MISSIONE 4
ISTRUZIONE
RICERCA

# Fair Implementation Profiles (FIPs) Report on the first release.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

| | |
|---|---|
| Deliverable no. | D2.8 |
| Work package | WP2 – ACCESS TO FACILITIES, FAIR DATA AND RELATED SERVICES |
| Intermediate Objective | |
| Deliverable type | X Document, report<br>☐ Websites, patent filings, videos, etc.<br>X Other: please specify FAIR Implementation Profile case study |
| Dissemination level | X Public<br>☐ Restricted |
| Estimated delivery (bimester) | B4 |
| Actual delivery date | |
| Authors (Partner-OU) | Andrea Tarallo, Ilaria Rosati (CNR-IRET Lecce) |
| Reviewed by | Gianmarco Ingrosso, Carmela Cornacchia |
| Comments | |

# INDEX

# 1 LIST OF ACRONYMS

FIP: FAIR Implementation Profile

FER: FAIR Enabling Resource

FIC: FAIR Implementation Community

RI: Research Infrastructure

# 2 INTRODUCTION AND PURPOSE OF THIS DOCUMENT

The Italian Integrated Environmental Research Infrastructures System (ITINERIS) Project started in November 2022, and it will build the Italian Hub of Research Infrastructures (RIs) in the environmental scientific domain providing access to data and services and supporting the Country to address current and expected environmental challenges. ITINERIS coordinates a network of national nodes from 22 RIs (18 from the environmental domain, 2 from agri-food with strong link with the environment and 2 from the Physical Sciences and Engineering domain, supporting services for the marine domain). The participating RIs are:

- ESFRI Landmarks (Italian nodes): ACTRIS, EMSO, Euro-Argo, ICOS and LifeWatch, (environmental domain) and ANAEE (Health&Food domain);
- ESFRI projects (Italian nodes): DANUBIUS-RI, DiSSCo, e-LTER, (environmental domain), and EMPHASIS and EUIBISBA (Health&Food domain);
- EU RIs: ECORD, EUFAR, Eurofleets, JERICO and SIOS (environmental domain);
- National RIs: ATLAS, CeTRA, Laura Bassi, and SMINO (environmental domain), Geosciences, and LNS (Physical Sciences and Engineering domain) those last as support services to the marine domain.

The deliverable 2.8 has been planned to be produced along with the deliverable 2.7, State of the Art review of FAIR-enabling best practises, to assess the technological choices for the implementation of RIs FAIRness by using the FAIR Implementation Profile (FIP) Wizard[1]. Following the extensive review carried out with the deliverable 2.7, a great variety emerged in the capacity and autonomy of data management and governance aspects. Indeed, some RIs are in an initial construction phase, while others have more than 10 years of operation and an advanced FAIR readiness level. In addition, their governance structures show a high degree of heterogeneity. The ITINERIS project brings together nationally funded RIs, European ESFRI and non-ESFRI RIs, that in turn can be structured with distributed national nodes with varying degrees of independence. Considering all these aspects, we decided to design a roadmap to obtain a full release of the FIPs before the end of the project. The roadmap is outlined in this document, in which we present:

1. a review of the existing FIPs available by the FIP Wizard for the RIs that are relevant for the ITINERIS project;
2. the release of the FIP for LifeWatch Italy, the Italian distributed centre of LifeWatch ERIC;
3. a plan of implementation for future activities including a training to the dedicated personnel of the RIs to guide them towards the release of the FIP.

