MISSIONE 4
ISTRUZIONE RICERCA

State of the Art review of FAIR-enabling best practices

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

Deliverable no.	D2.7
Work package	WP2 – ACCESS TO FACILITIES, FAIR DATA AND RELATED SERVICES
Intermediate Objective	
Deliverable type	X Document, report ☐ Websites, patent filings, videos, etc. ☐ Other: please specify .....
Dissemination level Estimated delivery (bimester)	X Public ☐ Restricted B4
First delivery date	30/06/2023
Final delivery date	31/10/2023
Authors (Partner-OU)	Enrica Nestola, Gianmarco Ingrosso, Gregorio Sgrigna, Andrea Tarallo, Ilaria Rosati (CNR-IRET, Lecce)
Reviewed by	Caterina Bergami (CNR-ISMAR), Lorenzo Corgnati (CNR-ISMAR), Carmela Cornacchia (CNR-IMAA), Claudio Dema (CNR-IMAA), Mauro Di Fenza (CNR-IBBR), Cristina Di Muri (CNR-IRET), Elisabetta Erba (UniMI), Claudia Fratianni (INGV), Marcello Magaldi (CNR-ISMAR), Carlo Mantovani (CNR-ISMAR), Mauro Mazzola (CNR-ISP), Lucia Mona (CNR-IMAA), Alessandro Oggioni (CNR-IREA), Elena Partescano (OGS), Davide Raho (CNR-IRET), Pietro Roversi (CNR-IBBA), Simone Sabbatini (CMCC), Rosalia Santolieri (CNR-ISMAR), Angela Saraò (OGS), Katrin Schroeder (CNR-ISMAR), Giulio Verazzo (CNR-ISP), Daniele Zannoni (UNIVE), David Zuliani (OGS).
Comments	Due to the nature of the deliverable, this document shows two release dates: an initial planned release and a final release updated to the latest available information. This document may undergo several updates and releases as the evolving technologies.

INDEX

1 List of acronyms

ACTRIS: Aerosols, Clouds, and Trace Gases Research Infrastructure
AnaEE: Analysis and Experimentation on Ecosystems
API: Application Programming Interface
ASC: Atmospheric Simulation Chambers
ASCII: American Standard Code for Information Interchange
ATLaS:Advanced Technologies for LandSlides
BODC: British Oceanographic Data Centre
CAS: Central Authentication Service
CC0: Creative Commons Zero (No Rights Reserved)
CC0 1.0: Creative Commons Zero 1.0 Universal (No Rights Reserved)
CC BY 4.0: Creative Commons Attribution 4.0 International License
CC BY-SA 4.0: Creative Commons Attribution-ShareAlike 4.0 International License
CC BY-NC-SA 4.0 International: Creative Commons Attribution Non Commercial-ShareAlike 4.0 International License
CeTrA: Centre for Trace Analysis
CMEMS: Copernicus Marine Environment Monitoring Service
CNR: Consiglio Nazionale delle Ricerche
CONiSMA: Consorzio nazionale interuniversitario per le scienze del mare
CF: Climate Forecast
CSV: Comma Separated Values
CSW: Catalog Service for the Web
DCAT-AP: Data Catalog vocabulary Application Profile
DiSSCo: Distributed System of Scientific Collections
DO: Digital Object
DOI: Digital Object Identifiers
DMP: Data Management Plan
DSG: Discrete Sampling Geometries
DVAS: ACTRIS Data Discovery, Virtual Access and Services unit
ECORD: European Consortium for Ocean Research Drilling
EDMO: European Database of Marine Organization
eLTER: Integrated European Long-Term Ecosystem, critical zone and socio-ecological Research
EML: Ecological Metadata Language
EMPHASIS: European Infrastructure for Multi-scale Plant Phenomics and Simulation
EMSO: European Multidisciplinary Seafloor and water column Observatory
EMODnet: European Marine Observation and Data Network
ENEA: Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile
ENVRI: Environmental Research Infrastructures
EOSC: European Open Science Cloud
ERIC: European Research Infrastructure Consortium
ESFRI: European Strategy Forum on Research Infrastructures
EUDAT: European Data Infrastructure
EUFAR: European Facility for Airborne Research
EVIOR: European Virtual Infrastructure in Ocean Research
FAIR: Findable, Accessible, Interoperable, and Reusable

FDO: FAIR Digital Objects
FDSN: International Federation of Digital Seismograph Networks
FIPs: Fair Implementation Profiles
GCMD: Global Change Master Directory
GDAC: Global Data Assembly Centres
GTS: Global Telecommunication System
HCMR: Hellenic Centre for Marine Research
HGP: Human Genome Project
HTTPS: HyperText Transfer Protocol Secure
IADC: Italian Artic Data Center
IBISBA: Industrial Biotechnology Innovation and Synthetic Biology Accelerator
ICOS: Integrated Carbon Observation System
IIM: Istituto Idrografico della Marina
IMAA: Istituto di Metodologie per l'Analisi Ambientale
INGV: Istituto Nazionale di Geofisica e Vulcanologia
INFN: Istituto Nazionale di Fisica Nucleare
IODP: International Ocean Discovery Program
IRET: Istituto di Ricerca per gli Ecosistemi Terrestri
ISPRA: Istituto superiore per la protezione e la ricerca ambientale
ITINERIS: Italian Integrated Environmental Research Infrastructures System
LNS: Laboratori Nazionali del Sud
JERICO: Joint European Research Infrastructure of Coastal Observatories
JRU: Joint Research Unit
JSON: JavaScript Object Notation
KML: Keyhole Markup Language
LIDAR: Light Detection And Ranging
MMD: MET Norway Metadata Format
MOD: Metadata for Ontology Description
N/R Laura Bassi: Nave da ricerca Laura Bassi
NCEI: National Centre for Environmental Information
NetCDF: Network Common Data Form
NODCs: National Oceanographic Data Centers
NRT: Non-Real Time
NVS: NERC Vocabulary Server
OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting
OASIS: Archive System of Instrumental Seismology
ODC-By: Open Data Commons Attribution Licence
ODC-ODbL: Open Data Commons Open Database License
ODV: Ocean Data View
OGC:Open Geospatial Consortium
OGC-SOS: Open Geospatial Consortium - Sensor Observation Service
OGC-WMS: Open Geospatial Consortium - Web Map Service
OGS: Istituto Nazionale di Oceanografia e di Geofisica Sperimentale
ORFEUS: Observatories & Research Facilities for European Seismology
OU: Operative Unit
PDDL: Public Domain Dedication and License
PID: Persistent Identifiers
PURL: Persistent Uniform Resource Locator

QuakeML: A flexible, extensible, and modular XML representation of seismological data
RDF: Resource Description Framework
RDBMS: Relational Database Management System
REST: Representational State Transfer
RI: Research Infrastructure
RINEX: Receiver Independent Exchange Format
ROOS: Regional operational oceanographic Systems
RRT: Real-Time
SAC: Seismic Analysis Code
SIOS: Svalbard Integrated Arctic Earth Observing System
SMINO: Sistema di monitoraggio terrestre dell'Italia Nord Orientale
SOAP: Simple Object Access Protocol
SOPs: Standard Operating Procedures
SSO: Single Sign-On
SZN: Stazione Zoologica Anton Dohrn
URI: Uniform Resource Identifier
URL: Uniform Resource Locator
URN: Universal Resource Names
UUID:Universally Unique Identifier
VRE: Virtual Research Environment
WFS: Web Feature Service
WMO: World Meteorological Organisation
WMS: Web Map Service
WP: Work Package
XML: Extensible Markup Language

2 INTRODUCTION

In the last decades the scientific community has been experiencing an unprecedented influx of digital research outputs. The huge amount of gathered research outputs has evidenced that efficient management is necessary for reproducing and advancing research, fostering collaboration, and ensuring research outputs' long-term value. In response to these challenges, the FAIR principles, an acronym for Findable, Accessible, Interoperable, and Reusable, have emerged as a revolutionary framework designed to guide and optimise the integration and reusing of data and other digital assets (Wilkinson et al., 2016; Devaraju et al., 2021). Since their introduction, the FAIR principles have revolutionised, together with the open science policy (Hill, 2019), research management, by promoting better practices and encouraging the distribution of digital objects (DOs, see section 4 for further details on this concept) from various sources, including science, public administration, and the private sector to society (Peters-Von Gehlen et al., 2022). However, the FAIR principles are aspirational, providing a set of guidelines, attributes, and behaviours to attain FAIRness rather than explicitly outlining how to reach this ideal state (Wilkinson et al., 2019). It should be noted that while the article by Wilkinson et al. (2016) provides crucial guidance on the "where" we need to go, the subsequent critical step is to determine the "how." This involves specifying the technical steps required in this direction. In this context, several significant international efforts focused on FAIR principles have emerged, such as ENVRI-FAIR[1], GO FAIR[2], FAIRsFAIR[3], FAIR IMPACT[4], ENVRI-Hub NEXT[5] and, at the national level, the ITINERIS project. These initiatives aim to identify the most effective strategies for progressing in the direction outlined by the vision of a European Open Science Cloud[6] (EOSC).

As reported by van Reisen et al. (2020), the implementation of FAIR principles to date has reached a more advanced level and has been more effectively applied in the biomedical field, primarily in the areas of health and biotechnology. This is mainly due to intrinsic necessity, imposed by the subject matter itself: for the study and sequencing of DNA, for example, see the Human Genome Project - HGP (Gibbs, 2020), it was essential to be able to share and interact with a significant volume of data. Furthermore, in the context of public health, certain policy choices were made even before the advent of the COVID-19 pandemic, which promoted the international free sharing of virus sequences as advantageous (Maxwell et al., 2023; Babady et al., 2022).

Starting from the experience gained in these areas, the field of environmental research could take inspiration and likewise implement those successful strategies that have made data and many related services (software, algorithms, etc.) more easily findable, accessible, interoperable, and reusable. In this context, research infrastructures (RIs) play a crucial role as they serve as the fundamental framework providing resources and support for data collection, storage, analysis, and dissemination, thereby enabling scientific progress. However, as highlighted by Kirkpatrick et al. (2023), for researchers and those involved in data management, the FAIR aspect of data and related services can sometimes be perceived merely as another item on a long list of unfunded obligations or as a resource-intensive

[1] https://envri.eu/
[2] https://www.go-fair.org/
[3] https://www.fairsfair.eu/
[4] https://fair-impact.eu/
[5] https://www.egi.eu/project/envri-hub-next/
[6] https://eosc-portal.eu/

requirement. This issue, as indicated by the authors of the same study, is also related to the excessive generality of most FAIR implementation guidelines. For this reason, it is of great importance to be able to identify and describe in detail and in practical terms the FAIR strategies implemented by the more mature RIs. In this way, less advanced RIs can easily draw inspiration from the FAIR solutions already identified in the same domains.

3 Purpose of this document

This deliverable represents the "*State of the Art review of FAIR-enabling best practices"* which is part of activity 2.3 of the Work Package (WP) 2. The overall objective of this document is to provide an up-to-date analysis of the FAIR implementation practices adopted by more consolidated national RIs in each environmental subdomain (i.e. atmosphere, marine, terrestrial biosphere and geosphere landsurface). Based on this analysis, the document will offer a general overview of the strategies and methodologies adopted by RIs to put into practice the FAIR principles. Our analysis is aimed to facilitate the transfer of best practices from the most advanced RIs to the emerging and evolving ones, thus contributing to strengthening the national RIs landscape in the environmental domain.

3.1 The Italian Research Infrastructures cluster of the ITINERIS project

The ITINERIS project coordinates a network of national nodes from 22 RIs which can be classified according to four different categories (Fig. 1):

I. ESFRI Landmarks (Italian node): ACTRIS, EMSO, Euro-Argo, ICOS, LifeWatch, and ANAEE;
II. ESFRI Projects (Italian node): DANUBIUS, DiSSCo, eLTER, EMPHASIS, and IBISBA;
III. EU RIs: ECORD, EUFAR, Eurofleets, JERICO, and SIOS;
IV. National RIs: ATLAS, CeTRA, Laura Bassi, SMINO, Geosciences, and LNS.

According to the ESFRI Roadmap 2021[7], the majority of the ITINERIS RIs (17 on 22) fall within the "Environment" domain, three belong to "Health and Food" domain and have a strong link with the environment (e.i., ANAEE, EMPHASIS, and IBISBA), while the remaining two which support the services for the marine domain (e.i., Geosciences and LNS) may be ascribed to the "Physical Science and Engineering" domain. The ITINERIS project is structured into eight Work Packages (WPs). Of these, four WPs correspond to four environmental subdomains: atmosphere (WP4), marine (WP5), terrestrial biosphere (WP6), and geosphere landsurface (WP7) (Fig. 1).
Clearly, considering the closely intertwined nature of environmental subdomains, some of the 22 RIs are simultaneously represented in two or more subdomains (e.g. ICOS is present in the atmosphere, terrestrial biosphere, and marine subdomains) (Fig. 1). In this deliverable, we have chosen to present each RI within a specific domain. This decision aims to improve the clarity of the analysis and enhance the readability of the results, making it easier for readers to comprehend.

[7] https://roadmap2021.esfri.eu/media/1295/esfri-roadmap-2021.pdf

Consequently, in this document, the RIs have been grouped as follows:

- **Atmosphere**: ACTRIS ERIC, ICOS ERIC, SIOS, EUFAR, CeTrA;
- **Marine**: DANUBIUS, EMSO ERIC, Euro-Argo ERIC, EuroFleets, Geoscience, JERICO, LNS, N/R Laura Bassi;
- **Terrestrial biosphere**: AnaEE ERIC, DiSSCo, eLTER, EMPHASIS, IBISBA, LifeWatch ERIC;
- **Geosphere landsurface**: ATLaS, ECORD, SMINO.

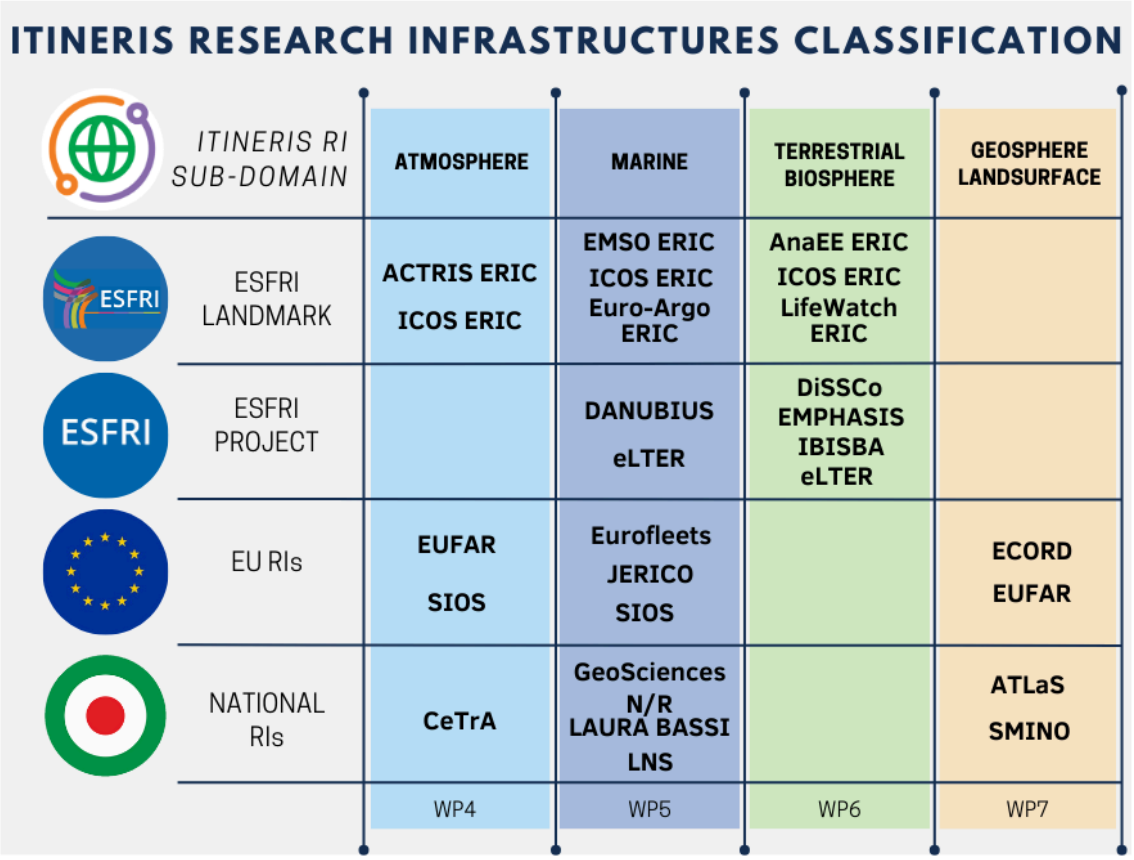

Figure 1. The 22 Research Infrastructures (RIs) involved in the ITINERIS project collocated within the four environmental subdomains (atmosphere, marine, terrestrial biosphere and geosphere landsurface). The RIs are reported and classified according to the ESFRI and as EU or National RIs. Cross-domain RIs were simultaneously reported in each subdomain of interest (e.g. ICOS is reported in atmosphere, marine and terrestrial biosphere).

4 Analysis Approach

We conducted an analysis of all 22 RIs involved in the project using a three-step process: a) online survey[8], b) one-to-one online interviews and c) in-depth analysis of the acquired documentation. The survey has been structured to provide an initial overview of the digital resources that each RI provides and/or manages, along with the accompanying documentation. It consisted of 10 questions aimed to gather preliminary information about the state of the art of FAIR-enabling best practices. The survey was sent to the FAIR Implementation Working Group, consisting of representatives from each RI. After a preliminary analysis, consultations were scheduled between the specific personnel hired for activity 2.3 (one dedicated person for each subdomain) and the representatives of each RI. The one-to-one meetings provided the appropriate context to thoroughly examine all the answers provided during the survey. Finally, an in-depth analysis was conducted on the website of each RI and on the reference documentation in order to extract all the information useful for describing the state of the art of FAIRness within the RIs. Upon completion of this process, all extracted information underwent thorough review by the FAIR Implementation Working Group representatives.
In order to investigate the application of FAIR principles in the 22 RIs, a series of FAIR-enabling practices actually employed by the RIs were examined. All categories of the identified practices are presented in Figure 2, wherein the direct correlation between FAIR principles and practices can be discerned.

