergence where needed most. This task will continue to closely interact with the other tasks in the WP7 to provide valuable input to the knowledge base (T7.2) and the development support (T7.3) during the consultation. For training activities, we will do so in collaboration with WP6. Clearly, the task of cluster-level technology convergence for effective and efficient FAIR principles implementation is enormous and ultimately relies on the RIs, within ENVRI-FAIR and beyond.

8 References

Auer, S., Dietzold, S., & Riechert, T. (2006). OntoWiki–a tool for social, semantic collaboration. In International Semantic Web Conference (pp. 736-749). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11926078_53

Bailo, D. (2020). MS18 Design of the service catalogue. Zenodo. https://doi.org/10.5281/zenodo.3955526

CODATA-ICSTI Task Group on Data Citation Standards and Practices (2013). Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. Data Science Journal, 12, pp.CIDCR1–CIDCR7. http://doi.org/10.2481/dsj.OSOM13-043

Fiebig, M., Lund Myhre, C., Boulanger, D., Rivier, L., Vermeulen, A., Häggström, I., … Tukiainen, S. (2020). ENVRI-FAIR D8.3 Atmospheric subdomain implementation plan (Version 1). Zenodo. https://doi.org/10.5281/zenodo.3885239

[19] https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/
[20] https://github.com/go-fair-ins/GO-FAIR-Ontology/blob/master/Diagrams/Matrix.pdf
[21] https://github.com/go-fair-ins/GO-FAIR-Ontology/blob/master/Diagrams/Core.pdf

Hellström, M., Johnsson, M., Fiore, N., Vaira, L., & Konijn, J. (2020). ENVRI-FAIR MS23: First face-to-face FAIR data management training event for ENVRI data centre staff completed. Zenodo. https://doi.org/10.5281/ZENODO.3906125

Koers, H., Bangert, D., Hermans, E., van Horik, R., de Jong, M., & Mokrane, M. (2020). Recommendations for Services in a FAIR Data Ecosystem. Patterns, 1(5), 100058. https://doi.org/10.1016/j.patter.2020.100058

Lebo, T., Sahoo, S., McGuinness, D. (2013). PROV-O: The PROV Ontology. W3C Recommendation. https://www.w3.org/TR/prov-o/

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

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

Schultes, E.,Magagna B., Hettne K. M., Pergl R., Suchánek, M., Kuhn, T. (2020). Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence, CMOMM4FAIR workshop, co-located to ER 2020 Vienna, to appear

Thijsse, P., Schaap, D., Exter, K., Vermeulen, A., Pfeil, B., Carval, T., … Rodero, I. (2019). ENVRI-FAIR D9.2 Marine subdomain implementation plan (Version 1). Zenodo. https://doi.org/10.5281/zenodo.3885326

Wilkinson, M., Dumontier, M., Aalbersberg, I. ... (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Zhao, Z., Liao, X., Martin, P., Maduro, J., Thijsse, P., Schaap, D., Stocker, M., Goldfarb, D., & Magagna, B. (2019). Knowledge-as-a-Service: A Community Knowledge Base for Research Infrastructures in Environmental and Earth Sciences. 2019 IEEE World Congress on Services (SERVICES). 2019 IEEE World Congress on Services (SERVICES). https://doi.org/10.1109/services.2019.00041

9 Appendix 1: Glossary

ACTRIS	Aerosols, Clouds, and Trace gases Research InfraStructure network
Catalogue (Metadata)	A collection of metadata, usually established to make the metadata available to a community. A metadata catalogue has an access service.
CERIF	Common European Research Information Format
CODATA	Committee on data for Science and Technology
DANUBIUS	The international center for Advanced studies on river-sea systems
DiSSCo	Distributed Systems of Scientific Collections
DOI	Digital Object Identifier
EISCAT	EISCAT Scientific Association
EMSO	European Multidisciplinary Seafloor and Water Column Observatory
ENVRI	(1) The ENVRI Community of Environmental Research Infrastructures. (2) FP7 project on Implementation of common solutions for a cluster of ESFRI infrastructures in the field of Environmental Sciences.
EPOS	The European Plate Observing System
EURO-ARGO	European ARGO programme (ARGO are a type of marine survey device)
H2020	Horizon 2020, European level research funding scheme
IAGOS	In-service Aircraft for a Global Observing System
ICOS	Integrated Carbon Observation System
ICSU	The International Council for Science
INSPIRE	Infrastructure for Spatial Information in the European Community

Knowledge Base	(1) A store of information or data that is available to draw on. (2) The underlying set of facts, assumptions, and rules which a computer system has available to solve a problem.
LifeWatch	European e-Science infrastructure for biodiversity and ecosystem research
LTER	The Long-term Ecological Research Network
Metadata	Data that describes other data. Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier.
NetCDF	A file format
Ontology	(In computer science and information science) an ontology is a formal naming and definition of the types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain of discourse.
OWL	Web Ontology language
Provenance	The pathway of data generation from raw data to the actual state of data
RDA	Research Data Alliance
RDF	Resource Description Framework
RI	Research Infrastructure
Semantics	The encoding of meaning using a formal language.
SIOS	Svalbard Integrated Arctic Earth Observing System
Syntax	In computer science, the syntax of a computer language is the set of rules that defines the combinations of symbols that are considered to be a correctly structured document or fragment in that language.
W3C	World Wide Web Consortium
WP	Work Package