[1] https://fip-wizard.ds-wizard.org/wizard/dashboard

# 3 WHAT IS A FIP

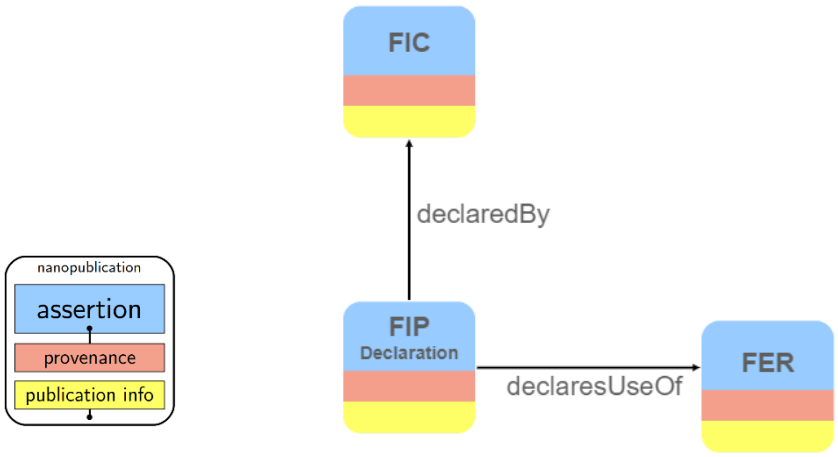


***Figure 1.*** *Simplified view of the FAIR Implementation Profile (FIP) ontology (modified from [https://osf.io/ymjwp](https://osf.io/ymjwp)). Each box represents a nanopublication. The colour-code for the nanopublication is to graphically show the three minimum parts in which a nanopublication is divided: assertion (blue), provenance (red), and publication info (yellow). FIC stands for FAIR Implementation Community, FER for FAIR Enabling Resource.*

In its essence, a FIP is a list of declaration of using certain technology choices, known as FAIR Enabling Resources (FERs), selected by a specific community of practice (e.g. RI, Data Repositories managers, etc.), called FAIR Implementation Community (FIC), to implement one or more FAIR Guiding Principles (Figure 1). All these assertions are represented by nanopublications (for more information about nanopublication visit the FIP Wiki[2]). FIP was conceived during the ENVRI-FAIR project[3] (see also Schultes et al., 2020) to provide a means to track the progress of FAIR data services made by the involved RIs. In addition, the FIP was also used to measure the grade of convergence among them, in simple words the number of common FERs between two RIs. A FIP consist of 21 questions that are compiled by a responsible person for the (meta)data management in a specific FIC through the FIP Wizard[4]. When a question is answered with an existing resource, it is considered a FER, indicating that a specific FAIR sub-principle is fulfilled.

# 4 REVIEW OF THE EXISTING FIPS

In order to make the review of the existing FIPs, we downloaded all the FIPs from the nanopublication API query[5] that assembles the results as a .csv file from a sequence of SPARQL queries. This customised query allows the direct download of all the available FIPs that are published through the FIP Wizard[3]. The table was subsequently cleaned by selecting only the RIs participating in ITINERIS plus the available SeaDataNet FIPs (i.e. SeaDataNet-

[2] https://gofair-foundation.github.io/fip/MintingNanopublications.html
[3] https://envri.eu/home-envri-fair/
[4] https://fip-wizard.ds-wizard.org/
[5] https://github.com/peta-pico/dsw-nanopub-api/blob/main/tables/new_matrix.csv

CDI and SeaDataNet-Sextant). SeaDataNet has been analysed along the ITINERIS RIs because the marine subdomain RIs make use of such infrastructure. We took into consideration the most updated version of each FIP (last visit 09.10.2023). Only the FERs that were declared as "available" and "in use" by each declaring community have been selected. From this cleaned table we produce a new matrix[6]. Subsequently, the matrix was used to provide a cross table to show the number of FERs in common between each couple of RIs belonging to ITINERIS as a mean of level of convergence (Figure 5). In the cross table shown in Figure 5 SeaDataNet FIPs were excluded because, even if it is used by the "marine" community, it is not an ITINERIS RI. RIs represented by more than one community (i.e. presented more than one valid FIP, such as ACTRIS ERIC, AnaEE ERIC, and LifeWatch ERIC)) were merged to build the cross table.