After completing the information gathering and analysis, it was decided to showcase the results only for the RIs with a minimum FAIR readiness level. This decision stems from the fact that some RIs are at an early stage of their construction and therefore did not have enough information to be analysed. It also aligns with the purpose of this document to extract best practices, which requires a focus on the RIs with an advanced level of development. For these reasons, 13 out of the 22 participating RIs have been analysed and named as follows:

- **Atmosphere**: ACTRIS, ICOS, SIOS;
- **Marine**: EMSO, Euro-Argo, EuroFleets, JERICO;
- **Terrestrial biosphere**: AnaEE, eLTER, IBISBA, LifeWatch;
- **Geosphere landsurface**: ECORD, SMINO.

Hereafter, the results are organised into four sections, which refer to the four foundational principles: Findability (5.1); Accessibility (5.2); Interoperability (5.3); and Reusability (5.4). Each section initially provides a brief description of the principle and it is then divided into subsections that describe, for each subdomain, the choices and practices adopted by the RIs for its implementation. In Figure 2, it is clearly depicted how each "subsection" (last column) is linked to the 15 sub-principles in which the FAIR principles are elaborated.

[8] Survey_State of the Art of FAIR-enabling best practices".xlsx

FAIR PRINCIPLES AND FAIR-ENABLING PRACTICES

FAIR principles		FAIR-enabling practices
FINDABILITY	**F1.** (Meta)data are assigned a globally unique and persistent identifier	PERSISTENT IDENTIFIER (PID)
	F2. Data are described with rich metadata (defined by R1 below)	METADATA SCHEMA [A]
	F3. Metadata clearly and explicitly include the identifier of the data they describe	
	F4. (Meta)data are registered or indexed in a searchable resource	SEARCHABLE RESOURCES
ACCESSIBILITY	**A1.** (Meta)data are retrievable by their identifier using a standardised communication protocol	COMMUNICATION PROTOCOL
	A1.1 The protocol is open, free, and universally implementable	
	A1.2 The protocol allows for an authentication and authorisation procedure, where necessary	AUTHENTICATION & AUTHORISATION SERVICE
	A2. Metadata are accessible, even when the data are no longer available	METADATA LONGEVITY PLAN
INTEROPERABILITY	**I1.** (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	(META)DATA FORMAT [B]
	I2. (Meta)data use vocabularies that follow FAIR principles	SEMANTIC ARTEFACTS [C]
	I3. (Meta)data include qualified references to other (meta)data	N/A
REUSABILITY	**R1.** (Meta)data are richly described with a plurality of accurate and relevant attributes	(META)DATA SCHEMA [A]
	R1.1. (Meta)data are released with a clear and accessible data usage licence	(META)DATA LICENCE
	R1.2. (Meta)data are associated with detailed provenance	(META)DATA SCHEMA [A]
	R1.3. (Meta)data meet domain-relevant community standards	SEMANTIC ARTEFACTS [C] + (META)DATA FORMAT [B]/SCHEMA [A]

Figure 2 – The figure illustrates the connection between individual FAIR principles, as reported in the GO FAIR initiative[9] (F1; F2; F3; F4; A1; A1.1; A1.2; A2; I1; I2; I3; R1; R1.1; R1.2; R1.3) and the practices for their implementation within the RIs, as identified through the survey completed by the Fair Implementation Working Group. In the "FAIR-enabling practices" column, the superscript letters indicate that entries marked with the same letter share the same definition, as described in the respective subsequent sections. Namely: A) Metadata Schema, section 5.1.2; B) Metadata format, section 5.3.1; and C) Semantic artefacts, section 5.3.2.

[9] https://www.go-fair.org/

As already specified in the ITINERIS project's Data Management Plan (Deliverable 2.10), also this document will refer to DOs since the RIs do not solely manage research data. Providing an exact and precise definition of "DO" can be challenging due to the lack of consensus among the different research groups. The term "digital object" is used in various initiatives that address the implications of sharing and reusing information in contexts featuring diverse standards and sources. Nevertheless, these definitions often emerge from separate communities with unique prerequisites, frameworks, and terminologies. Consequently, the definitions of DOs tend to be incompatible, exacerbating the interoperability challenges these initiatives aim to resolve.
In this deliverable, the definition proposed by Kahn and Wilensky (1995; 2006) was adopted, whereby a DO is defined as "a sequence of bits, or a set of sequences of bits, incorporating a work or portion of a work or other information in which a party has rights or interests, or in which there is value, each of the sequences being structured in a way that is interpretable by one or more of the computational facilities, and having as an essential element an associated unique persistent identifier". The diagram in Figure 3 indicates the simple structure of this model. The content of a DO is encoded as a structured bit-sequence and stored in repositories. It is referenced by a globally unique, persistent and resolvable identifier (PID), and is described by metadata (descriptive, scientific, system, provenance, rights, etc.) (Fig. 3). Metadata descriptions themselves are DOs. Moreover, DOs can be aggregated into collections which are also DOs with a content consisting of the references to its components. This definition makes DO a generic concept, abstracting away from the many possible types of content of a DO, and covering the whole domain of digital data entities. Specifically, in this deliverable, the term "DO" refers to research data, semantic artefacts, research codes, software, policy documents, training materials, and metadata records that a RI can potentially generate and manage along the entire lifecycle of the ITINERIS project.

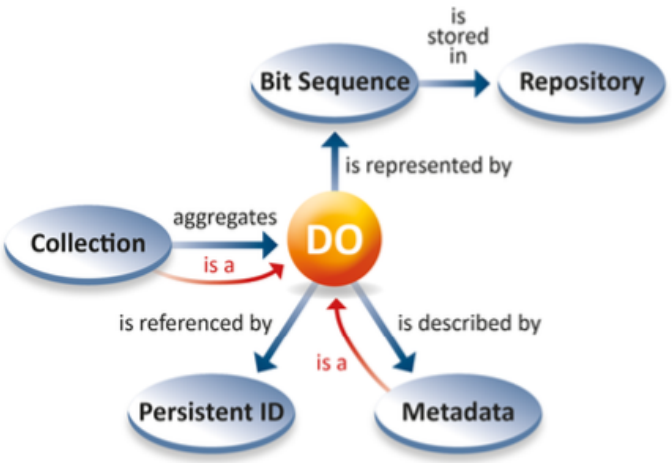

Figure 3. Relations between information associated with a digital object (DO; Wittenburg et al., 2019).

Due to the heterogeneity in terms of governance, structure and aims of the ITINERIS RIs, we give an overview of the RIs analysed for the scope of this report (see "*Analysis approach*" section for more details) and their Italian contribution. The overview aims to facilitate the readability of the following sections.

ACTRIS ERIC

The Aerosols, Clouds, and Trace Gases Research Infrastructure (ACTRIS) is a pan-European research infrastructure, producing high-quality data and information on short-lived atmospheric constituents and on the processes leading to the variability of these constituents in natural and controlled atmospheres. Its main aim is to provide facilities, high-quality data and related services to a wide range of users, both public and private. The ACTRIS community boasts world-class expertise in various atmospheric science fields and offers advanced technology for atmospheric observation.

ACTRIS operations are performed by ACTRIS Central Facilities (CFs) that include the Data Centre[10], the Head Office and six Topical Centres: the Centre for Aerosol In Situ - European Centre for Aerosol Calibration (CAIS-ECAC)[11]; the Centre for Aerosol Remote Sensing (CARS)[12]; the Centre for Cloud In Situ (CIS)[13]; the Centre for Cloud Remote Sensing (CCRES)[14]; the Centre for Reactive Trace Gases In Situ (CiGas)[15]; the Centre for Reactive Trace Gases Remote Sensing (CREGARS)[16]. The ACTRIS CFs represent the key operative entities of this RI and have a fundamental role as they provide services to the users according to the ACTRIS access policy as well as operation support to the National Facilities to increase their performance as well as participate in ACTRIS governance and management. The CFs link the National Facilities, i.e. the observational and exploratory platforms, which are operated at the national level and produce the majority of the ACTRIS data. ACTRIS Data Centre is a distributed data centre, which offers scientists and other user groups with free and open access to all ACTRIS data, complemented with access to innovative and mature data products, together with tools for quality assurance (QA), data analysis and research. It is organised in six units, i.e., five units (In-Situ, ARES, CLU, GRES, ASC) with complementary topic expertise and one unit with integrating activities (DVAS). Regarding data acquisition, ACTRIS Data Management Plan (DMP) reports that data comprises ACTRIS variables resulting from measurements at National Facilities that fully comply with the standard operating procedures (SOP), measurement recommendations, and quality guidelines established within ACTRIS. Regarding data production and provision, there are 4 levels of ACTRIS data: a) level 0 (Raw sensor output, metadata necessary for next level); b) level 1 (calibrated and quality assured data with minimum level of quality control); c) level 2 (approved and fully quality controlled ACTRIS data product or geophysical variable); d) level 3 (elaborated data products derived by post-processing of ACTRIS Level 0, 1, 2 data, and data from other sources).

ACTRIS-IT[17] is the Italian Joint Research Unit that coordinates the national component of the distributed research infrastructure ACTRIS, coordinated by the National Research Council. It promotes and facilitates the Italian participation in the implementation process of the ACTRIS research infrastructure by fostering the national scientific research in the field of aerosols, clouds and trace gases, and their crucial role in understanding climate processes. ACTRIS-IT is based on a number of different facilities distributed throughout the national territory including observational sites with multi-parameter instrumentation, laboratories and experimental simulation chambers. For the scope of the current document,

[10] https://www.actris.eu/topical-centre/data-centre
[11] https://www.actris-ecac.eu/
[12] https://www.actris.eu/topical-centre/cars
[13] https://www.actris.eu/topical-centre/cis
[14] https://www.actris.eu/topical-centre/ccres
[15] https://www.actris.eu/topical-centre/cigas
[16] https://www.actris.eu/topical-centre/cregars
[17] http://www.actris.it/

it is necessary to highlight that Italy strongly contributes to the implementation of the ACTRIS Lidar Calibration Centre and the ACTRIS Data Centre. CNR-IMAA hosts one of the 8 units of the Centre for Aerosol Remote Sensing[18] (CARS-AHL-CNR), which is responsible for the QA/QC of the aerosol high-power lidar measurements; and the Aerosol Remote Sensing Data Centre[19] unit (DC-ARES), which is the main responsible for the Single Calculus Chain for aerosol remote sensing (SCC), the centralised processing suite for the processing of aerosol lidar data.

ICOS ERIC

The Integrated Carbon Observation System (ICOS) is a European research infrastructure with the aim of providing accessible and high quality data to improve understanding of the carbon cycle in the European Continent and other greenhouse gas emissions and absorption in a standardised manner. High-standard measurements are keys for supporting policy-making in the fight against climate change and its impacts in society. ICOS comprises more than 170 stations spread across 16 countries over three domains: land, ocean and atmosphere. Hundreds of scientists are engaged in the activities of ICOS. Stations are provided with different instrumentation and processing pipelines in the three domains to serve the overall goal of providing greenhouse gas observations: for example, GHG concentrations in the atmosphere, carbon and energy fluxes between the atmosphere, the land surface and the oceans. ICOS also includes: i. three Thematic Centres on terrestrial ecosystems (Ecosystem Thematic Centre[20]), oceans and sea (Ocean Thematic Centre[21]) and atmosphere (Atmosphere Thematic Centre[22]); ii. two analytical laboratories (Central Analytical Laboratories[23]); iii. the Carbon Portal[24], a 'one-stop shop' for all ICOS data products. As observed in other Research Infrastructures (RI) as well, ICOS also categorises its research data and products into multiple levels based on the type of processing. Specifically, the following levels are present: Level 0 (Raw data); Level 1 (Intermediate observational data); Level 2 (Final quality-controlled observational data); Level 3 (Elaborated products).

Italian participation in ICOS is supported and promoted by a Joint Research Unit (JRU) made up of 15 institutions including research centres and universities, coordinated by the National Research Council. The ICOS Italy Network consists of 17 stations, 10 dedicated to terrestrial ecosystems, 4 focused on the ocean, and 3 dedicated to atmospheric studies. Furthermore, ICOS Italy coordinates, through the Euro-Mediterranean Centre on Climate Change (CMCC) and the University of Tuscia, the Ecosystem Thematic Centre (ETC). The ETC, whose activities are carried out in three different countries (Italy, Belgium and France), is coordinated from the Italian headquarters[25]. All the data from the ecosystem stations of the European network converges to the ICOS Carbon Portal[26] located in Lund (SE), while the metadata and related services are managed in the ETC servers. The ETC is in charge of the data processing and quality control, methodological developments, training and coordination of the network, among many other services. .

[18] https://www.actris.eu/topical-centre/cars
[19] https://www.actris.eu/topical-centre/data-centre/ares-aerosol-remote-sensing-data-centre-unit
[20] https://www.icos-cp.eu/observations/ecosystem/etc
[21] https://www.icos-cp.eu/observations/ocean/otc
[22] https://www.icos-cp.eu/observations/atmosphere/atc
[23] https://www.icos-cp.eu/observations/atmosphere/cal
[24] https://www.icos-cp.eu/observations/carbon-portal
[25] http://gaia.agraria.unitus.it/icos/home
[26] https://data.icos-cp.eu/portal/

SIOS

The Svalbard Integrated Arctic Earth Observing System (SIOS) is a collaborative effort to develop and maintain a regional observational system for long-term measurements in and around Svalbard, addressing Earth System Science questions related to Global Change. SIOS is a distributed research infrastructure organised as a consortium of 29 member institutions from 10 countries. It coordinates, develops and optimises research infrastructure owned by the member institutions. The observing system and research facilities offered by SIOS build on the extensive observation capacity and diverse world-class research infrastructure provided by many institutions already established in Svalbard.

The SIOS Knowledge Centre (SIOS-KC) is the central hub of SIOS. It is located in the Svalbard Science Centre in Longyearbyen and offers coordinated services for the international research community. The SIOS-KC is currently funded by the Research Council of Norway through 2026. The activities are based on three pillars: project management and communication, joint activities, and services. These latter include:

- *Access programme*[27] and *logistical services*[28] - coordinated access to the various scientific facilities and equipment owned by SIOS members. SIOS also offers and coordinates logistical support;
- *Remote sensing services*[29] - single-point of contact for ground-, airborne-, and satellite-borne information for Svalbard;
- *Data management services*[30] - a unified interface to data that are produced using SIOS related infrastructure and third-party datasets that are of relevance to the SIOS scientific community.

The SIOS Data Management System relies on the principles of distributed data management. Datasets that are relevant for SIOS, as well as their associated metadata, are managed by several physically distributed data centres and the information from these is collected in the SIOS Data Access Portal[31]. Each Data Centre has its own set of Data Management Facilities for ingestion of new data (and associated metadata), maintenance of the data sets (including metadata) and for data curation. The SIOS Data Management Service enables dataset discovery through standardised metadata, and retrieval, visualisation & transformation of data. The Italian contribution to SIOS is the Italian Artic Data Center[32] (IADC) managed by the National Research Council. The IADC is a repository for scientific data acquired by Italian research activity in the Arctic region. It is based on the ERDDAP data server that provides a simple way to download subsets of scientific datasets in common file formats, make graphs and maps. Metadata are also available by a catalogue[33] based on GeoNetwork 3.12.3.0.

EMSO ERIC

The European Multidisciplinary Seafloor and water-column Observatory (EMSO) is a pan-European research infrastructure focused on exploring the oceans to understand their

[27] https://sios-svalbard.org/RIAccess
[28] https://sios-svalbard.org/Logistics
[29] https://sios-svalbard.org/RemoteSensing
[30] https://sios-svalbard.org/Data
[31] https://sios-svalbard.org/metsis/search
[32] https://data.iadc.cnr.it/erddap/index.html
[33] https://metadata.iadc.cnr.it/geonetwork/srv/eng/catalog.search#/home

integral role in the Earth system. EMSO contributes to studies across diverse disciplines, from biology to computer science, in environments ranging from polar to tropical. EMSO comprises 14 Regional Facilities, including 11 deep-sea observatories and 3 shallow-water test sites, spread across Europe from the North East Atlantic to the Black Sea. These observatories, equipped with various sensors, continually record biogeochemical and physical parameters useful to monitor natural hazards, climate change, and marine ecosystems. EMSO offers data and services to a large and diverse group of users, from scientists and industries to institutions and policy makers, for addressing environmental policies and significant social issues.
At present, EMSO ERIC is a consortium of 8 European countries sharing a common strategic framework of scientific facilities (data, instruments, computing and storage capacity). Italy hosts the ERIC office at the INGV in Rome. INGV also coordinates the Italian Joint Research Unit (JRU) composed of INGV, SZN, INFN, CNR, OGS, ISPRA, ENEA, CONiSMA and IIM. The activities of the involved Italian research institutes are carried out in the regional facilities of the South Adriatic Sea, Western Ionian Sea, and Western Mediterranean Sea. Data gathered by regional facilities are harvested centrally and served from one single ERDDAP server; each partner serves data through their own ERDDAP server and references datasets as a distributed federation system. The central EMSO ERIC ERDDAP server then provides those datasets as if they are locally hosted. In addition to a web interface, ERDDAP also provides a RESTful API that allows users to programmatically interact with the data using scripting languages such as Python, R or Matlab, and can be used to provide data to dashboards and other applications. Finally, EMSO also provides research code and software to open, harmonise, plot and analyse marine datasets. The creation of thematic VRE is under development.

Euro-Argo ERIC
Euro-Argo ERIC is Europe's contribution to the International Argo Programme, overseeing European in situ ocean observations. It operates about 800 floats, which is ¼ of the global Argo array, collecting temperature and salinity data every 10 days from deep oceans. This data supports climate change research and operational services. Euro-Argo also aims to expand the Argo network to deeper oceans and new biogeochemical parameters. Integrated with the Copernicus Marine Environment Monitoring Service (CMEMS) and European Marine Observation and Data Network (EMODnet), it became an ERIC in 2014, focusing on enhancing its real-time data systems for oceanography and meteorology. It also offers research codes[34], software[35], best practice documents[36], and specialised services, such as joint float procurement, monitoring of the European fleet and training programs. Virtual labs and thematic VRE are also developed under the framework of the EOSC-Blue Cloud. The Euro-Argo ERIC is composed of a Central Infrastructure (the ERIC Office) and distributed national facilities. Argo Italy is the Italian component of Euro-Argo. It focuses mainly on the Italian seas, such as the Mediterranean, and the Black Sea being an important component related to it. OGS coordinates and is responsible for the activity of the Mediterranean and Black Sea Argo Regional Center (MedArgo). The activity of Argo-Italy also extends over the Southern Ocean and the Ross Sea. Data that reflects the state of the ocean in real time is distributed by a powerful data-management system supported by robust quality-control processes. The international Argo Data System is

[34] https://github.com/euroargodev; http://www.argodatamgt.org/Documentation/Tools
[35] https://argo.ucsd.edu/data/argo-software-tools/
[36] http://www.argodatamgt.org/Documentation

based on two Global Data Assembly Centres (GDACs), eleven national Data Assembling Centres (DACs) and six Argo Regional Centres (ARCs), located across the world and managed by the Argo Data Management Team[37]. Two national DACs are based in Europe: The French DAC (operated by Coriolis to process data of float deployed by Bulgaria, Finland, France, Germany, Greece, Italy, Netherlands, Norway, Poland, Portugal, Spain and EU floats) and the UK DAC (operated by the British Oceanographic Data Centre (BODC) to process all UK and Irish data as well as from Mauritius and Saudi Arabia. After receiving data from the satellite operators, the DACs carry out decoding and quality-control processes according to a set of real-time automatic tests agreed by the international Argo programme (Argo Quality Control Manual). The data is then sent to the two GDACs and to the World Meteorological Organisation (WMO)'s Global Telecommunication System (GTS). These data flows enable near real-time exchange of information. The GDACs synchronise their database every day and process data for operational users or research communities. The data is also archived in the Global Argo Data repository operated by NCEI (National Centre for Environmental Information) in the US. For the delayed-mode streams, the Euro-Argo ERIC contributes with four delayed-mode operators working from four different national institutions - BSH in Germany, Coriolis in France, OGS in Italy and BODC in the UK - which process all the European floats. Three ARCs are coordinated by the Euro-Argo ERIC. The ARCs offer regional analysis of regional Argo data to assess its internal consistency with recent shipboard data. Euro-Argo coordinates 3 of these centres: the North Atlantic ARC (NA-ARC, managed by Ifremer), the Mediterranean and Black Seas ARC (MedArgo, managed by OGS) and the Southern Ocean ARC (SOARC, managed by an international collaboration of partner institutions, namely BSH, CSIRO, SOCCOM/MBARI and BODC). Details on Euro-Argo data management can be found in the dedicated DMP[38].

Eurofleets

The Eurofleets is a marine Research Infrastructure that promotes open access to an integrated and advanced research vessel fleet, designed to meet the evolving and challenging needs of the user community. It focuses on granting seamless access to research vessels for European and global researchers, emphasising support for early-stage, female scientists, and those from under-resourced areas. Eurofleets is committed to developing tools for deep ocean exploration and data management, fostering industry collaboration to drive innovation. The project champions open data access and active stakeholder engagement to guide the European research fleet's future. Furthermore, it prioritises education, training, and outreach, emphasising ocean literacy and gender diversity in science. The CNR represent the italian lead partner in the framework of the "Piano Nazionale Infrastrutture di Ricerca (PNIR) 2021 – 2027" [39].

Regarding the DOs management, specific data management plans produced during an oceanographic cruise are deployed in synergy with SeaDataNet and the European network of National Oceanographic Data Centers (NODCs)[40]. Scientific cruise teams formulate cruise data management plans for review by SeaDataNet NODCs, who will coach each scientific team before, during, and after transnational access cruises. OGS represents the

[37] http://www.argodatamgt.org/
[38] https://zenodo.org/records/4561618
[39] https://www.mur.gov.it/sites/default/files/2021-10/Decreto%20Ministeriale%20n.1082%20del%2010-09-2021%20-%20PNIR%202021%20-%202027.pdf
[40] https://www.eurofleets.eu/data-management/

NODC for Italy that deals with the review process of the produced data management plans. All metadata and data are available at European level with SeaDataNet services for exchange and wider dissemination by other several European and international portals, such as CMEMS, EMODnet, Blue-Cloud, GEOSS, and IOC-IODE portals. Moreover, the Eurofleets website features the EVIOR platform (European Virtual Infrastructure in Ocean Research) which will give access to information and data from the Research Vessels (RV), cruise plans, cruises while sailing, Cruise Summary Reports, and finally the resulting data sets, included in the SeaDataNet CDI service. Moreover, the SeaDataNet SWE (Sensor Web Enablement) toolkit has been adopted and upgraded for transfer of the Eurofleets Automatic Reporting System (EARS) information from RVs to shore for storage in a Data Hub at the Spanish National Research Council (CSIC), and further distribution by means of SOS (Sensor Observation Service) for visualisation in a dashboard on the EVIOR platform. The data management work is performed by three reference NODCs on data files prepared by the principal investigator. These centres are the Hellenic Centre for Marine Research (HCMR) in Greece, the OGS in Italy and the Belgian Marine Data Centre (BMDC). They decide amongst themselves on what criteria the datasets will be assigned. The decision model is described in the Eurofleets DMP[41].

JERICO

JERICO is an integrated pan-European multidisciplinary and multi-platform research infrastructure dedicated to a holistic monitoring of coastal marine system changes. It is seamlessly bridging existing continental, atmospheric and open ocean RIs, thus filling a key gap in the ESFRI landscape. JERICO establishes the framework upon which coastal marine systems are observed, analysed, understood and forecasted. JERICO enables open-access to state-of-the-art and innovative facilities, resources, FAIR data and fit-for-purpose services, fostering international science collaboration. JERICO is spearheading the creation of an integrated central hub e-infrastructure, named JERICO-CORE, to facilitate the discovery, access, management, and interaction with a broad spectrum of coastal ocean resources. These encompass services, datasets, software, best practices, manuals, publications, organisations, projects, observatories, equipment, support, training, and similar entities. Moreover, this e-infrastructure will include a VRE resulting from the partnership with EOSC Blue-Cloud. The CNR and OGS represent the Italian institutes involved in the management of JERICO marine infrastructures located among the national waters.

As described by JERICO'S DMP[42], data management is organised at two different levels:

- For real time data (NRT/RT mode), the data providers are advised to contact the regional leaders of the seven geographical domains of the Copernicus Marine Environmental Monitoring Service (CMEMS) In Situ Thematic Centre (TAC) in order to push their data into the Regional Operational Oceanographic Systems (ROOS[43]) data centres. The relevant QC/QA procedures can be applied either by the provider or by the above regional data centres before being delivered to EMODnet and CMEMS portfolio when it is applicable.

41 https://www.eurofleets.eu/download/Deliverables/D1.3-Eurofleets-Data-Management-Plan-V1.1-RBINS.pdf

42 https://www.jerico-ri.eu/download/jerico-s3_deliverables/JERICO-S3-D6.1-Data-Management-Plan_final_2.pdf

43 https://noos.eurogoos.eu/rooss/

- For delayed mode data, it is recommended to deliver the data to the NODCs where they will be further processed and validated before being pushed to SeaDataNet. Populating the SeaDataNet CDI service will ensure that the delayed mode validated observation data time series (e.g. monthly /quarterly) will become available automatically for the relevant EMODnet thematic portals (chemistry, physics, biology) as input for their EMODnet data products and services, for inclusion in SeaDataNet data products such as climatologies, and in the overall SeaDataNet CDI data offer.

AnaEE ERIC

AnaEE (Analysis and Experimentation on Ecosystems) is an ERIC dedicated to the study of ecosystems and their response to environmental changes. It integrates, in a single, distributed, RI, all the steps of the scientific experimental methods, modelling and experimentation, fostering a deeper understanding of ecosystem functioning and resilience.
AnaEE ERIC is constituted by a Central Hub, in charge of the overall coordination of the ERIC, and of three service centres: i) the Technology Centre (TC), in charge of the research and development portion of the ERIC; ii) the Data and Modelling Centre (DMC), based in Italy, responsible for the processing and provision of data and metadata; iii) the Interface and Synthesis Centre, responsible for the overall integration of the results obtained thanks to AnaEE ERIC and their connection with the society at large.
Anaee ERIC also includes 8 national nodes: Belgium, Czech Republic, Denmark, Finland, Italy, France, Bulgaria, and one intergovernmental organisation, the International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM). These offer, through the AnaEE Platforms catalog[44], access to platforms with different facilities and technologies, including controlled environment chambers, mesocosms, and data management tools. The platforms are gathered in four categories: open-air platforms, enclosed platforms, analytical platforms, and modelling platforms.
The DMC is hosted by the Council for the research in agriculture and the analysis of agrarian economy (CREA) in Italy. Among other duties, it is in charge of feeding the data and model portals[45]. The data and models portals are composed by the API Portal[46] and the Data portal[47]. The API portal is based on the Azure API Management system. It allows users to register RESTful APIs of data products that may be of interest for the environmental research community. The Data portal is a CKAN instance (see section 5.1.3), and currently comprises 11 datasets of several file formats (see section 5.3.1).

eLTER

eLTER (Integrated European Long-Term Ecosystem, critical zone and socio-ecological Research) is an ESFRI RI, currently in its preparatory phase. eLTER is a pan-European in-situ research infrastructure whose mission is to study long-term ecological changes in terrestrial, freshwater and transitional/coastal ecosystems through a holistic "whole system" approach. eLTER is one of the regional groups of ILTER, the international umbrella network encompassing 39 International LTER networks, mostly on a national scale. eLTER consists of 26 separately governed and funded national LTER networks with

[44] https://isia.cnrs.fr/catalog/platform/list/anaee_europe
[45] https://www.anaee.eu/services/centres/data-and-modelling-centre
[46] https://developer.anaee.eu/
[47] https://data.anaee.eu/

varying historical backgrounds[48]. These networks and their sites comply with the eLTER criteria[49]. All eLTER member networks are also members of the ILTER network. eLTER's data collection spans terrestrial, freshwater, and marine ecosystems over 500 different sites and 50 LTSER (Long-Term Socio-ecological Research) platforms in Europe and beyond. This RI enables researchers to access and share long-term ecological information, mainly by providing access through the Dynamic Ecological Information Management System - Site and dataset registry (DEIMS-SDR)[50]. It allows discovery of all ILTER sites and it contains for each site information on: location, ecosystems, facilities, observed properties, related resources (e.g. datasets, data products, activities, sensors) and data management procedures (e.g. data policy). A growing number of datasets and data products can be directly accessed starting from the associated sites.
The eLTER Data Integration Portal (eLTER-DIP)[51] has been developed in order to provide access, through metadata, to Remote Sensing data as well as sensor time-series data collected by the national nodes. eLTER-DIP consists of a GeoNetwork-based metadata catalogue and of a visualisation tool. The DIP allows surfing among the DEIMS resources as well. Data is mainly described through B2Share, the data portal maintained by the European Data Infrastructure (EUDAT), while data is findable and accessible through eLTER Digital Asset Register (DAR)[52]. At national network level, other DOs are sporadically stored in institutional or generalist repositories, such as ZENODO[53]. The metadata of those DOs are generally visible in DEIMS, and related through DOIs with the associated research sites. In alternative, some resources are available by contacting the owner/producer. eLTER has an Italian node, LTER-Italy[54], coordinated by the CNR. Currently, DEIMS-SDR lists 78 active and recognised sites in Italy[55] (organised in parent sites), managed locally by universities and research centres. Beyond exposing their metadata on DEIMS, DIP and DAR, LTER Italy also offers a storage space for several types of DOs, including datasets, in a dedicated ZENODO community[56].

IBISBA

IBISBA (Integrated Biological Infrastructure for the Life Sciences) is an ESFRI RI currently in its preparatory phase, comprising at the moment two national nodes: Italy and France. Its main objective is to support industrial biotechnology by providing a single access point to researchers from academia and industry across the globe to integrate services for end-to-end bioprocess development in areas such as synthetic biology, metabolic engineering, and bioprocess optimization. By federating state-of-the-art research and development facilities, IBISBA promotes data standardisation and best practices as core elements of service reproducibility and interoperability. In doing so, IBISBA accelerates the production and translation of cutting-edge knowledge into innovation for biomanufacturing (https://bit.ly/496iBlk). Most of the facilities contributing to IBISBA are identified as parts of nationally based RIs that are earmarked for funding by national

[48] https://elter-ri.eu/national-lter-networks
[49]https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5da5b2e53&appId=PPGMS
[50] https://deims.org/
[51] https://dip.lter-europe.net/#/home
[52] https://catalogue.lter-europe.net
[53] https://zenodo.org/communities/elter?q=&l=list&p=1&s=10&sort=newest
[54] http://www.lteritalia.it/wordpress/?page_id=1034
[55] https://deims.org/networks/7fef6b73-e5cb-4cd2-b438-ed32eb1504b3
[56] https://zenodo.org/communities/lter-italy?q=&l=list&p=1&s=10&sort=newest

agencies. The owners of facilities are responsible for implementing strategy, staffing and research quality. The primary role of National nodes is to structure National research infrastructure in the field of biotechnology, building research service capability and interoperability, and a well-organised stakeholder community. Additionally, in the framework of IBISBA, National nodes will endeavour to develop research service offers to IBISBA, the aim being to compile a comprehensive catalogue of top-quality integrative services for international R&D&I clients. It is expected that the different National nodes will develop unique selling points, differentiating to some extent their service offer from those of other National Nodes in the IBISBA network. IBISBA-IT is the Italian node of the RI, and is coordinated by the CNR-IBBR-NA. It has the specific mission of "*developing new molecules and processes through enzyme/protein discovery and engineering, and development of new biotransformations and bioprocesses. These goals will be addressed through modern and holistic approaches including 'omics' sciences, bioinformatics, molecular engineering, and bioreactor technologies*"[57]. It is made up of ten partners: four CNR institutes from the Department of Biology, Agriculture and Food Sciences (DISBA): CNR-IBBR, CNR-ISPAAM, CNR-ISA and CNR-IBBA; five Universities: University of Naples, University of Bologna, University of Insubria, University of Milano Bicocca, University of Tuscia; and one foundation, the Istituto Insubrico di Ricerca per la Vita (FIIRV).
The IBISBA Knowledge Hub[58] is a key part of the IBISBA ecosystem of facilities and serves as a complete data management platform for the IBISBA projects. It is an online portal for recording the progression of the projects as well as storing, sharing and organising the data and models being generated from the IBISBA projects. It uses the ISA (Investigation, Study, Assay) structure to organise data providing different levels of accessibility and confidentiality. The portal allows the management of DOs as linked research objects. Each research object is an aggregation of DOs that are produced in a single research worK (Soiland-Reyes et al., 2022). IBISBA Knowledge Hub has been developed in the framework of IBISBA 1.0[59], the first H2020 funded project that launched the function of the RI through its first Transnational Access program. The Hub is based on the SEEK platform that is offered by the FAIRDOM association, which is a spin-off from the ELIXIR infrastructure.

LifeWatch ERIC
LifeWatch is a distributed ERIC dedicated to biodiversity and ecosystem research to explore new frontiers in ecological science and support society in addressing planetary challenges. It offers a comprehensive suite of e-science resources and services for scientists studying biodiversity, ecological processes, and the impacts of environmental changes. It comprises three common facilities, that are in charge of: i) the coordination of all legal, administrative and institutional activities, and the management of the distributed ICT e-Infrastructure operations and implementation (Statutory Seat and ICT e-Infrastructure Technical Offices in Spain); ii) the provision of tools and services to user communities, the active involvement of the European scientific community, and general communication activities (Service Centre in Italy); iii) the coordination and management of the requirements and needs analysis, the design and implementation of the scientific case studies and the products of the Virtual Laboratories (vLab & Innovations Centre in

[57] www.ibisba.it
[58] https://hub.ibisba.eu/
[59] https://www.ibisba.eu/EU-Projects/IBISBA-1.0/Deliverables

Netherlands). LifeWatch ERIC is composed of eight National nodes (Portugal, Spain, Belgium, the Netherlands, Italy, Slovenia, Bulgaria, and Greece).
LifeWatch ERIC is an e-infrastructure focusing on providing digital facilities for the curation, access, analyses and modelling of data. The gathering of data and related services is demanded by the National nodes that are free to set up their digital infrastructure and contribute in-kind to the functioning of the ERIC. LifeWatch ERIC effort is, then, towards integration of the data provided, and building Virtual Research Environments (VRE) and their components (e.g. web services, workflows) in order to provide high quality computing facilities for the analyses and modelling of these data. All DOs produced by LifeWatch ERIC, together with those produced by national nodes and other initiatives connected to them, are accessible from the LifeWatch metadata catalogue[60].
Italian participation in LifeWatch is supported and promoted by a Joint Research Unit (JRU) made up of 35 institutions among universities, research centres, national agencies and private entities, and it is led by the CNR. LifeWatch Italy activity focuses on the development and management of tools and services designed to cover the entire data lifecycle, from acquisition to visualisation and analysis. Data and their associated metadata are accessible by the Data Portal[61], based on Metacat, a repository software for preserving data and metadata in a standardised way that helps scientists to find, understand and effectively use data sets they manage or that have been created by others. The use of semantic artefacts for the annotation of meta(data) promotes their harmonisation and integration. The discoverability and accessibility of data and other DOs is also provided by the LifeWatch Italy Semantic Platform[62] based on a main semantic model that describes different resources and their metadata, as well as on meaningful links between controlled vocabularies and these resources. The semantic artefacts produced by LifeWatch Italy are deposited on EcoPortal[63], a repository for semantic resources maintained by LifeWatch ERIC. LifeWatch Italy also produces and maintains web services and VREs for the analysis of FAIR data, and other platforms for the training and networking of the scientific community and other stakeholders. Most of these digital facilities are hosted at the University of Salento and managed by the Research Institute on Terrestrial Ecosystem (CNR-IRET) in Lecce, Italy.

SMINO
SMINO (Northeast Italy monitoring system) is a nationally significant infrastructure included in the National Plan for Research Infrastructures (PNIR 2021-2027). It is directly managed by OGS through the Center for Seismological Research (CRS), one of the four scientific and technological research facilities in which OGS is organised. It aims to enhance the safety of citizens by increasing seismological knowledge in Northeastern Italy and improving the capacity to respond to earthquakes (Bragato et al., 2021). SMINO is composed of the integration of three networks: 1) the Seismometric network of Northeastern Italy[64], 2) the Accelerometric network of Northeastern Italy and 3) the Friuli Regional Deformation Network (FReDNet)[65]. These networks consist of several distributed stations equipped with high-quality sensors and GNSS

[60] https://metadatacatalogue.lifewatch.eu/srv/eng/catalog.search#/home
[61] https://dataportal.lifewatchitaly.eu/data
[62] https://semantics.lifewatchitaly.eu/resource/Start
[63] https://ecoportal.lifewatch.eu/
[64] https://rts.crs.inogs.it/
[65] https://frednet.crs.ogs.it/ItalianSite/XFReDNetHome.htm

(GPS+GLONASS+Galileo+Beidou) receivers for the seismic and geodetic monitoring of crustal deformations in Northeastern Italy. Networks data are acquired in real-time at the CRS (OGS) and their processing contributes to the rapid estimation of the impact of seismic events. Information on seismic events is disseminated to the public through a dedicated web portal[66]. Additionally, data are also available through EIDA Italia[67], the European Integrated Data Archive italian node, which stores and distributes seismic waveforms collected by italian and foreign data providers and managed from INGV.