## 4.1 Declared FAIR Enabling Resources

According to the FIP Ontology, a FIC is a "*self-identified collection of people and/or organizations with the aim to implement the FAIR Principles*"[7]. The definition is rather vague on purpose because it needs to include all the shadows between well-characterised communities, such as data infrastructure, to more faded ones, such as the community of people that contribute to a specific research topic, for instance the scientists that study the effect of climate changes on community composition. In the case of ITINERIS, the FIC can be the community of a RI or the community of its thematic/distributed centres, and the FIP is the detailed list of the technological implementations (i.e. what FERs they implement) adopted to satisfy each of the FAIR sub-principles. As a result, the number of communities taken into consideration in this report is higher than the number of RIs considered for this analysis. The final number of communities is 18: six Aeorosol, Clouds and Trace Gases Research Infrastructure data centre unit of ACTRIS[8] (Cloud remote sensing data centre unit "ACTRIS_CLU", Data discovery, virtual access and Services unit "ACTRIS_DVAS", Trace gases remote sensing data centre unit "ACTRIS_GRES", In-Situ data centre unit "ACTRIS_InSitu", Aerosol remote sensing data centre unit "ACTRIS-ARES", Atmospheric simulation chamber data unit "ACTRIS-ASC"), ICOS, sios, the Euro-Argo ERIC Global Data Centres "ArgoGdac", EMSO, the The Common Data Index service of SeaDataNet "SeaDataNet-CDI", the product catalogue service and system for SeaDataNet "SeaDataNet-Sextant", Anaee, AnaEE_CREA (the centre responsible for the Data & Modeling Centre), DiSSCo, eLTER-RI, LWERIC_Ecosystem, which correspond to the digital services offered by the LifeWatch ERIC Service Centre, and lw-marine, which correspond to the digital services offered by the Belgian node. The communities represent nine ITINERIS RIs (ACTRIS ERIC, AnaEE ERIC, EMSO ERIC, Euro-Argo ERIC, ICOS ERIC, LifeWatch ERIC, DiSSCo, eLTER and SIOS) plus SeaDataNet. FIPs from RIs belonging to the geosphere were not available.

The total number of FERs declaration is 252. Some FERs may cover more than one FAIR sub-principle. For instance, like exemplified in Figure 2, a FER (e.g. Darwin Core Archive) might be declared as a knowledge representation language and as a structured vocabulary. Moreover, the FIP separate the questions for data and metadata. This means that the same

[6] https://zenodo.org/records/10808170

[7] https://peta-pico.github.io/FAIR-nanopubs/fip/index-en.html#https://w3id.org/fair/fip/terms/FAIR-Implementation-Community

[8] https://www.actris.eu/topical-centre/data-centre

FER (e.g. Darwin Core Archive in Figure 2) might be used for data and metadata FAIRness. As a result, the total number of unique declared FERs in use is 145 (Figure 3).