ECORD

ECORD[68], the European Consortium for Ocean Research Drilling, is a component of the broader International Ocean Discovery Program (IODP). IODP focuses on a wide range of applied issues, such as climate and ocean change, biodiversity and origin of life, the Earth in motion including the study of earthquakes processes, and the Earth structure and dynamics in relation with its surface environment.

ECORD's role is to coordinate European and Canadian participation in IODP and to extend the scientific capability of the programme by providing support for mission-specific platform (MSP) operations. Specifically, ECORD is responsible for funding and implementing mission-specific platforms (MSP) for IODP expeditions. The ECORD Council is the funding entity, which coordinates a common ECORD approach to IODP policy with membership from 15 members (14 European countries and Canada). On behalf of the ECORD Council, the ECORD Managing Agency (EMA), administered by the Centre National de la Recherche Scientifique (CNRS), manages the participation of ECORD's members in IODP, represents the link between ECORD and the other IODP members, provides the central services for funds, and oversees the other ECORD entities.

The Italian participation in the international drilling programs IODP and ICDP (International Continental Scientific Drilling Program) is supported by the "CNR ECORD-IODP and ICDP Commission". This commission is dedicated to scientific and management coordination, as well as the promotion of activities supporting research and dissemination in the field of scientific drilling. Specifically, Italian participation in IODP (IODP-Italy) is linked to the membership in the ECORD consortium, made possible by dedicated funding from the Ministry of University and Research (MUR) to the ECORD Infrastructure managed by the National Research Council (CNR).

The ECORD Science Operator (ESO)[69] is the consortium of European scientific institutions formed to undertake Mission specific platform (MSP) operations for ECORD on behalf of the IODP. The MARUM, University of Bremen, provides the ESO Laboratory and Curarion Manager, who is responsible for all aspects of core curation and analytical facilities during offshore MSP operations. The Bremen Core Repository[70] (BCR) is the ECORD facility for core curation and management. The MARUM is also involved in data management tasks provided by PANGAEA (IODP-MSP data portal, Drilling Information System (DIS)).

[66] http://www.crs.ogs.it/bollettino_new/
[67] https://eida.ingv.it/en/
[68] https://www.ecord.org/
[69] https://www.ecord.org/about-ecord/management-structure/eso/
[70] https://www.marum.de/en/Research/IODP-Bremen-Core-Repository.html

5 Current state of the RIs FAIRness

5.1 What RIs do to be Findable?

The Findability principle declares that (meta)data should be easy to find for both humans and computers. Machine-readable metadata is essential for the automatic discovery of data and related services, so this is an essential component of the FAIRification process.
Each DO should be identified by a unique PID and described by machine-readable, rich, standardised metadata that clearly includes the persistent identifier (Jacobsen et al., 2020). Moreover, the metadata record should be indexed in a catalogue and contain the PID of the described DO. In the following sections, we addressed three major aspects: strategies for the use of PIDs, metadata schema adopted, and identification of the searchable resources in use.

5.1.1 Use of Persistent Identifiers (PIDs)

One best practice to guarantee the Findability of DOs is the use of PIDs. PIDs are defined as unique, machine-readable codes assigned to digital resources to make them easily discoverable and referenceable (Hellström, et al., 2020; de Castro et al., 2023). PIDs are intentionally created to remain constant regardless of any changes in the location or storage of the associated DOs. In essence, the allocation of these PIDs to DOs is crucial for ensuring their long-term findability and accessibility. This practice establishes an unwavering reference for researchers, facilitating the efficient sharing and reuse of DOs over the long term. The most widely adopted PIDs are, for example, the Digital Object Identifiers (DOI) for research data (Klump and Huber, 2017), the Open Researcher and Contributor ID (ORCID) for researchers, and the Research Organization Registry (ROR) for institutional affiliation.

Atmosphere - ACTRIS adopts some of the most widely used PIDs in DOs identification strategies, with difference among the data centre units and the required function. PIDs are mainly resumable as: the DOI and Handle for data, the ROR for institutional affiliation, the ORCID for persons, and the Handle for instruments. Furthermore, ACTRIS adopts different PID identifications depending on the data level and data type, using both DOI and Handle System as PIDs.
ICOS also adopted for data both DOI and the Handle System as PIDs. These PIDs contain the sha-256 checksum[71] of the DO and resolves into a landing page that contains the relevant metadata and a link to the DO. However, the difference in usage between the two PIDs is not specified. Metadata are presented through the web search interface DataCite Commons.
Meanwhile, SIOS, while sharing the use of DOI with the others RIs for data management, is distinguished by the additional use of UUID (Universally Unique Identifier) for metadata: a unique identifier is assigned to each metadata through the GeoNetwork[72] web interface. A UUID is a 128-bit identifier typically represented as a sequence of 32 hexadecimal digits, often separated into groups using hyphens (Singh&Huhns, 2005). We

[71] https://docs.precisely.com/docs/sftw/spectrum/ProductUpdateSummary/ProductUpdateSummary/source/about_sha256.html
[72] https://geonetwork-opensource.org/

point out that the UUID is not properly a PID, but can be listed as an additional strategy adopted by RIs[73].

Marine - Euro-Argo was the first network to apply a unique ID for each float (unique WMO numbers) and has worked with the Research Data Alliance (RDA) to develop a strategy for managing DOIs for continuously increasing datasets (DOI with a monthly tag for the GDAC - monthly snapshots)[74]. Initially, according to the suggestion of DataCite to include dynamic data, a DOI was set to describe the overall data set. In addition, specific DOIs were assigned to each monthly data snapshot that contains all the Argo data available at the time of the snapshot creation. In March 2016, to satisfy the Argo group that did not want to inflate the number of DOIs and to follow the new recommendations of the RDA, a new single Argo DOI was published using SEANOE (SEA scieNtific Open data Edition)[75]. This unique DOI quotes either the global data set or a specific snapshot. Each monthly snapshot is uploaded to SEANOE, which assigns a URL and a key. The metadata attached to an Argo float, containing information such as serial number of the sensors and other technical information, support analysis from the Argo GDAC, a strategy that has shown its efficiency when anomalies were detected. Moreover, Euro-Argo data management assigns DOIs to all its published documents. ORCID is also assigned to data owners who publish through SEANOE.
Eurofleets and JERICO make their data findable in at least two different data portals, EMODnet data portal and SeaDataNet. JERICO data are also findable through the CMEMS data portal. SeaDataNet uses the SEANOE to facilitate scientists to publish their research data. Each data set published by SEANOE has a DOI and assigns an ORCID to the authors of the datasets. Data published in SEANOE are also automatically duplicated to the EMODnet Data Ingestion portal. Finally, the use of PID for EMSO is not centralised and varies across the different regional facilities. However, the DOI resulted as the preferential PID for published datasets.

Terrestrial biosphere - IBISBA implemented FAIRDOM-SEEK[76], an open source web-based cataloguing and commons platform, for sharing heterogeneous digital research outputs, preserving associations between them as linked research objects[77]. By integrating the SEEK and openBIS platforms, it produces a FAIR data and model management facility for Systems Biology. Individual research assets (or aggregates of data and models) are identified with unique, resolvable, and persistent HTTP URLs (called SEEK-ID), which can be registered with DOIs for publication. All the resources published on the IBISBA Hub (DOs and associated metadata) are assigned a SEEK-ID. The resources that are deposited elsewhere, such as documentations and publications, are resolved using a DOI and/or a PubMed ID. AnaEE, eLTER and LifeWatch use a similar strategy. They implemented the use of UUID. To be resolvable, a UUID must be binded to a URL, which indicates where the object can be found. This is the case for the metadata exposed on the AnaEE Data Portal, including “packages” as well, i.e. a container for digital assets.

[73] https://faircookbook.elixir-europe.org/content/recipes/findability/identifiers.html#:~:text=A%20UUID%20can%20be%20used,persistent%20objects%20across%20a%20network.&text=Key%20Fact%20about%20UUID%3A,is%20required%20to%20administer%20them
[74] http://www.argodatamgt.org/Access-to-data/Argo-DOI-Digital-Object-Identifier
[75] https://www.seanoe.org/html/argo-floats-example-of-doi-attribution-policy.htm
[76] https://docs.seek4science.org/, http://fair-dom.org/about
[77] https://www.researchobject.org/

A similar approach is used by eLTER, that implemented the DEIMS.ID[78], a UUID-like PID bounded to an URL and automatically generated for describing the entities registered in DEIMS-SDR (e.g. networks, sites, sensors, activities). Entities associated with sites in DEIMS-SDR are identified by the ROR ID for institutional affiliation, and the ORCID for persons. Datasets are mainly identified by DOIs, pointing to an external data repository that provide this service (for instance a DOI generated by ZENODO or GBIF), or to B2SHARE[79]. By using eLTER Digital Assets Register (DAR)[80] is possible to access the metadata on datasets through landing pages, identified by UUIDs. LifeWatch Italy assigns UUIDs to datasets stored in the LifeWatch Italy Data Portal, and attach them as resolvable URLs. People are indicated in the metadata fields with ORCID, when available. One aspect to be taken into consideration with UUIDs is that the long-term maintenance of such PIDs is entirely dependent on the long-term persistence of the infrastructure that generates them. Additionally, for all of them a DOI might be assigned to some resources. For instance, in LifeWatch Italy DOI are assigned through the LifeWatch ERIC metadata catalogue only for DOs that are provided under CC-BY licence and are complemented by well-described and rich metadata. Since LifeWatch Italy also manages semantic artefacts, those are stored on the LifeWatch ERIC EcoPortal where they are assigned an URI as PID. Concepts and classes are identified with URI as PID as well, unless they are mapped with external resources.

Geosphere landsurface - Among all the websites from which SMINO data can be downloaded (see section 5.1.3), the Bulletin of the Seismometric Network of North Eastern Italy is the only one that associates a DOI for each dataset from 2015. The others do not report any information or expose any persistent identifier, except for the DOI related to the FReDNet Data Center itself [81].
As for ECORD, PANGAEA uses multiple internationally endorsed persistent identifier systems for individual scientific resources and information. They are the ORCID for the unambiguous identification of individual researchers, the ROR for institutional affiliation, the CORDIS[82], a persistent identifier for projects, and the Crossref Funder registry[83] for research funders. Moreover, PANGAEA also uses community-specific PIDs such as the International Generic Sample Number[84] (IGSN) and the Integrated Taxonomic Information System[85] (ITIS) for organisms (Felden et al., 2023).

5.1.2 Metadata schema

Metadata describes a broad range of information that allows DOs to be understood (Melton and Buxton, 2006). It provides a context for research findings, ideally in a machine-readable format, and requires the use of standardised sets of terms. To ensure that DOs are thoroughly described with complete metadata and that metadata clearly includes the PID, careful attention should be devoted to the choice of metadata schema (Jacobsen et

[78] https://deims.org/docs/deimsid.html
[79] https://eudat.eu/services/userdoc/b2handle
[80] https://catalogue.lter-europe.net/elter/documents
[81] see for example https://frednet.crs.ogs.it/DOI/
[82] https://cordis.europa.eu/projects/en
[83] https://www.crossref.org/services/funder-registry/
[84] https://www.igsn.org/
[85] https://www.itis.gov/

al., 2020). Metadata schema, in fact, not only enables effective data discovery and retrieval but also facilitates the understanding and evaluation of the DO.

Atmosphere - The three examined RIs employ distinct metadata schemas. Namely ACTRIS adopts the standard ISO 19115-2, the NetCDF CF-1.7, ECMA262-3 and the NcML. ICOS also adopts the metadata schema ISO 19115, and differently from all the other RIs also adopts specific templates for metadata based on the BADM[86] (Biological, Ancillary, Disturbance and Metadata) protocol: a detailed and standardised protocol in common with AmeriFlux, Fluxnet and other networks. Additional information are reported by Fluxnet network documentation[87]. Furthermore, for ocean data accessible through ERDDAP the RI follows the FGDC-STD-001-1998 and ISO19115-2/19139 metadata standards. SIOS, through its Italian node (IADC), uses the open-source GeoNetwork as an interface for data and metadata collection. The metadata standards adopted through this system are ISO 19115/19139. Additionally, as a specific metadata schema based on ISO 19115, SIOS adopts the MET Norway[88] schema. Furthermore, we point out that the support for DCAT-AP and schema.org is in progress both for ICOS and SIOS, an advancement that will facilitate data discovery and accessibility. Finally, with EUDAT's B2FIND search service, the RIs included in the BlueCloud project (ICOS and SIOS) or directly registered on EUDAT CDI (ICOS), aligned to the EUDAT Core Metadata Schema[89] (based on DataCite metadata schema) in order to transfer metadata information in a common schema.

Marine - As highlighted in Table 1, the currently most consolidated marine subdomain infrastructures use mutually compliant metadata schemas. Specifically, EMSO uses the FGDC-STD-001-1998 and ISO19115-2/19139 metadata standards throughout the ERDDAP page for the dataset exposed. EMSO Metadata Specifications[90] defines how (meta)data should be structured to be compliant with EMSO's data policy and it is applied to both NetCDF files and datasets served through ERDDAP. NetCDF CF is, indeed, the international standard for (meta)data and is the one adopted by all key European and international ocean-data management infrastructures (GDACs, CMEMS, EMODnet, SeaDataNet, etc.). All the considered marine RIs use the same core NetCDF CF structure with some specific infrastructure extensions (see for example the Euro-Argo user manual for NetCDF CF metadata schema specification[91]). Moreover, all the metadata of Euro-Argo, Eurofleets, and JERICO exposed on SeaDataNet portal are based on the ISO 19115 and ISO 19139 standards, with XML formats and exchange schema's (XSD), especially for making these INSPIRE compliant and for making the exchange between the national providers and the central portal operators more efficient by automatic harvesting. Different metadata schema of SeaDataNet are also used for reporting on cruises or field experiments at sea (Cruise Summary Reports, CRS), marine data sets and collections held within European research laboratories (European Directory of Marine Environmental Data, EDMED), marine research projects (European Directory of Marine Environmental Research Projects, EDMERP), marine observing and monitoring programmes, stations and platforms (European Directory of the Ocean Observing Systems, EDIOS), and common

[86] http://www.icos-etc.eu/icos/documents/templates
[87] https://fluxnet.org/badm-data-product/
[88] https://htmlpreview.github.io/?https://github.com/metno/mmd/blob/master/doc/mmd-specification.html
[89] https://docs.eudat.eu/b2find/#service-description; https://schema.eudat.eu/eudatcore/; https://schema.eudat.eu/eudatcore/ https://gitlab.eudat.eu/eudat-metadata/eudat-core-schema/-/blob/master/eudat-core.xsd
[90] https://github.com/emso-eric/emso-metadata-specifications/tree/develop
[91] https://archimer.ifremer.fr/doc/00187/29825/

marine data (Common Data Index, CDI)[92].
Euro-Argo, EMSO, and SeaDataNet in 2021, relying on joint ENVRI-FAIR[93] cross cutting developments, are developing a DCAT-AP standardised catalogue, description of its data services using the DCAT-AP protocol. Finally, the RIs included in the BlueCloud project and SEANOE (Euro-Argo, EMSO, JERICO, EuroFleets - through SeaDataNet and EMODnet) are planning to align to the EUDAT Core Metadata Schema XSD[94] (based on DataCite metadata schema) in order to transfer metadata information through different EUDAT CDI services (e.g, EUDAT's B2FIND search service).

Terrestrial biosphere - In the terrestrial biosphere subdomain, there is not a clear convergence toward a single implementation choice. Rather, each RI uses a different metadata schema, and most of them have more than one schema since they expose metadata for different DOs. AnaEE implemented DCAT, which is recommended by the W3C[95] in the framework of the ENVRI-FAIR project. To deal with geospatial information, metadata exposed on the eLTER DIP implemented the use of ISO 19115. The dataset metadata model for the DEIMS is based on Ecological Metadata Language (EML) metadata specification, version 2.1.1[96]. For the other types of resources described in DEIMS, eLTER seems to have developed different customised metadata models[97]. The metadata of the resources in the DEIMS-SDR are available as RDF and can be consulted through a TripleStore[98]. For data deposited in B2SHARE, eLTER uses a metadata block specific for its community[99]. LifeWatch ERIC metadata catalogue exposes metadata for dataset, research sites, VREs, workflows and services. A customised version of the Ecological Metadata Language v 2.2.0 is used for the datasets (60 metadata attributes), whereas for the other DOs one different customisation of the ISO1913 is used for each DO. IBISBA implemented the metadata standard recommended by the SEEK platform (O'Donohue 2023), which underpins the IBISBA Hub. Since IBISBA is using the model of Research Objects[100], it implements the Research Object Crate[101] (RO-Crate), a community effort to establish a lightweight approach to packaging Research Objects. It is based on Schema.org annotations in JSON for Linking Data (JSON-LD)[102]. For all of them, metadata are displayed in indexed metadata catalogues, and the PID of the resource that they point at is correctly displayed, when available. Notably, all the metadata catalogues are maintained at the European level, and only LifeWatch Italy preserves the metadata in the same national data portal in which the data are stored as well. Metadata from LifeWatch Italy are in turn also exposed on the ERIC metadata catalogue. LifeWatch is also the only RI that manages a repository of semantic artefacts, namely EcoPortal, used to create, store and manage semantic artefacts created by the LifeWatch Italy community and other semantic resources for the biodiversity and ecosystem domain. EcoPortal uses the

[92] https://www.seadatanet.org/Standards/Metadata-formats
[93] https://zenodo.org/records/10362974
[94] https://docs.eudat.eu/b2find/#service-description;
https://gitlab.eudat.eu/eudat-metadata/eudat-core-schema/-/blob/master/eudat-core.xsd
[95] https://www.w3.org/TR/vocab-dcat-2/
[96] https://deims.org/models/?id=dataset
[97] https://deims.org/models/
[98] http://sparql.lteritalia.it/
[99] https://b2share.eudat.eu/communities/LTER
[100] https://www.researchobject.org/
[101] https://www.researchobject.org/ro-crate/
[102] https://www.ibisba.eu/content/download/5486/47375/file/IBISBA_D7.3_IBISBAHub_Workflow_Repository_and_Portal.pdf

Metadata for Ontology Description and publication Ver. 2.0[103] (MOD) (Dutta et al., 2015) as metadata schema.

Geosphere landsurface - To retrieve the information about metadata schema employed within SMINO, all the different websites used to archive data were analysed (see section 4.1.3). The Bulletin of the Seismometric Network of North Eastern and the EIDA Italia are based on DataCite. Conversely, there is a lack of information regarding the metadata schema employed by the other websites, as indicated in Table 1. In general the choice of a metadata schema is often closely tied to the type of repository selected for storing DOs. In the case of PANGAEA, which is the long-term archive for MSP expeditions conducted by ECORD, the metadata schema adopted within the repository is the ISO 19115 [104].

5.1.3 Searchable resources

In the pursuit of maximising the findability of scientific DOs, the question of where the *objects* are indexed is paramount. F4 stipulates that DOs should be registered or indexed in a clear and unequivocal manner and the searchable resource provides the way by which a DOs can be discovered (Weigel et al., 2020). This means that DOs need to be stored or indexed in a public resource such as a data archive or institutional repository (Boeckhout et al., 2018). Repositories, (meta)data catalogues, and archives serve for organising, storing, and providing access to DOs and their associated metadata. Anyway, considering that the RIs analysed in this deliverable have very different degrees of development (see section 3.1), in this section, we are not only referring to the structured repository but also to all those systems that support the RIs in making the DOs findable (i.e., web interfaces, data discovery platform).

Atmosphere - The analysed RIs present different approaches for the management and sharing of data and related services. As described in section 4, ACTRIS has a distributed data centre but it is organised in six European thematic data centres which manage and offer access to data and data products (ACTRIS Data Center)[105]. One of these thematic centres, DC-ARES[106] is responsible for the collection and analysis of atmospheric aerosol remote sensing data. It is managed by the CNR-IMAA and make available their data through EARLINET data portal[107] and ACTRIS data portal. Beyond data, ACTRIS ARES provides a public page[108] for accessing tool packages along with their documentation to support researchers in managing and processing data. We also highlight the availability of software and Python code, both centrally managed (ACTRIS ERIC) and related to another data centre unit (CLU - Cloud remote sensing data centre Unit). These are respectively Python codes usable in the Jupyter environment (akin to a VRE) called ACTRIS Jupyter Hub[109], and a group of Python software designed for data processing, known as CloudnetPy[110]. On the other hand, ICOS presents a centralised management of data and related services, which are accessible through the ICOS Carbon Portal[111]. Then, similarly to

[103] https://www.isibang.ac.in/ns/mod/index.html
[104] https://wiki.pangaea.de/wiki/Data_set
[105] https://www.actris.eu/topical-centre/data-centre
[106] https://www.actris.eu/topical-centre/data-centre/ares-aerosol-remote-sensing-data-centre-unit
[107] https://data.earlinet.org/earlinet/login.zul
[108] https://repositories.imaa.cnr.it/public/
[109] https://github.com/ACTRIS-Data-Centre/actris-jupyter-hub
[110] https://github.com/actris-cloudnet/cloudnetpy
[111] https://data.icos-cp.eu/portal/

ACTRIS, ICOS also provides VREs through the Jupyter environment (Collaborative Jupyter Hub[112]). Furthermore, ICOS oceanic data are also available through the Blue-Cloud Data[113] Discovery and Access service, one of the two main components of the Blue-Cloud technical framework. Additionally, regarding the oceanic component of ICOS, there is a dedicated ERDDAP[114] data server. SIOS data and their associated metadata are managed by distributed data centres, whereby the Italian node contributes to the SIOS Data Access Portal[115] through the ERDDAP server of the IADC[116]. Nevertheless, as mentioned in section 5.1.2, they are also adopting GeoNetwork as a metadata catalogue. Furthermore, all three researched RIs maintain official GitHub profiles for codes, software, and services, all profiles managed at the centralised/European level (ACTRIS[117], ICOS[118] [119], SIOS[120]). Additionally, ICOS and SIOS are developing VREs through the Blue-Cloud portal, while ACTRIS will implement it as well as central level. Finally, excluding SIOS, both ACTRIS and ICOS provide semantic resources; all relevant information will be addressed in section 5.3.2.

Marine - JERICO and Eurofleets deposit their data in EMODnet and in the distributed Pan-European Marine Data Infrastructure SeaDataNet. Both these services have sought to address a solution to the challenges posed by the fragmented distribution of marine data. EMODnet acts as a European central hub, facilitating access to marine data from various sources and across seven discipline-based themes, integrating areas like hydrography, geology, and biology. SeaDataNet is a European distributed Marine Data Infrastructure for the management of large and diverse sets of data deriving from *in situ* observation of the seas and oceans. The online access to *in-situ* data and metadata products is provided through a unique portal interconnecting different interoperable node platforms constituted by the SeaDataNet data centres, managed by national research institutes, universities, and other professional entities. Integration of JERICO data is also in place with CMEMS repositories.
Euro-Argo and EMSO have opted to establish their own dedicated data portals, serving the specific needs of their marine community. The EMSO data is managed by a specific ERDDAP server that allows users to download subsets of gridded and tabular marine data and metadata in standard formats. The Euro-Argo fleet monitoring tool allows any user to visualise Argo profiling float metadata, ocean measurements, trajectories and technical parameters. It gives access to a fleet dashboard and also provides detailed information on a specific Argo float webpage (by World Meteorological Organization number). However, integration of Argo data from the Argo GDACs is already in place with the SeaDataNet, CMEMS, EMODNet repositories.
At present, the SEANOE data repository is the ingestion channel for data from Euro-Argo, SeaDataNet and EMODnet. Yet, the marine RIs considered in our deliverable are working to be findable also via the SEANOE searchable services. Common searchable resources for all the considered marine RIs are in development within EUDAT's B2FIND search service

[112] https://jupyter.icos-cp.eu/hub/login?next=%2Fhub%2F
[113] https://data.blue-cloud.org/search
[114] https://erddap.icos-cp.eu/erddap/info/index.html?page=1&itemsPerPage=1000
[115] https://sios-svalbard.org/metsis/search
[116] https://data.iadc.cnr.it/
[117] https://github.com/actris
[118] https://github.com/ICOS-Carbon-Portal
[119] https://github.com/icos-etc/
[120] https://github.com/SIOS-Svalbard

and Blue-Cloud Data Discovery service. Finally, regarding searchable resources for VRE, code, and software, Euro-Argo[121] and EMSO[122] have an official GitHub profile, instead all the considered marine RIs are developing thematic VREs that can be available through the Blue-Cloud services.

Terrestrial biosphere - Each RIs developed, or is developing, their own data portal and metadata catalogue at the centralised European level. The case of LifeWatch is rather unique in the RIs horizon, as the management and the storage of research data is done at the level of the national nodes that are in charge of implementing their own data portals. LifeWatch Italy currently have a data portal[123] that mainly collect datasets produced by the Italian scientific community, even if its geographic coverage extends beyond Italy. LifeWatch Italy data portal also exposes and provides the datasets metadata. Currently, LifeWatch Italy is releasing a new version of the data portal and a new metadata catalogue. In the new LifeWatch Italy metadata catalogue, beside metadata related to datasets, also metadata on scripts, VREs, workflows, services and research sites will be exposed. Metadata are then harvested and exposed on the centralised LifeWatch ERIC Metadata Catalogue[124] as well. It is a standard-based information management system based on GeoNetwork 3.10 (User manual V1.3[125]), designed and implemented to enable access to several resources (e.g. VREs, datasets, services, workflows and research sites) from a variety of providers through descriptive metadata. Moreover, the LifeWatch ERIC Metadata Catalogue allows (upon validation and verification) the creation of DOIs for resources that do not have it, by exploiting the GeoNetwork – DataCite connection. LifeWatch ERIC also maintains EcoPortal, a repository implemented to host semantic artefacts (mainly thesauri and ontologies) of the ecological and biodiversity domain. Notably, in a recent development, LifeWatch Italy upgraded the current version of EcoPortal with the aim of aligning it to the most advanced instances of the OntoPortal Alliance. It integrated a tool that automatically assesses the level of FAIRness of semantic artefacts within the portal (Ontology FAIRness Evaluation, O'FAIRe; Amdouni et al., 2022). On top of this, it is integrated with a collaborative editing tool (Vocbench) that allows users and managers of semantic resources deposited in EcoPortal to edit them, facilitating the mapping and the extension of such resources.
eLTER primary searchable resource is DEIMS-SDR[126]. As detailed in Wohner et al., 2019, DEIMS-SDR is an extended fork of the DEIMS system that was developed by the United States LTER network (Gries et al., 2010), which was an installation profile for the Content Management System Drupal7, and deployed as a GUI-based web application. It features both contributed modules and custom modules tailored to the description of ecological data. DEIMS-SDR exposes metadata for several entities: Sites, Datasets, Sensors, Activities. The datasets are then stored in different repositories external to eLTER. The main repository is a "community" nested in B2SHARE[127]. In parallel, the Italian node of eLTER (LTER-Italy) uses a ZENODO collection to store and centralise different types of DOs (datasets, softwares, reports, etc). Metadata from the Italian community points also to

[121] https://github.com/euroargodev
[122] https://github.com/emso-eric
[123] https://dataportal.lifewatchitaly.eu/data
[124] https://metadatacatalogue.lifewatch.eu/srv/eng/catalog.search#/home
[125] https://training.lifewatch.eu:9001/pluginfile.php/352/mod_resource/content/1/LifeWatch%20ERIC%20Metadata%20Catalogue_About.pdf
[126] https://deims.org/
[127] https://b2share.eudat.eu/communities/LTER

GBIF. With the aim of providing a visualisation platform of all the resources that are distributed over different portals, eLTER developed the Data Integration Platform (DIP)[128] (see Minić et al., 2018 for a complete overview). The development is funded by two European Union's Horizon 2020 projects: eLTER-Plus[129] and e-shape[130]. The landing page of the DIP incorporates a dynamic map that can show different base maps like Google or OpenStreetMap and layers which can be filtered by sites, stations, and on a complete set of filters related to the Essential Biodiversity Variables (Jetz et al., 2019). It enables exploration and visualisation of sensor time series data. It also allows to show information about sites, linking to the DEIMS UUIDs.
eLTER-DIP also provides a metadata catalogues, which is built on a GeoNetwork deployment. GeoNetwork is a free and open source cataloguing application for spatially referenced resources which allows to generate metadata through a web-interface in a variety of formats, enhancing DOs discoverability. DIP primarily points to the metadata exposed by the national networks in different data portals. The advantage is that the metadata exposed in DIP points directly to the data they describe. All the data uploaded to EUDAT B2SHARE service, belonging to theeLTER network is planned to be searchable through eLTER DIP using B2FIND search (Minić et al., 2018).
Moreover, eLTER is developing the Digital Asset Registry (DAR)[131], a web-based catalogue of digital assets, which include datasets, web services (e.g., OGC WMS/WFS), models and code/computational notebooks (e.g., Jupyter). Other types of assets can be added if requested. A relevant feature of the current implementation is the recording of the provenance of data. The datasets and their sources contribute to the catalogue, providing all the information needed to integrate datasets in higher-level data products. The assets recorded in the DAR can be searched with web services (API) or via a faceted search interface and connected to the eLTER sites through DEIMS-SDR. The AnaEE portal[132] is an instance of CKAN[133] while the IBISBA Hub[134] is based on FAIRDOM-SEEK[135], both open source data management systems. However, they are tailored to different communities and use cases, with CKAN being more general-purpose, while FAIRDOM-SEEK serving the specific needs of systems biology research. CKAN is a data management system that provides a powerful platform for cataloguing, storing and accessing datasets and their metadata, with a rich front-end, full API (for both data and catalogue), visualisation tools and more. FAIRDOM-SEEK is a mature web-based resource for organising, sharing and publishing heterogeneous scientific research datasets, models or simulations, protocols, workflows, samples, publications and other research digital outcomes. It preserves associations between them, along with information about the people and organisations.

[128] https://dip.lter-europe.net/#/home
[129] https://elter-ri.eu/elter-plus
[130] https://e-shape.eu/
[131] https://catalogue.lter-europe.net
[132] https://data.anaee.eu/
[133] https://ckan.org/
[134] https://hub.ibisba.eu/
[135] https://seek4science.org/

Geosphere Landsurface - SMINO has various websites that enable (meta)data retrieval. In fact, data are available through the Real Time Seismology (Saraò et al., 2009) of NE Italy[136], the Bulletin of the Seismometric Network of North Eastern Italy[137], the Archive System of Instrumental Seismology (OASIS)[138], and the Friuli Regional Deformation Network (FReDNet)[139]. These websites are all accessible from the dedicated SMINO webpage[140], but there is not a unified data portal for the RI. Furthermore, the data are available through the EIDA Italia[141]. Data managed by SMINO infrastructure are mainly real-time seismic events, site and waveform data, geodetic data, seismic waveform data and their relative metadata. Going into detail, the Real Time Seismology of NE Italy and the Bulletin of the Seismometric Network of North Eastern Italy are interconnected. The former provides real-time data, whereas the latter serves as a storage archive following a validation process. Both websites show information in a relational manner. Going forward, OASIS is the information system developed at OGS with the aim of organising, archiving, and providing access to the whole set of its seismological instrumental data. OASIS comprises a database, a double archive of digital waveforms, and a web interface, set up on a specifically devoted and redundant hardware infrastructure. OASIS is built on the open source DBMS MySQL and is structured into two parts: sites and waveforms. The archive of continuous waveforms is managed through the Boulder Real Time Technologies (BRTT) Antelope system, by its proprietary database named DATASCOPE. The web application was developed in Java language and Tomcat[142] is the web technology on which OASIS runs. The OASIS website provides descriptions of networks and archived data, consultation and query tools and data download functionality (Priolo et al., 2015).
FReDNet is a network of continuously operating GNSS (Global Navigation Satellite System) reference stations in northeast Italy, operated by the Center for Seismological Research (CRS) of OGS. The FReDNet Data Center (FReDNet DC) is a permanent archive and distribution centre primarily for digital data, located at CRS. A data server is in direct contact with the tracking sites and collects from them the GPS raw data (Zuliani, 2004; Zuliani et al. 2018).
On the other hand, EIDA implements a suite of standard FDSN web services to provide standardised and open access to data. The specifications and the usage of parameters of each service care is described by Strollo et al. (2021).
As described in the section 4, ECORD coordinates the European participation in IODP, which manages a vast amount of scientific data and (physical) samples from ocean drilling expeditions. IODP cores samples, drilled from the seafloor, are stored and curated at core repositories funded by the platform providers and based on their ocean of origin. In our analysis, we refer to the BCR-Bremen Core Repository, the unique repository managed at the European level, which stores data from the Atlantic and Arctic Oceans, as well as the Mediterranean, Black, and Baltic Seas. All BCR samples are entered into the curatorial database Drilling Information System (IODP BCR XDIS)[143], that is accessible to the general public for post-moratorium samples to identify where samples are located. The Drilling Information System is a modular structure of databases, tailored user applications

[136] rts.crs.inogs.it
[137] www.crs.inogs.it/bollettino_new
[138] oasis.crs.inogs.it
[139] frednet.crs.inogs.it
[140] https://www.ogs.it/it/sistema-di-monitoraggio-terrestre-dellitalia-nord-orientale-smino
[141] https://eida.ingv.it
[142] http://tomcat.apache.or
[143] https://xdis.marum.de/

as well as web services and instruments including appropriate interfaces to DIS. The main focuses are the data acquisition on drill sites (ExpeditionDIS), and the curation of sample material e.g., in core repositories (CurationDIS). The concept of the DIS is that it is used only during the lifetime of a drilling project, and that on completion of the project the data are transferred to a long-term data management and access system (Conze et al., 2007). Taking this into account, the document exclusively outlines the last as the reference repository for the ECORD infrastructure. The ECORD Science Operator selected PANGAEA as the long-term archive for expedition and post-expedition scientific data resulting from Mission Specific Platform (MSP) expeditions of the IODP (Conze et al., 2007). The technical architecture of PANGAEA follows a three-tiered client/server architecture with several clients and middleware components controlling the information flow and quality. A RDBMS (PostgreSQL) is used for the information storage management. For better performance high volume and binary data are stored in consistent formats on hard disk arrays and tape archives (Felden et al., 2023). Moreover, ECORD data deposited in PANGAEA are findable through the EUDAT B2Find search service[144].

Table 1. State of the art for Findability FAIR principles practices adopted by the most mature Research Infrastructures within the four environmental subdomains (atmosphere, marine, terrestrial biosphere and geosphere landsurface).