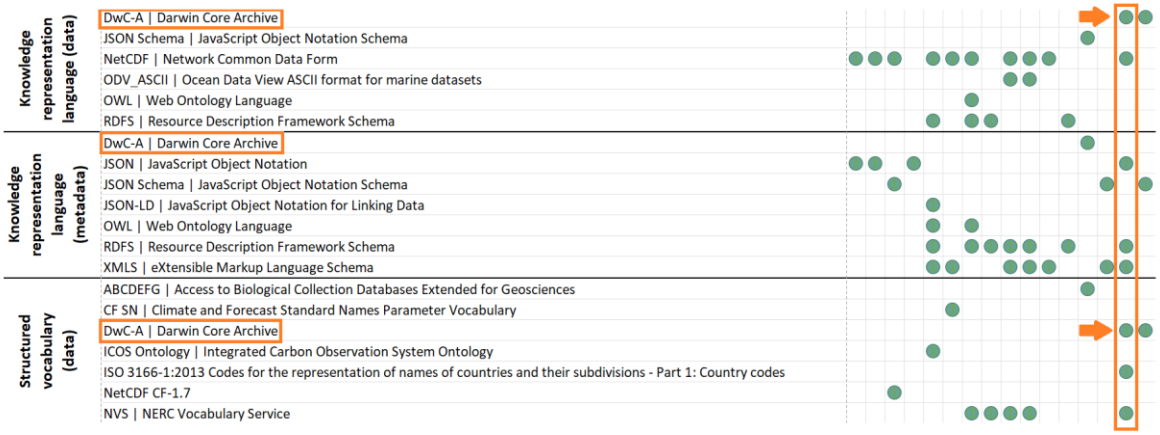


***Figure 2.*** *Snapshot of the matrix. In the orange squares is highlighted the same FER (Darwin Core Archive) that is declared by the same community (orange arrows) as a Knowledge representation model and a structured vocabulary. Another RI declared the use of the Darwin Core Archive as a Knowledge representation model for metadata. In this case, one FER is declared three times. This explain the discrepancy between FERs declaration (247) and the number of unique declared FERs in use (143).*