FINDABILITY				
Research infrastructures		**PID**	**Metadata schema**	**Searchable resource**
ATMOSPHERE	ACTRiS	Data: DOI, Handle Institutional Affiliation: ROR Researcher: ORCID Instruments: Handle	ISO 19115-2; NetCDF CF-1.7; ECMA262-3; NcML	ACTRIS data center https://www.actris.eu/topical-centre/data-centre ARES data portal (EARLINET) https://data.earlinet.org/earlinet/login.zul ARES tool packages: https://repositories.imaa.cnr.it/public/ ACTRIS GitHub: https://github.com/actris
	ICOS	Data: DOI, Handle System	Data: ISO19115; FGDC-STD-001-1998 and ISO19115-2/19139; BADM protocol	ICOS Carbon Portal: https://www.icos-cp.eu/data-services/about-data-portal ICOS ERDDAP (meta)data portal: https://erddap.icos-cp.eu/erddap/info/index.html?page=1&itemsPerPage=1000 ICOS GitHub: Carbon Portal - https://github.com/ICOS-Carbon-Portal ; ETC - https://github.com/icos-etc/ BlueCloud data discovery https://data.blue-cloud.org/search

[144] https://b2find.eudat.eu/organization/pangaea

				BlueCloud VRE https://blue-cloud.org/blue-cloud-virtual-research-environment B2FIND https://b2find9.cloud.dkrz.de
	SIOS	Data: DOI Metadata: UUID	Data: ISO 19115; ISO 19115/19139 MET Norway	SIOS data portal https://sios-svalbard.org/metsis/search IADC data portal: https://data.iadc.cnr.it IADC metadata portal: https://metadata.iadc.cnr.it/ SIOS GitHub: https://github.com/SIOS-Svalbard
MARINE	emso ERIC	Data: DOI (strongly suggested)	Data: FGDC-STD-001-1998 ISO19115-2/19139 NetCDF CF Metadata application profile: DCAT-AP	EMSO ERIC data portal: https://data.emso.eu EMSO GitHub: https://github.com/emso-eric
	Euro Argo	Data : DOI Researcher: ORCID (both via SEANOE service) Instruments: WMO	Data: NetCDF CF SeaDataNet CDI metadata based upon ISO19115/19139 EUDAT Core Metadata Schema (DataCite based) Researcher: ORCID record schema Metadata application profile: DCAT-AP	Euro-Argo ERIC data access https://www.euro-argo.eu/Argo-Data-access EMODNet data portal https://emodnet.ec.europa.eu/en SeaDataNet data portal https://cdi.seadatanet.org/search SEANOE https://www.seanoe.org/html/about.htm B2FIND https://b2find9.cloud.dkrz.de BlueCloud data discovery https://data.blue-cloud.org/search BlueCloud VRE https://blue-cloud.org/blue-cloud-virtual-research-environment Euro-Argo GitHub: https://github.com/euroargodev

		Data : DOI Researcher: ORCID (both via SEANOE service)	Data: NetCDF CF SeaDataNet CDI metadata schema based upon ISO19115/19139 EUDAT Core Metadata Schema (DataCite based) Researcher: ORCID record schema Metadata application profile: DCAT-AP (SeaDataNet)	EMODNet data portal https://emodnet.ec.europa.eu/en SeaDataNet data portal https://cdi.seadatanet.org/search CLIVAR and Carbon Hydrographic Data Office (GO-SHIP) https://cchdo.ucsd.edu SEANOE https://www.seanoe.org/html/about.htm BlueCloud data discovery https://data.blue-cloud.org/search BlueCloud VRE https://blue-cloud.org/blue-cloud-virtual-research-environment B2FIND https://b2find9.cloud.dkrz.de
		Data : DOI Researcher: ORCID (both via SEANOE service)	Data: NetCDF CF SeaDataNet CDI metadata schema based upon ISO19115/19139 CMEMS metadata schema based upon NetCDF CF convention EUDAT Core Metadata Schema (DataCite based) Researcher: ORCID record schema Metadata application profile: DCAT-AP (SeaDataNet)	EMODNet data portal https://emodnet.ec.europa.eu/en SeaDataNet data portal https://cdi.seadatanet.org/search Copernicus Marine Service In Situ TAC data portal http://www.marineinsitu.eu/ SEANOE https://www.seanoe.org/html/about.htm BlueCloud data discovery https://data.blue-cloud.org/search BlueCloud VRE https://blue-cloud.org/blue-cloud-virtual-research-environment B2FIND https://b2find9.cloud.dkrz.de
TERRESTRI		Metadata: UUID/URN Data: DOI, UUID/URN	Metadata application profile: DCAT-AP	AnaEE data portal https://data.anaee.eu/

ALBIOSPHERE	eLTER RI	Metadata: UUID Data: DOI Institutional Affiliation: ROR Researcher: ORCID	Data: ISO 19115 (GeoNetwork DIP); Customised metadata model (B2SHARE and DEIMS)	DEIMS-SDR (https://deims.org/search/datasets) Data Integration Portal (DIP) https://dip.lter-europe.net/#/metadata-catalogue Digital Asset Registry (DAR) Digital Asset Registry (DAR) https://catalogue.lter-europe.net/ B2Share eLTER repository https://b2share.eudat.eu/communities/LTER ZENODO LTER-Italy repository https://zenodo.org/communities/lter-italy/
	IBISBA	All the resources: SEEKID Documents and publications: DOI, PubMed ID	RO-Crate	IBISBA Hub https://hub.ibisba.eu/home/
	LifeWatch ERIC	LifeWatch ERIC metadata (that describe datasets): DOI Datasets, semantic artefacts, concepts: UUID/URI	Dataset on LifeWatch Italy data portal: EML2.2.0 VRE, workflows and services on the LifeWatch ERIC metadata catalogue: ISO 19139 Semantic artefacts on EcoPortal: MOD Ver. 2.0	LifeWatch Italy data portal https://dataportal.lifewatchitaly.eu/data LifeWatch ERIC metadata catalogue https://metadatacatalogue.lifewatch.eu/srv/eng/catalog.search#/home LifeWatch ERIC repository of semantic resources (EcoPortal) https://ecoportal.lifewatch.eu/
GEOSPHERE LANDSURFACE	Sistema di monitoraggio terrestre dell'Italia Nord Orientale (SMINO)	Data: DOI	Data: DataCite	Real Time Seismology of NE Italy https://rts.crs.inogs.it/ Bulletin of the Seismometric Network of North Eastern Italy http://www.crs.ogs.it/bollettino_new/ Archive System of Instrumental Seismology (OASIS) http://oasis.crs.inogs.it/ Friuli Regional Deformation Network (FReDNet) https://frednet.crs.ogs.it/ European Integrated Data Archive (EIDA) https://eida.ingv.it
	ECORD European Consortium for Ocean Research Drilling	Data: DOI Researchers: ORCID Institutional affiliation: ROR Projects: CORDIS Research funders: Crossref Funder registry Organisms: IGSN and ITIS	Data: ISO 19115	PANGAEA https://iodp.pangaea.de/ B2FIND https://b2find9.cloud.dkrz.de

5.2 What RIs do to be Accessible?

The Accessibility principle declares that DOs should be available and accessible to both people and automated systems. For this reason, the DOs and, in particular, the metadata records should be retrievable through appropriate and standardised communication protocol (Solle, 2020). In turn, that protocol should allow for authentication and authorization, where necessary. Finally, the associated metadata records should remain accessible even when the datasets they describe are not easily accessible.

5.2.1 Communication protocol

Standardised communication protocol establishes the rules and conventions for how data and metadata are exchanged and shared between information systems like users (clients) and the RIs web servers. The choice of the protocol can significantly impact the streamline data transmission and enhance data security (Jacobsen et al., 2020). For this reason, FAIR principles encourage free and open-sourced protocols to facilitate data retrieval and reuse. Moreover, despite not being explicitly considered in the statement of principle A.1, the next section will also describe the methods, services and tools useful to access the DOs (i.e., API, SPARQL, etc).

Atmosphere - In the context of accessibility and starting from the perspective of communication protocols, the three infrastructures pertaining to the atmosphere subdomain, for which we provide information, conform to standard HTTPS transfer protocols. Furthermore, all three RIs provide additional information in their respective DMPs. ACTRIS data is accessible through THREDDS (Thematic Realtime Environmental Distributed Data Service), using OPeNDAP, OGC WMS and WCS, HTTP[145]. The ACTRIS Data Discovery, Virtual Access and Services unit (DVAS) makes available REST APIs. API for ACTRIS Italian unit (ARES) are distinguished depending on Data and Metadata[146]. Similarly, for ICOS, all data is accessible through the ICOS Carbon Portal[147], that allows both human user access and machine-to-machine access. In the latter case, a RESTful API is available for data downloads, while a Python package is provided for metadata. Both metadata and data are accessible via APIs using standard web browsers, internet tools like wget or curl, as well as JavaScript or Python code. All data is made available through standard HTTPS transfer, and can be found through the interactive search interface on the Carbon Portal or through a data access mode offered by an "endpoint" service, based on the SPARQL language, as well as through other portals such as B2FIND. Finally, for SIOS, more general information is provided, indicating that data is made available by data centres contributing to the SIOS Data Management System (SDMS), and initially, there is no centralised data management. Moreover, SIOS provides a RESTful API service[148]. At Italian level, an ERDDAP data portal is provided, it acts like a gateway to uniform and standardise communication and data retrieval between data centres and users. To do so, IADC offers Rest APIs[149] and also provides Data Access Forms to help users to create the OPeNDAP requests[150].

[145] https://www.earlinet.org/index.php?id=285
[146] https://data.earlinet.org/api/services/restapi?_wadl
[147] https://www.icos-cp.eu/data-services/about-data-portal
[148] https://sios-svalbard.org/rest/stations/data.json
[149] https://data.iadc.cnr.it/erddap/rest.html
[150] https://data.iadc.cnr.it/erddap/index.html

Marine - European marine research infrastructures have streamlined their communication protocols to ensure efficient data and metadata access and sharing. The EMSO adopted the ERDDAP[151] server to facilitate both human interaction (OPeNDAP's Data Access Protocol[152], via HTTP communication protocol and WMS web service) and machine-to-machine communication (e.g, ERDDAP REST API[153]). For EMSO a specific web API is also under development for (meta)data access.

Euro-Argo employs a diverse set of servers with different communication protocols such as GDAC via both FTP and HTTPS; ERDDAP server with OPeNDAP's Data Access Protocol (via HTTP communication protocol) and WMS web service; THREDDS server, with OPeNDAP, OGC WMS, WFS, WCS, and HTTP[154]. Finally, a REST API (OpenAPI - swagger)[155], based on Apache 2.0, to ensure diverse levels of data and metadata accessibility is available. On the other hand, both Eurofleets and JERICO have standardised their communication methods when presenting (meta)data in SeaDataNet CDI using HTTP and HTTPS protocols, OGC WMS, WFS, WCS web services and APIs[156]. Moreover a SPARQL endpoint is also available for discovery of the Eurofleets and JERICO metadata in RDF format. Finally, for Eurofleets and JERICO (meta)data released through EMODnet are available through CSW, OGC WMS, WFS, WCS web service[157].

Terrestrial biosphere - HTTPS is by far the most common protocol implemented by the RIs in the terrestrial biosphere subdomain, and APIs are implemented to search, retrieve and deposit (meta)data. AnaEE and eLTER use CSW over HTTPS, which provides a means for discovering, querying, and managing metadata about geospatial data resources. Other web portals include RESTful API (IBISBA[158], LifeWatch ERIC metadata catalogue[159], eLTER metadata catalogue DIP[160], eLTER DEIMS-SDR[161], LifeWatch EcoPortal[162,163]). Metadata harvesting within the eLTER DIP relies on harvesting capabilities of a GeoNetwork server. It is used to collect metadata from any source within eLTER Network compatible with at least one of this mechanisms for repository interoperability: CSW 2.0.2 ISO Profile, OAI-PMH, Z39.50 protocols, Thredds, Webdav, Web Accessible Folders, ESRI GeoPortal, Other GeoNetwork node. An eLTER triple store is implemented for SPARQL queries[164], where all the metadata available in the DEIMS-SDR are described with RDF through ontologies (e.g. SSN, SOSA, SmOD, FOAF ontologies).

[151] https://coastwatch.pfeg.noaa.gov/erddap/information.html
https://www.opendap.org/pdf/ESE-RFC-004v1.2.pdf
[152]https://www.opendap.org/pdf/ESE-RFC-004v1.2.pdf
[153] https://ioos.github.io/erddapy/erddapy.html
[154] http://www.argodatamgt.org/Access-to-data/Thredds-data-serve
[155] https://dataselection.euro-argo.eu/swagger-ui.html; https://fleetmonitoring.euro-argo.eu/swagger-ui.html#
[156] https://cdi.seadatanet.org/#api
[157] https://emodnet.ec.europa.eu/en/emodnet-web-service-documentation
[158] https://hub.ibisba.eu/api
[159] https://metadatacatalogue.lifewatch.eu/doc/api/index.html
[160] https://dip.lter-europe.net/#/metadata-catalogue
[161] https://deims.org/api
[162] https://data.ecoportal.lifewatch.eu/documentation
[163] http://ecoportal.lifewatch.eu:8080/
[164] http://sparql.lteritalia.it/

Geosphere landsurface - As described above, SMINO has different access points to its related DOs and the communication protocols can vary. In particular, the communication protocol is HTTPS for the Real Time Seismology of NE Italy, FReDNet and EIDA while is HTTP for the Bulletin of the Seismometric Network of North Eastern Italy and OASIS. Going in details, OASIS allows the access to the site information and data downloads through the web portal[165].
Then, the FReDNet network informs that data related to its stations can be freely downloaded via HTTPS (HyperText Transfer Protocol Secure) and/or FTP (File Transfer Protocol soon to be decommissioned) in RAW and RINEX formats. Moreover, the FReDNet network reports that the RTK (Real-Time Kinematic) services are freely available through the Ntrip protocol via the caster managed by the Department of Seismological Research Center - CRS[166]. Ntrip (Networked Transport of RTCM via Internet Protocol) is a protocol for real-time transmission of correction data to GPS or GNSS receivers over the internet and it is crucial for achieving high-precision positioning.
EIDA data access is based on a portfolio of web services which continue to be extended through collaborative work among the data centres coordinated within ORFEUS and in synergy with the European Plate Observing System (EPOS) developments (Strollo et al. 2021). Waveform and station metadata from EIDA can be accessed through modern APIs and interfaces[167]. The API documentation of the FDSNWS Dataselect service[168] describe in detail how specific data series can be requested.
As for ECORD, direct access to DOs is provided via HTTPS content negotiation and REST APIs[169]. All PANGAEA APIs are available through a single entry point, using several techniques: REST-based endpoints (OAI-PMH, Elasticsearch) and SOAP APIs for well structured, proprietary access. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a protocol developed by the Open Archives Initiative which can be used to harvest (or collect) the metadata descriptions of PANGAEA so that services (like data portals) can be built.

5.2.2 Authentication & authorisation service

Openness is just one aspect of accessibility, and it may not be applicable or suitable for all types of DOs. In fact, it is worth mentioning that even DOs with high levels of protection and confidentiality can still adhere to FAIR principles (Landi et al., 2020). In some cases, it makes sense to require users to create a user account for accessing a repository. This approach not only ensures the verification of DO owners or contributors but also allows for the establishment of access rights tailored to individual users.

Atmosphere - Data access is open for 2 on the 3 examined infrastructures. Specifically, for ACTRIS, the ARES unit employs a Single Sign-On (SSO) authentication system based on the Central Authentication Service[170] (CAS) project, supporting various authentication protocols (CAS, SAML, OAuth, OpenID) and offering both username/password and Google credential authentication for access to ARES products. Resuming, the access to ACTRIS data is facilitated through a portal that just requires a quick registration process

[165] http://oasis.crs.inogs.it
[166] https://frednet.crs.ogs.it/EnglishSite/XFReDNetHomeENG.htm
[167] https://orfeus-eu.org/data/eida/webservices/
[168] https://www.orfeus-eu.org/swagger/dist/index.html?url=https://www.orfeus-eu.org/data/eida/webservices/dataselect/dataselect.yaml
[169] https://ws.pangaea.de/
[170] https://www.apereo.org/projects/cas

(EARLINET), as described in section 5.2.1. Nevertheless, the removal of the authentication process from the portal is currently under development. Once this step is completed, accessing the data is relatively straightforward. For ICOS all data are open, the only restriction concerns the raw data, which cannot be directly downloaded from the landing pages. Users must first contact the thematic centre to ensure the data interpretation and communicate their specific data needs. Finally, SIOS promotes free and open access to data to any person or any organisation who requests them[171], without any request for authentication.

Marine - Marine research infrastructures such as EMSO and Euro-Argo offer open access to their DOs. This means that anyone can freely access the datasets, metadata, and services provided by different infrastructures, promoting transparency and collaboration in the marine research community. On the other hand, accessing data from Eurofleets and JERICO requires a different approach. Specifically, access to their data through platforms like SeaDataNet and EMODnet is allowed by Marine-ID authentication and authorization service. This integrated system handles registration, authentication, and accounting for marine data users. Essentially, Marine-ID is designed to ensure data security and integrity while providing an efficient way for users to access it. The integrated authentication consists to implement a CAS SSO and manage the CAS API. The CAS API allows service providers to rely on Marine-ID, to authenticate the users connecting the service and get some personal attributes (first name, email address, affiliation, etc).

Terrestrial biosphere - In principle, all the (meta)data for which the licence allows the reuse, are openly available at the RIs portal without any authentication and authorisation protocol. In other words, any RI applies an Open data policy. However, all of them only allow authenticated users to upload DOs, and to use advanced services to analyse them. Among the terrestrial biosphere RIs, only LifeWatch Italy and ERIC allow the authentication through KEYCLOAK[172], an Identity Provider which mediates as SSO to the Data Portal, enabling login with Google and ORCID for LifeWatch Italy and EOSC-Federated for LifeWatch-ERIC[173]. The integration of a widely used external authentication and authorization services ensure a seamless user experience across RI portals and offers several advantages. These include simplifying user registration, minimising the necessity for multiple login credentials, enhancing data traceability, and improving authorship recognition by automating metadata attribute filling like "Creator" and "Contact." Regarding authorisation policies, LifeWatch ERIC and LifeWatch Italy use a stereotype-based system (hence grouping users according to predefined privilege levels) (de Natale et al., 2021). eLTER manages user authentication with two dedicated services, one for the DIP and one for DEIMS.

Geosphere landsurface - The websites through which SMINO enable data retrieval do not require authentication or authorisation service. The exception is OASIS, which informs that waveform data can be accessed with three levels of permission properties. The first level is for internal users and leaves free access to all OASIS data. The second level is devoted to project partners and it makes specific data-sets available through restricted login. Finally, the third level provides free access to a subset of the whole dataset only.

[171] https://sios-svalbard.org/sites/sios-svalbard.org/files/common/SIOS_Data_Policy.pdf
[172] https://www.keycloak.org/
[173] https://docs.egi.eu/users/aai/

Another specific case is the FReDNet network. To access the data, it is required to expressly declare that the user read and accepted the conditions that govern the access and use of the FReDNet network data. When declared that the conditions governing data access have been accepted, the website provides two possible alternatives for accessing the data. The first option is through HTTPS/FTP, where you manually search for and download the data via FTP. The second option is a simplified access with a Guided RINEX Data File Construction and immediate download.
Conversely, PANGAEA does not require any authentication offering a full open access to their DOs.

5.2.3 Metadata longevity plan

To make sure that metadata remains accessible even when the DOs they describe are not, a sustainable metadata longevity plan should be designed. Ideally, this point should be addressed within the DMP of each RI in order to implement specific strategies for the management and preservation of metadata (Li & Sugimoto, 2017). Frequently, when a preservation plan is available, it does not necessarily refer to metadata and a general reference to data is more common.

Atmosphere - As a European Research Infrastructure and more recently as ERIC RI, ACTRIS is designed as a long-term commitment by the member countries towards monitoring of short-lived components in the atmosphere. ACTRIS has a time horizon of at least 20 years from ERIC signature in 2023. With this background, the ACTRIS DC units commit to maintaining data and metadata records over this timeframe. Among the three atmosphere RIs, ICOS outlines the most robust metadata longevity plan. The RI is indeed designed for long-term sustainability, with a projected lifespan of at least 20-25 years, ensuring data operation and availability until 2040. Additionally, as indicated in the DMP, trusted repositories at B2SAFE, part of EUDAT at CSC and KFA Jülich, will preserve all ICOS data objects for the foreseeable future. B2SAFE will be included in the EOSC service portfolio, and funding is secured from the ICOS budget, which is guaranteed for the long term (more than 20 years). On the other hand, SIOS DMP states that cost estimates for long-term (meta)data preservation are not currently available[174]. SIOS utilises existing (meta)data streams and data centres, ensuring that these (meta)data are preserved since they serve purposes beyond SIOS.

Marine - Among marine RIs, only Euro-Argo has established its (meta)data longevity plan. The U.S. NODC volunteered to provide for the long-term archival of Argo data and information. JERICO and Eurofleets strategy is to make their DOs available to the entire marine community through the infrastructures of the major European integrators (SeaDataNet and EMODnet). JERICO longevity plan is also based on CMEMS. Consequently, the long-term preservation of the DOs and metadata is in charge of those entities.

Terrestrial biosphere - The preservation of the AnaEE data is ensured by the Data and Model Centre data storage, as detailed in the DMP[175]. As for eLTER, the technical data management is supported by the different institutions involved in the ICT core group,

[174] https://sios-svalbard.org/system/files/common/Documents/SIOS_Data_Management_Plan.pdf
[175] https://www.anaee.eu/sites/default/files/Mediatheque/Resources/reportsdocuments/std_-_v_7.1_-_appendix_6_-_16.10.2020.pdf

aggregated in the framework of several Horizon projects (H2020 eLTER-Plus and eLTER-PPP and HE eLTER-EnRich). Long-term preservation of data is therefore ensured by institutional commitment. At the best of our knowledge, IBISBA and LifeWatch do not have any related documentation explicitly dealing with (meta)data preservation. However, considering that LifeWatch is an ERIC and IBISBA is working towards the ERIC status[176], both have a long-term operational lifetime. IBISBA, that is currently in its preparatory phase[177], recently published the deliverables 8.2 and 8.5 in which a clear commitment to long-term data and metadata preservation is foreseen (Andrabi 2023; Admin, I. B. I. S. B. A., & Goble, C. 2022).

Geosphere landsurface - No metadata preservation policy has been found for SMINO. Conversely, PANGAEA website, where ECORD archive the DOs collected during the ocean drilling campaigns, state that "*the system guarantees long-term availability of its content through a commitment of the Alfred Wegener Institute, Helmholtz Center for Polar- and Marine Research (AWI) and the Center for Marine Environmental Sciences, University of Bremen (MARUM)*"[178]. Moreover, PANGAEA has a dedicated Preservation Plan webpage[179] where extensive information for data and metadata preservation are available.

Table 2. State of the art for Accessibility FAIR principles practices adopted by the most mature Research Infrastructures within the four environmental subdomains (atmosphere, marine, terrestrial biosphere and geosphere landsurface).

ACCESSIBILITY				
Research Infrastructure		**Communication protocol, methods and services**	**Authentication & authorisation service**	**Metadata longevity plan**
ATMOSPHERE	ACTRiS	HTTPS; THREDDS (OPeNDAP, OGC WMS, OGC WCS); ACTRIS DC REST API for (meta)data	Authentication system CAS (SAML, OAuth, OpenID)	ACTRIS DC units commit to maintaining metadata records for 20 years after ERIC signature in 2023
	ICOS	HTTPS; SPARQL endpoint	Open data	Data operation and availability ensured until 2040. B2SAFE funding is secured from the ICOS budget (guaranteed for > 20 years)
	SIOS	HTTPS; ERDDAP (OPeNDAP, WMS, REST API)	Open data	N/A

[176] https://www.ibisba.eu/News-and-Events/A-Major-Step-Towards-IBISBA-ERIC
[177] https://www.ibisba.eu/EU-Projects/PREP-IBISBA
[178] https://www.pangaea.de/about/terms.php
[179] https://wiki.pangaea.de/wiki/Preservation_Plan

MARINE	emso ERIC	ERDDAP (OPeNDAP, WMS, REST API) Specific REST API for (meta)data *in develop*	Open Data	N/A
	Euro Argo	GDAC (FTP, HTTPS) ERDDAP (OPeNDAP, WMS, REST API) THREDDS (OPeNDAP, OGC WMS, WFS, WCS) REST API	Open Data	U.S. National Oceanographic Data Centre (NODC) volunteered to provide for the long-term archival of Argo data and metadata
	Eurofleets+	***SeaDataNet*** HTTPS, HTTP, OGC WMS, OGC WFS, OGC WCS, SPARQL endpoint REST API for (meta)data ***EMODnet*** CSW OGC WMS, WFS, WCS ERDDAP THREDDS	Marine ID service	Long-term preservation of the data and metadata are covered by SeaDataNet and EMODnet
	JERICO RESEARCH INFRASTRUCTURE	***SeaDataNet*** HTTPS, HTTP, OGC WMS, OGC WFS, OGC WCS, SPARQL endpoint REST API for (meta)data ***EMODnet*** CSW OGC WMS, WFS, WCS ERDDAP THREDDS ***CMEMS*** HTTPS, HTTP, OGC WCS, OPENDAP, FTP, CMEMS API	Marine ID service	Long-term preservation of the data and metadata are covered by SeaDataNet, EMODnet and CMEMS.

TERRESTRIAL BIOSPHERE	AnaEE ERIC	CSW HTTPS	Open Data GeoNetwork account in the internal catalogue database	Preservation of the AnaEE infrastructure datasets is ensured by the DMC data storage and its mirror storage in the CC IN2P3. https://www.anaee.eu/sites/default/files/Mediatheque/Resources/reportsdocuments/std_-_v_7.1_-_appendix_6_-_16.10.2020.pdf
	eLTER RI	OGC CSW HTTPS REST API OAI-PMH OGC WFS OGC WMS OGC SOS Z39.50	Open Data	The technical data management is supported by the different institutions involved in the ICT core group, aggregated in the framework of several Horizon projects (H2020 eLTER-Plus and eLTER-PPP and HE eLTER-EnRich)
	IBISBA	HTTPS	Open Data	N/A
	LifeWatch ERIC	HTTPS REST API SPARQL endpoint	Open Data	N/A
GEOSPHERE LANDSURFACE	Sistema di monitoraggio terrestre dell'Italia Nord Orientale (SMINO)	HTTPS HTTP FTP Ntrip REST API (EIDA)	Open data; different level of permission restricted login (OASIS); required to accept the terms governing data access (FReDNet)	N/A
	ECORD European Consortium for Ocean Research Drilling	HTTPS; OAI-PMH; SOAP API	Open data	PANGAEA Preservation Plan https://wiki.pangaea.de/wiki/Preservation_Plan

5.3 What RIs do to be Interoperable?

The Interoperability principle emphasises the importance of enabling the accurate exchange and integration of DOs across different systems and platforms. In short, Interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged (IEEE 1990). This includes either the capability of exchanging information, which is technical or syntactic interoperability, and the ability to exchange information with unambiguous, shared meaning, i.e. the capability of the recipient to use that information, which is the semantic interoperability (Benson et al., 2016). It encompasses the adoption of standardised and machine-interpretable languages, formats, and protocols for the structured organisation of information. Furthermore, this principle encourages the use of common and shared semantic artefacts (which could range from simple taxonomies to ontologies) to represent the content and the meaning of the DOs. Finally, it expresses that (meta)data should include qualified references to other (meta)data, thus suggesting to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about DOs (see the GO FAIR Foundation website[180] for a detailed explanation). The issue of metadata interoperability has been tackled early in European projects (Wittenburg et al., 2004). However, metadata is just one component. Achieving data and other DOs interoperability is fundamental. Currently, the EOSC is tackling this aspect, as it is essential for building a federated space for digital services in Europe[181] (European Commission DG for Research and Innovation 2022). Starting from the EOSC Future Interoperability Framework (Corcho et al. 2021), the Semantic Interoperability Task Force[182] and the Technical Interoperability of Data and Services Task Force[183] are now operating to propose recommendations and examples of adaptation hints for major existing solutions.

5.3.1 (Meta)data format

"*The main goal of this principle is to provide a "common understanding" of DOs by means of a language for knowledge representation to be used to represent these objects*"[184]. Knowledge representation languages are formal systems used to represent and encode knowledge in a structured and machine-readable format. In terms of syntactic interoperability, it means that data and metadata are provided in a common and shared format that can be read by machines as well. The format and extension are dependent on the type of DO. For instance, for metadata JSON and RDF/XML are common standard formats among domains. For datasets, taking apart formats adopted for specific use, NetCDF is adopted particularly among the atmospheric and marine subdomains, while CSV is common in terrestrial biosphere and geosphere landsurface. Furthermore, the analysed RIs manage other several types of DOs, such as semantic artefacts. These can be available in XML or resource-specific formats (e.g., OWL/XML, or Turtle). In the

[180] https://www.go-fair.org/fair-principles/

[181] https://eosc-portal.eu/eosc-interoperability-framework/about-eosc-interoperability-framework-governance-eosc-if

[182] https://eosc.eu/advisory-groups/semantic-interoperability

[183] https://eosc.eu/advisory-groups/technical-interoperability-data-and-services

[184] https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/

following section, we describe the (meta)data formats used within the RIs when users upload their data, and the formats that are available by the RIs portals for download or machine access.

Atmosphere - ACTRIS presents all metadata in JSON format (also accessible via API), while NetCDF4 format has been adopted for the data files. Data can be provided in real-time (RRT) and non-real-time (NRT). All provided data conform to the NetCDF4 format, following Climate Forecast (CF) 1.7 conventions. The purpose of the CF conventions is to require conforming datasets to contain sufficient metadata that they are self-describing in the sense that each variable in the file has an associated description of what it represents, including physical units if appropriate, and that each value can be located in space (relative to earth-based coordinates) and time. ICOS metadata are recorded in the following formats: JSON, RDF/XML, RDF/Turtle, and XML. Then, the data format, as a general rule, is always presented as semicolon-separated ASCII files. SIOS metadata are downloadable as XML. Additionally, SIOS also adopts the MET Norway Metadata Format[185] (MMD), which is an XML metadata format for storing information about scientific datasets. Regarding the data, these are also provided by following NetCDF CF formats.

Marine - The GDAC data management team suggested that all ocean observations should be universally available through CF-compliant NetCDF files (Obaton et al., 2022). Over the past ten years, the discrete sampling geometries (DSG) were added to the CF standard. These geometric forms are intended to represent *in situ* features like time series, vertical profiles, and surface trajectories. Such a format is already common for data outputs from the Euro-Argo marine system. JERICO and Eurofleets, within the European SeaDataNet and EMODnet groups, have chosen the Ocean Data View (ODV) ASCII format. They have also used SeaDataNet NetCDF for observational data (like profiles, time series, and sea surface variables) and NetCDF for outputs like climatology. Within CMEMS, JERICO has chosen the CF-compliant NetCDF data format. EMSO opted for the ERDDAP tool, which allow scientists to work in their format of choice, but makes the data available through interoperable formats, such as NetCDF and Web services (but also such as HTML table, CSV, TXT, MAT, JSON, etc.).
Regarding the metadata, the most common format is NetCDF CF Metadata Conventions[186] (Euro-Argo, EMSO, JERICO, and Eurofleets). EMSO uses the XML metadata format. However, ERDDAP tool allow the export of metadata also into .csv, .htmlTable, .itx, .json, .jsonlCSV1, .jsonlCSV, .jsonlKVP, .mat, .nc, .nccsv, .tsv, .xhtml extensions. The metadata of JERICO and Eurofleets published on SeaDataNet have XML and RDF formats. Finally, for all the considered marine RIs that expose their data on SEANOE, the metadata are also available in TXT, RIS, XLS, RTF, BIBTEX formats.

Terrestrial biosphere - IBISBA serves as a versatile platform that manages a wide range of DOs. These DOs encompass data files, models, SOPs, workflows, publications, documents, file templates, and collections. It shares with AnaEE the notable features of accepting data in various file formats, allowing users to upload and download files just as

185 https://htmlpreview.github.io/?https://github.com/metno/mmd/blob/master/doc/mmd-specification.html#mmd-to-iso-mapping

186 http://cfconventions.org/

they are. However, it is worth noting that for both the RIs, while the DO files can be downloaded directly, the metadata associated with them cannot be downloaded separately. To allow full interoperability, they need to be available as distinct machine-actionable files alongside the data. LifeWatch Italy and eLTER share the same technology for their metadata catalogues (GeoNetwork). However, while for LifeWatch Italy metadata are stored and downloadable in XML format and the DOs corresponding to datasets are either in CSV UTF-8 encoding or TXT extensions, for eLTER the situation is more articulated. eLTER allows users to download metadata in different formats. Metadata of entities, such as sites, activities, sensors, through DEIMS-SDR, are provided in JSON format, while metadata of dataset provided by DIP through GeoNetwork are provided in 19139 and RDF XML formats. All files uploaded to the B2Share repository through DEIMS are available for end users through the DIP search. Metadata and data search is done using B2FIND EUDAT service. It uses APIs for communication and service responds with messages formatted using JSON (JavaScript Object Notation). The data that are accommodated belong to different formats (ASCII, ESRI Shapefile Format, NetCDF, OGC-SOS, OGC-WMS, PNG, ZIP, etc.). Finally, the metadata for digital products hosted on ZENODO is not integrated into the metadata catalogue. For such files, both data and metadata align with ZENODO's implementation. LifeWatch is the only RI that manages and exposes semantic artefacts in a systematic way through the EcoPortal. Those are mainly ontologies and thesauri in OWL and SKOS respectively. SKOS thesauri are downloadable in RDF, RDF/XML, and CSV, while OWL ontologies are downloadable in OWL/XML or RDF/XML, and CSV. Metadata are available as JSON, JSON-LD, N-triple, and RDF/XML file formats.

Geosphere landsurface - Given the several websites where data is deposited, SMINO presents different formats for (meta)data. The Real Time Seismology of NE Italy offers the possibility to download data in KML and CSV formats, but no information about metadata could be retrieved. The format for the Bulletin of the Seismometric Network of North Eastern Italy is QuakeML. This format is considered a flexible, extensible and modular XML representation of seismological data which is intended to cover a broad range of fields of application in modern seismology. The relative metadata are available in JSON and XML. As for the Archive System of Instrumental Seismology (OASIS), the user can download both “event waveforms" and “continuous waveform”. As for “event waveform”, both instrument-corrected and uncorrected data in SAC or ASCII format can be downloaded, as well as the dataless file which contains the whole information on instruments for processing data correctly. “Continuous waveform” data are provided in SAC and/or MiniSEED format together with the dataless files. MiniSEED is a stripped down version of SEED containing only waveform data while a dataless SEED file is the metadata counterpart to miniSEED. It contains only station and channel metadata, without any time series values. From the FReDNet Network, it is possible to download RAW files, RINEX files and Site Logs. In particular, RINEX (Receiver Independent Exchange Format) is a standardised data format used for GNSS data exchange which facilitates precise positioning and post-processing applications. No information about metadata seems available on the website. Lastly, EIDA uses miniSEED for waveforms while JSON and XML format for metadata. Switching to ECORD, the data and metadata formats are those used by PANGAEA. PANGAEA's preferred data format is TAB-delimited text files (UTF-8 encoding). The repository offers the opportunity to download the data in many other types of formats. Metadata records are directly accessible in schema.org (in JSON-LD) and the other XML-based serialisations.

5.3.2 Semantic artefacts

Semantic artefacts are machine-readable, machine-interpretable and machine-actionable formalisations of concepts that can be used and exchanged to encode and decode information in a predictable way, thus enabling the discovery, integration and reuse of information by both humans and machines (Le Franc et al., 2022). Semantic artefacts may have a broad range of formalisations, from loose sets of terms such as lists, glossaries, and categorisation schemes to higher-order logic constructs such as thesauri and ontologies (Corcho et al., 2023; Zeng, 2008). To be useful, though, they must be used to describe the metadata fields or the data variables. This section describes the semantic artefacts used by the analysed RIs.

Atmosphere - Regarding semantic artefacts, ACTRIS has two semantic resources published on the SKOSMOS platform[187]: ACTRIS Vocabulary and ACTRIS Controlled Lists. The first one includes 2110 concepts and 327 alternate terms, while the second one contains 77 concepts and 6 alternate terms. These are standardised machine-actionable vocabularies to identify data and metadata entities, including variable names. ACTRIS semantic resources include the standard names of the Climate Forecast (CF)[188] convention, and the Global Change Master Directory (GCMD)[189] keywords. ACTRIS actively participates in the ENVRI framework to develop a commonly accepted vocabulary for variables in the subdomain, contributing to maintaining and completing the CF standard name and GCMD keywords vocabularies. ICOS has a well-structured open ontology describing metadata attributes of the ICOS Carbon Portal. This semantic resource was originally developed between 2009 and 2011 by the W3C Semantic Sensor Networks Incubator Group (SSN-XG[190]). Subsequently, the ontology was revised and modularised between 2015 and 2017 by the W3C/OGC[191] Spatial Data on the Web Working Group. Unlike the first two RIs, SIOS does not directly manage semantic resources. However, it makes use of terminologies, standard names, and controlled vocabularies from the Climate and Forecast Convention.

Marine - All the considered RIs of the marine subdomain use controlled vocabularies of the NERC Vocabulary Server (NVS), which is technically managed and hosted by the British Oceanographic Data Centre (BODC). The NVS is a SKOS-based platform that conforms to a set of standards designed to be compatible with Semantic Web tools and technologies, allowing its content to support the interoperability of datasets on the Web known as Linked Data[192]. Within the NVS system, collections, concepts, and concept schema are identified with URIs in a standardised syntax. Each collection corresponds to a controlled vocabulary or code list, each concept is a unique term with a definition, and each schema can be considered as a group of one or more related concepts. Concept schema also incorporates the semantic relationships between individual concepts. The NVS web user interface uses VocPrez and can be accessed through the link[193].
Euro-Argo makes use of the specific Argo vocabularies. Originally known as reference tables, they are used extensively to constrain the contents of the Argo NetCDF files, and to

[187] https://vocabulary.actris.nilu.no/skosmos/en/
[188] https://cfconventions.org/
[189] https://www.earthdata.nasa.gov/learn/find-data/idn/gcmd-keywords
[190] http://www.w3.org/2005/Incubator/ssn/
[191] https://www.w3.org/2015/spatial/wiki/Semantic_Sensor_Network_Ontology
[192] https://www.w3.org/standards/
[193] http://vocab.nerc.ac.uk

enhance the interoperability of the Argo data system. The transfer of existing reference tables to NVS collections, as well as the creation of new ones, is an ongoing effort managed by the Argo Vocabulary Task Team and supported by BODC. Argo vocabularies on the NVS can be accessed through this link[194].
EMSO, JERICO, and Eurofleets use the Common Vocabularies of SeaDataNet[195] for their (meta)data. The vocabulary services are technically managed and hosted by the BODC by means of the NVS. Specifically, the controlled vocabularies predominantly used are: P01, P02, P06, L05, L06, L22, L35 (by EMSO), P01, P02, P06, L05, L06, L22, L35 (by JERICO), and P02, C17 C19 (by Eurofleets).
Finally, the European Database of Marine Organization (EDMO)[196] registry, managed by SeaDataNet in collaboration with Eurofleet and other institutes, is used by all the considered RIs.