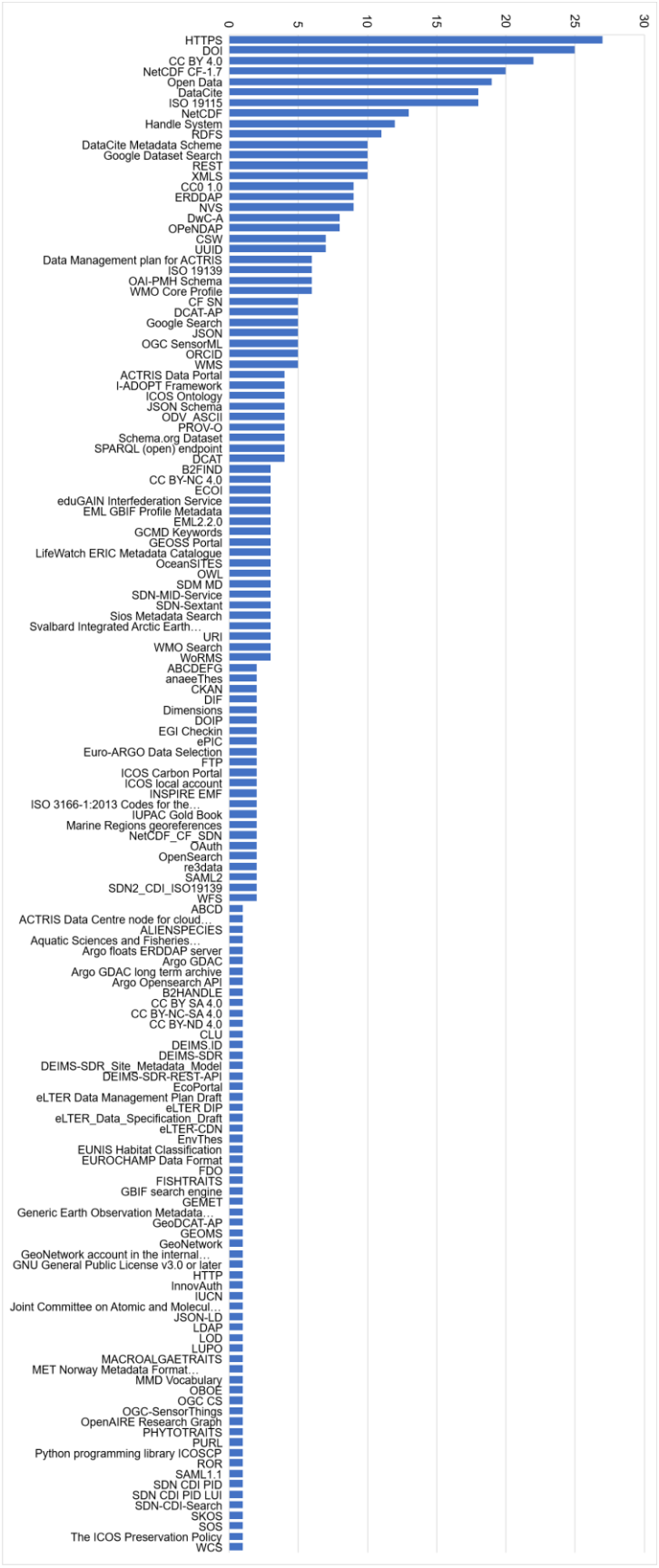


***Figure 3.*** *Histogram showing all the FAIR Enabling resources declared.*

63 FERs are used only once, with an additional 51 used fewer than 5 times (Figure 4). This data highlights the independent adoption of diverse technologies by RIs, occasionally leading to duplication of efforts in meeting overlapping FAIRness implementation needs. Consequently, this trend poses a challenge, contributing to a reduced interoperability among RIs.

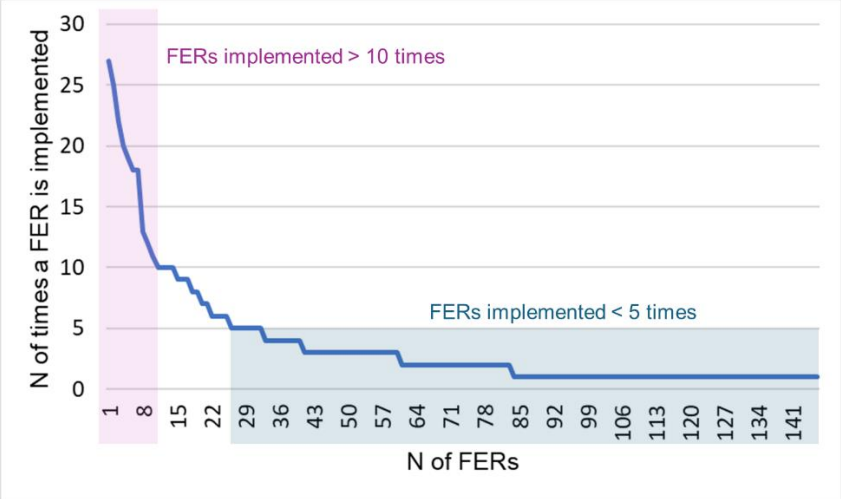


***Figure 4.*** *Number of times in which a FER is declared in all the FIP questionnaires taken into consideration (y axis) vs the number of FERs that are declared in that specific class.*

Only 10 FERs are used more than 10 times each, and are listed in Table 1. The most implemented FER is HTTPS (27 declarations) as standard protocol exchange to satisfy the FAIR sub-principle A1.1. HTTPS is the standard *de facto* for navigating the internet. It ensures secure communication by data encryption. This is essential for reliable and trustworthy data access. The second most declared is DOI (25 declarations). This can be explained by the fact that DOI is perceived by the RIs communities as a trusted infrastructure, already established more than 25 years now. It was mainly used for scholarly publications, so researchers are used to it. Additionally, it is assigned by multiple registration agencies worldwide, including those that provide DOI registration services to research data. The third most implemented FER is CC-BY licence 4.0. This is the recommended licence for the European Commission funded projects for both data and publications[9], and understandably the most used in this context.

***Table 1. The table shows all the FAIR Enabling resources declared more than 10 times, and the number of times they are used to answer a question in the FIP.***

| FAIR Enabling Resource | Times implemented | N of FIP questions in which the FER is declared |
|---|---|---|
| HTTPS \| Hypertext Transfer Protocol Secure | 27 | 2 |
| DOI \| Digital Object Identifier | 25 | 3 |
| CC BY 4.0 \| Attribution 4.0 International | 22 | 2 |
| NetCDF CF-1.7 | 20 | 7 |
| Open Data | 19 | 2 |

9 https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/open-access-obligations-horizon-europe-what-are-cc-licences-2021-11-15_en

| | | |
|---|---|---|
| DataCite \| DataCite Ontology | 18 | 3 |
| ISO 19115 \| Geographic information - Metadata | 18 | 4 |
| NetCDF \| Network Common Data Form | 13 | 2 |
| Handle System | 12 | 3 |
| RDFS \| Resource Description Framework Schema | 11 | 2 |

## 4.2 Level of convergence

In Table 2 we summarised the overall picture of the FICs and the number of FERs each FIC declared. The communities which declared more FERs are ACTRIS ERIC (70), LifeWatch ERIC (64), and ICOS (64), which are also the landmark ERIC, for which we expected a higher FAIR readiness level then the other RIs.

***Table 2. The table shows the total number of FAIR Enabling Resources that are declared by each FIP.***

| RI | Declaring community | Total FERs |
|---|---|---|
| ACTRIS ERIC | ACTRIS_CLU | 20 |
| | ACTRIS_DVAS | 19 |
| | ACTRIS_GRES | 29 |
| | ACTRIS_InSitu | 26 |
| | ACTRIS-ARES | 34 |
| | ACTRIS-ASC | 21 |
| ICOS ERIC | ICOS | 64 |
| LifeWatch ERIC | LWERIC_Ecosystem | 43 |
| | lw-marine | 29 |
| AnaEE ERIC | Anaee | 18 |
| | AnaEE_CREA | 11 |
| EMSO ERIC | EMSO | 22 |
| Euro-Argo ERIC | ArgoGdac | 39 |
| DiSSCo | DiSSCo | 16 |
| eLTER-RI | eLTER-RI | 40 |
| SIOS | sios | 33 |

To give a snapshot of the current convergence in the use of FERs among the ITINERIS RI, we built a cross matrix in which each RI' FIP is compared with all the others (Figure 5). Every time an RI has a FER declaration in common with another RI, it is counted as one. The number which is showed represent the total number of FER declarations in common between the two RIs that cross. The highest level of convergence is observed between ACTRIS ERIC and ICOS ERIC, and between ACTRIS ERIC and LifeWatch ERIC. This might be partially because those three RIs are also the RIs which declared more FERs (Table 2). ACTRIS ERIC, ICOS ERIC, Euro-Argo ERIC, and SIOS present a relative high level of convergence among them. Those RIs are all represented in the atmospheric subdomain of ITINERIS. eLTER-RI and LifeWatch ERIC stands out as to be convergent between them. Again, this reflects the fact that the two RIs belong to the terrestrial biosphere subdomain of ITINERIS. Both show a certain grade of convergence with the "atmospheric" RIs. Another

observed convergence is between EMSO ERIC and Euro-Argo ERIC. Both those RIs maintain an infrastructure for sea monitoring and are represented in the marine subdomain of ITINERIS.

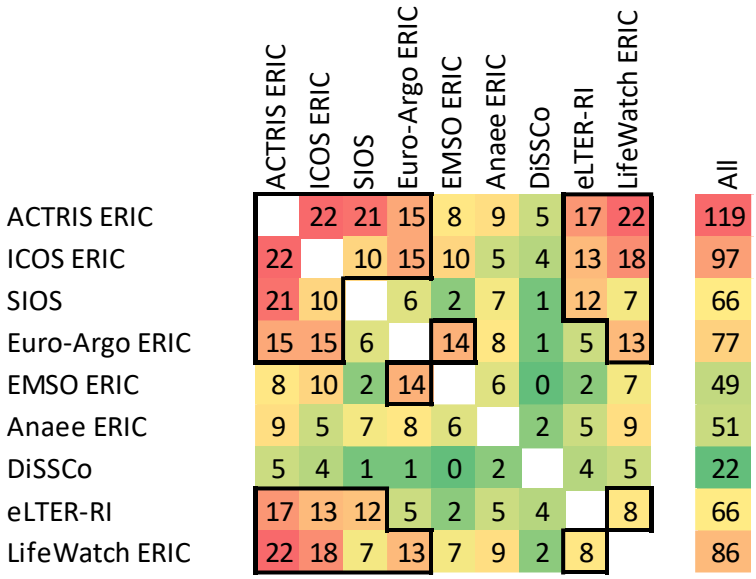


***Figure 5**. Heatmap showing the number of FERs in common between each pair of RIs.*

## 5 THE CASE STUDY OF LIFEWATCH ITALY

The LifeWatch Italy FIP was compiled in an online session by the WP2 representative for the Terrestrial Biodiversity subdomain together with the representative of the LifeWatch Italy FAIR Implementation WG in ITINERIS. Go-Fair Foundation, which is the organisation who manages and maintain the FIP Wizard tool advise that everyone that wish to compile a FIP should be do it in assistance with a facilitator. The WP2 representative was then instructed on the basic principles of the FIP in a FIP facilitator training course[10]. The LifeWatch Italy representative already knew the FIP Wizard, having participated in ENVRI-FAIR project and collaborated on the creation of the LifeWatch ERIC FIP. The FIP is publicly available[11] with a CC-BY licence in human and machine-readable formats. The total number of unique declared FERs in use is 28, while the total number of FERs declaration is 35.

## 6 STRATEGY FOR FUTURE RELEASE OF FIPS

The scheduled releases for RIs FIPs encountered two main challenges. First of all a governance issue. A limited number of national nodes indeed have a direct control on the implementation choices that are made for the RI. The national nodes that do not directly manage a digital infrastructure should update the FIP for the European RI that is not in their

[10] https://osf.io/bthf8

[11] https://fip-wizard.ds-wizard.org/wizard/projects/3ba95cc8-3174-44ab-ba0d-fe697513cea5

own hands. In this regard, the headquarter of an RI may not be necessarily interested in revising a FIP, since it is not a trivial effort. The same problems could apply when it comes to the RIs that have thematic/distributed centres placed in different countries, such as ACTRIS ERIC. Secondarily, the RIs involved in ITINERIS have varying years of development and trajectories. Some RIs have more than ten years of operation, while others are just starting, and cannot be compared in terms of FAIR readiness. In addition, some RIs, have obligations regarding FAIR data management, which are imposed by higher bodies (e.g., the ESFRI Forum for European RIs).

Looking ahead, our strategic considerations involve a comprehensive approach that extends through different levels. Regarding the RIs, we will target those with a more developed Italian node in terms of data infrastructures, such as LifeWatch Italy, as well as the national ones (ATLAS, CeTRA, Laura Bassi, SMINO, Geosciences, and LNS). We will also consider specific infrastructure components that will be developed within ITINERIS project, such as the various data centres that are planned, for instance the WP4 data centre for atmospheric data, the WP5 marine data centre, the data centre for Polar data, and others, including the ITINERIS Hub. The temporal aspects of the FIP releases for those components is tied to their developmental trajectories. Realistically, this will be achieved at the end of the project.

In this context, we envisaged a training initiative within WP3 involving the representatives from established RIs such as the RI managers, data managers, or data stewards from RIs and the various portals/hubs in ITINERIS. This comprehensive course will be designed to equip participants with the essential skills and knowledge necessary for the effective implementation of FIPs. Participants will gain hands-on experience in applying the acquired knowledge to their respective domains, ensuring a harmonized approach to FIP implementation across diverse infrastructures and data centres. This strategic initiative aims not only to empower data stewards but also to foster a community-wide adherence to FAIR principles, contributing to the overall success of the ITINERIS project.

## 7 ACKNOWLEDGEMENTS

Thanks to Barbara Magagna for her assistance in data extraction and analyses.

## 8 REFERENCES

Schultes E., Magagna B., Hettne K.M., Pergl R., Suchánek M., Kuhn T. (2020) Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In: Grossmann G., Ram S. (eds) Advances in Conceptual Modeling. ER 2020. Lecture Notes in Computer Science, vol 12584. Springer, Cham. https://doi.org/10.1007/978-3-030-65847-2_13 (preprint available at https://doi.org/10.31219/osf.io/2p85g).