Terrestrial biosphere - The RIs of the terrestrial biosphere subdomain make use of several semantic artefacts to describe their digital resources and the associated metadata. Those include the resources listed in the EMBL-EBI Ontology Lookup Service[197] (IBISBA), those listed in the NVS[193] (eLTER, LifeWatch Italy - e.g. P01), and Darwin Core[198] (LifeWatch Italy). Along with the use of the available resources, some RIs had developed their own resources in order to describe specific research areas. IBISBA maintains a glossary[199], mainly for internal use. eLTER Europe developed and maintains three semantic resources available through a SKOSMOS platform[200]: eLTER_CL, a thesaurus for controlled lists used by the eLTER community; EnvThes, a controlled vocabulary or terminology resource in the field of environmental science and biodiversity; eLTER Standard Observations, a resource to standardise observations and associated methods that is currently in the developmental stage. Metadata exposed on eLTER DIP are semantically enriched using terms from the EnvThes. LifeWatch ERIC maintains EcoPortal[201], a repository which exposes semantic artefacts and their associated metadata created by external users, as well as resources that are created by the LifeWatch community. LifeWatch Italy primarily uses the semantic artefacts stored in EcoPortal. One of the more interesting in this contest, it is the LUPO Ontology (LifeWatch Upper Ontology), specifically designed to deal with the RIs services and integrated into the LifeWatch Italy Semantic Platform[202]. The Italian node is responsible for the development of domain-specific thesauri (e.g. Fish Traits Thesaurus, Zooplankton Traits Thesaurus, Phytoplankton Traits Thesaurus, etc.) used to annotate (meta)data. Within the framework of the AnaEE-France, the AnaEE Thesaurus has been developed[203], but it seems it is not used to annotate data and metadata in the AnaEE Data Portal.

Geosphere landsurface - SMINO does not have any specific information regarding the use of semantic resources for data and metadata description for any of the sites where data

[194] https://vocab.nerc.ac.uk/search_nvs/cvl/?searchstr=argo&options=governance
[195] https://www.seadatanet.org/Standards/Common-Vocabularies
[196] https://edmo.seadatanet.org/search
[197] https://www.ebi.ac.uk/ols4
[198] https://dwc.tdwg.org/list/
[199] https://ibisba.github.io/handbook/terminology.html
[200] https://vocabs.lter-europe.net/en/
[201] https://ecoportal.lifewatch.eu/
[202] https://semantics.lifewatchitaly.eu/resource/Start
[203] https://agroportal.lirmm.fr/ontologies/ANAEETHES?p=summary

is accessible. As for ECORD, PANGAEA links each observation with terms/concepts from controlled and internationally recognised vocabularies and ontologies as a measure of standardisation (Felden et al., 2023). In general, this allows searching for synonyms or broader terms of concepts used in datasets, and to allow for filtering search results by using broader term relations. The semantic artefacts principally used within PANGAEA are mainly QUDT[204], PATO[205], EnvO[206], and ChEBI[207].

Table 3. State of the art for Interoperability FAIR principles practices adopted by the most mature Research Infrastructures within the four environmental subdomains (atmosphere, marine, terrestrial biosphere and geosphere landsurface).

INTEROPERABILITY				
Research Infrastructure		**Format**		**Semantic artefacts**
		data	**metadata**	
ATMOSPHERE	ACTRiS	NetCDF CF	JSON (accessible via API); NetCDF CF	ACTRIS Vocabulary and ACTRIS Controlled Lists (CF standard names and the keywords from GCMD)
	ICOS	Semicolon-separated ASCII files	JSON; RDF/XML; RDF/Turtle; and XML	Ontology of Integrated Carbon Observation System (ICOS Ontology)
	SIOS	NetCDF CF	MET Norway Metadata Format (MMD)	Use of CF terminologies, standard names, and controlled vocabularies
MARINE	emso ERIC	Data: .html table, ESRI .asc and .csv, Google Earth .kml, OPeNDAP binary, .mat, .nc, ODV .txt, .csv, .tsv, .json, and .xhtml. ia the ERDDAP RESTful web service.	XML, Metadata informations are available also in other file formats (.csv, .htmlTable, .itx, .json, .jsonlCSV1, .jsonlCSV, .jsonlKVP, .mat, .nc, .nccsv, .tsv, .xhtml) via the ERDDAP RESTful web service.	SeaDataNet Common Vocabularies (NVS - libraries P01, P02, P06, L05, L06, L22, L35)

[204] https://www.qudt.org/
[205] http://obofoundry.org/ontology/pato.html
[206] http://environmentontology.org/
[207] https://www.ebi.ac.uk/chebi/

		NetCDF CF, ASCII	NetCDF CF XML, RDF via SeaDataNet Metadata informations are available also in other file format (TXT, RIS, XLS, RTF, BIBTEX) via SEANOE	Argo vocabulary server (NVS - libraries R01, RR2, RD2, RP2, R03, R04, R05, R06, R07, R08, R09, R10, R11, R12, R13, R14, R15, RTV, RMC, R16, R18, R19, R20, R21, R22, R23, R24, R25, R26, R27, R28, TBC, R30)
		NetCDF CF, .html table, ESRI .asc and .csv, Google Earth .kml, OPeNDAP binary, .mat, .nc, ODV .txt, .csv, .tsv, .json, and .xhtml.	NetCDF CF XML, RDF, via SeaDataNet Metadata informations are available also in other file format (TXT, RIS, XLS, RTF, BIBTEX) via SEANOE	SeaDataNet Common Vocabularies (NVS - mainly libraries P02, C17 and C19)
		NetCDF CF, .html table, ESRI .asc and .csv, Google Earth .kml, OPeNDAP binary, .mat, .nc, ODV .txt, .csv, .tsv, .json, and .xhtml.	NetCDF CF 1.7 XML, RDF, via SeaDataNet Metadata informations are available also in other file format (TXT, RIS, XLS, RTF, BIBTEX) via SEANOE	SeaDataNet Common Vocabularies (NVS - mainly libraries P01, P02, P06, L05, L06, L22, L35)
TERRESTRIAL BI		NetCDF	XMLS | eXtensible Markup Language Schema	AnaEE Thesaurus
		ASCII, ESRI Shapefile Format, NetCDF, OGC-SOS, OGC-WMS, PNG, ZIP	JSON XML, RDF	eLTER_CL EnvThes eLTER Standard Observations resources listed in the NVS http://vocab.nerc.ac.uk/collection/

OSPHERE	IBISBA	FASTA, Genbank, SBML mzML, CSV (delimited text files), FASTQ, Excel, SBML V3, SBOL V2, InChI string, PDF, Word, protocols.io, CWL, Knime, Galaxy	JSON-LD	resources listed in the EMBL-EBI Ontology Lookup Service (https://www.ebi.ac.uk/ols4) glossary for internal use (https://ibisba.github.io/handbook/terminology.html)
	LifeWatch ERIC	Data: CSV UTF-8 TXT Semantic artefacts: RDF, RDF/XML, OWL/XML RDF/XML, CSV	Dataset: XML Semantic artefacts metadata: JSON, JSON-LD, N-triple, RDF/XML	-Fish Traits Thesaurus -Zooplankton Traits Thesaurus -Phytoplankton Traits Thesaurus -LUPO (available at https://ecoportal.lifewatch.eu/) Darwin Core https://dwc.tdwg.org/ NVS P01 http://vocab.nerc.ac.uk/collection/P01/current/
GEOSPHERE LANDSURFACE	Sistema di monitoraggio terrestre dell'Italia Nord Orientale (SMINO)	Real Time Seismology of NE Italy: KML and CSV Bulletin of the Seismometric Network of North Eastern Italy: QuakeML Archive System of Instrumental Seismology (OASIS): SAC, ASCII, miniSEED Friuli Regional Deformation Network (FReDNet): RAW and RINEX European Integrated Data Archive: miniSEED	no information JSON and XML Dataless SEED no information JSON and XML	no information

	ECORD European Consortium for Ocean Research Drilling	TAB-delimited text files	JSON and XML	mainly -QUDT, -PATO, -EnvO, -ChEBI

5.4 What RIs do to be Reusable?

The Reusability principle underscores the necessity for DOs to be formatted and annotated in a manner that facilitates their effective utilisation (Jacobsen et al., 2020). This entails providing comprehensive metadata and origin information, along with transparent documentation, as well as explicit permissions for reuse (Wolf et al., 2021). This principle promotes the idea that DOs should not be locked into specific projects or contexts but should be designed for broader and long-term usability. The clarity of licensing status is becoming more important with automated searches involving more licensing considerations. The conditions under which the DOs can be used should be clear to machines and humans. As described in Figure 2, the Reusability principle includes four sub-principles. In this section, only one of these will be addressed, namely Meta(data) licences (R1.1), which is realised through the following section. The analysis of the practices related to the remaining three have already been described in the previous sections, specifically: (Meta)data schema (section 5.1.2) for sub-principles R1, R1.2, and R1.3; (Meta)data format (section 5.3.1) and semantic artefacts (section 5.3.2) for R1.3.

5.4.1 Licence

The necessity of accompanying DOs with an explicit and readily accessible usage licence is essential for providing a legal framework that delineates the terms and conditions governing the utilisation of the DOs, ensuring clarity on what actions are permitted or restricted (Labastida and Margoni, 2020). A clear and accessible licence is crucial for enabling not only humans but also automated systems to ascertain the permissible uses of DOs, thus facilitating compliance and responsible reuse.

Atmosphere - All three examined infrastructures adopt a specific policy that includes a licensing scheme for data and/or metadata access. Regarding data, ACTRIS, ICOS and SIOS converge on the use of the CC BY 4.0 licence. However, it's important to note that ACTRIS related documentation[208] specifies the use of this licence depending on the data level and DO: level 0, 1, 2 data, and level 1 and 2 metadata have the CC BY 4.0 as mandatory. For level 3 data the CC BY4.0 is just recommended, nevertheless, the specific licence could be adopted if data contains information provided by third parties. For softwares, the licence to use is the AGPL (GNU Affero General Public License) while for vocabularies the reference licence is the CC0. Similarly, ICOS employs the Creative Common licence (CC0) for metadata and semantic resources (the ontology described in section 5.3.2), and the CC BY 4.0 for data. Furthermore, for softwares and code packages ICOS present other kind of licences: all softwares available on the dedicated Carbon Portal Github[209] present the GNU GPL v3.0 (General Public License). On the other hand, ETC

208 https://www.actris.eu/sites/default/files/inline-files/ACTRIS_ERIC_GA_approved_ACTRIS%20licensing.pdf

209 https://github.com/orgs/ICOS-Carbon-Portal/repositories?type=all

softwares[210] have three types of licences, depending on the code/software: GNU GPL v.2.0, v.3.0; and the BSD 3 Clause License. Finally, SIOS references the CC-BY-NC both for data and metadata.

Marine - Euro-Argo join the CC BY 4.0 licence for their (meta)data. EMSO strongly recommends data owners publish their data under the CC BY 4.0 licence. On the other hand, JERICO and Eurofleets, through the EMODnet portal, use a variety of open licences, including ODC-By, CC0 1.0, CC BY 4.0, CC BY SA 4.0, ODC-ODbL, and PDDL. Moreover, for these last two marine RIs, the Seadatanet portal ensures that more than 90% of their content is licensed under CC BY 4.0, while the remaining content has access restrictions which can be accessed "by negotiation". This aspect remains somewhat ambiguous, but its application likely depends on the data owner's availability to share their research products, going in clear contrast to the FAIR principles and open data sharing. On the other hand, all JERICO products in the CMEMS data portal are published under the Copernicus licence[211] which includes principles similar to those of open data, allowing users to freely access, use, modify, and share the data.

Terrestrial biosphere - LifeWatch Italy is the only RI in this subdomain that has a data policy available on the website[212]. The default data usage licence agreement for all uploaded datasets is a CC BY-NC-SA 4.0 licence. Individual data providers also have the option to negotiate their own data usage policy, including an embargo period, but they must be anyway consistent with the principles of the LifeWatch Italy Data Policy. The licence is always displayed on the metadata file. AnaEE data portal, IBISBA Hub, and eLTER metadata catalogues (DIP and DEIMS) present several degrees of licensing and permission to reuse, from CC0 to no permission to use unless the owner grants a licence. IBISBA Hub, which collects several kind of DOs, allows, among others, the licencing with Apache Software Licence 2.0[213], which is a free software licence which gives users permission to reuse the code. At the moment, one workflow is actually licence with the Apache Licence[214]. In the AnaEE Data Portal we found some data that are licenced with the Open Data Commons Attribution Licence (ODC-by), which is a licence agreement intended to allow users to freely share, modify, and use dataset and databases, and is subject only to the original author attribution. eLTER data reuse is defined in the metadata field “Legal constraints”. This field can be compiled by using the two concepts of the “Conditions Applying To Access and Use”[215] which comes from the INSPIRE metadata code list register[216]. Along with these concepts, the owner of the resource is allowed to add additional restrictions, for instance “Formal acknowledgement of the dataset providers”, or “Co-authorship on publications resulting from use of the dataset”. Most of the time, the licences are displayed and linked in the metadata. Metadata is always freely accessible, without any restriction. However, only the LifeWatch Italy data portal clearly states that the metadata are provided under a CC0 licence. For the other resources, the metadata licence is not clearly displayed.

[210] https://github.com/icos-etc/
[211] https://marine.copernicus.eu/user-corner/service-commitments-and-licence
[212] https://www.lifewatchitaly.eu/data-policy-ita/
[213] https://opensource.org/license/apache-2-0/
[214] https://hub.ibisba.eu/workflows/20 (last access 19/01/2024)
[215] http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse
[216] http://inspire.ec.europa.eu/metadata-codelist

Geosphere landsurface - Regarding SMINO, no data licence information is available on the Real Time Seismology of NE Italy website. Conversely, the Bulletin of the Seismometric Network of North Eastern Italy is published under CC BY 4.0. OASIS and FReDNet data are distributed under CC BY SA 2.0 . Lastly, EIDA declares that all the data available through the different services are distributed under CC BY 4.0. When not indicated, no specific information was found for the metadata licence. On the other hand, by submitting data and metadata to the PANGAEA platform the content will be published under the CC0 licence for metadata and CC BY for data, ensuring that authors of the data are cited.

Table 4. State of the art for Reusability FAIR principles practices adopted by the most mature Research Infrastructures within the four environmental subdomains (atmosphere, marine, terrestrial biosphere and geosphere landsurface).

REUSABILITY			
Research Infrastructure		**Metadata licence**	**DOs licence**
ATMOSPHERE	ACTRiS	Metadata (level 1 and 2): CC BY 4.0	Data: - level 0, 1, 2: CC BY 4.0; - level 3: CC BY 4.0 recommended; Softwares: AGPL; Vocabulary CC0
	ICOS	CC0	Data: CC BY 4.0 ICOS Ontology: CC0 Codes and softwares: GNU GPL v.2.0, v.3.0; BSD 3 Clause Licence
	SIOS	CC-BY-NC	Data: CC-BY-NC
MARINE	emso ERIC	CC BY 4.0	Data: use of CC-BY-4.0 is strongly recommended Software: different licences[217]
	Euro Argo	CC BY 4.0	Data: CC BY 4.0 Software: different licences (e.g., argpy license is EUPL 1.2) Argo vocabularies: CC BY 4.0
	Eurofleets+	CC BY 4.0	EMODnet Data: ODC-By, CC0 1.0, CC BY 4.0, CC BY SA 4.0, ODC-ODbL, PDDL SeaDataNet Data: more than 90% under CC BY 4.0. Other former access restrictions are set to "By negotiation"

[217] https://github.com/spdx/license-list-data/blob/main/licenses.md

	JERICO RESEARCH INFRASTRUCTURE	CC BY 4.0 Copernicus Licence	EMODnet Data: ODC-By, CC0 1.0, CC BY 4.0, CC BY SA 4.0, ODC-ODbL, PDDL SeaDataNet Data: more than 90% under CC BY 4.0. Other former access restrictions are set to "By negotiation" CMEMS Data: Copernicus Licence
TERRESTRIAL BIOSPHERE	AnaEE ERIC Analysis and Experimentation on Ecosystems	Freely accessible, licence is not displayed.	No licence provided, CC BY, CC BY SA, ODC-BY 1.0
	eLTER RI	CC BY NC	Several degrees of permission to reuse, from "no conditions to access and use" to no permission to use unless the owner grants a licence. Services (DIP): N/A
	IBISBA	Freely accessible, licence is not displayed.	Datafile: No licence - no permission to use unless the owner grants a licence, CC0, CC BY 4.0, CC0 1.0 SOPs: No licence - no permission to use unless the owner grants a licence, CC BY 4.0 Workflows: CC BY 4.0, Apache Software Licence 2.0 Documents: CC BY 4.0, No licence - no permission to use unless the owner grants a licence File templates: CC BY 4.0
	LifeWatch ERIC	CC0	CC BY 4.0 (with possibility of negotiation).
GEOSPHERE LANDSURFACE	Sistema di monitoraggio terrestre dell'Italia Nord Orientale (SMINO)	N/A	Real Time Seismology of NE Italy: no information Bulletin of the Seismometric Network of North Eastern Italy: CC BY 4.0 Archive System of Instrumental Seismology (OASIS): CC BY SA 2.0 Friuli Regional Deformation Network (FReDNet): CC BY SA European Integrated Data Archive: CC BY 4.0
	ECORD European Consortium for Ocean Research Drilling	CC0	CC BY

6 Exploring FAIR-enabling best practices in Italy: lessons learned and future prospects

Through a detailed analysis of FAIR-enabling practices adopted by the individual infrastructure involved in ITINERIS, we aim to provide an overview of the complex and heterogeneous landscape of the adopted FAIR practices at the national level. In this section, we present the main outcomes emerging from the analysis of each of the FAIR principles. Depending on the stage of development of each RI, we found that the strategies and practices adopted for the FAIRness of DOs are very heterogeneous.

6.1 *Findability*

The most used persistent identifier for (meta)data is the DOI. This is not surprising as Peterseil et al. (2023) already highlighted that 90% of the RIs participating in the FIP exercise of ENVRI-FAIR declared the use of the DOI to uniquely identify and resolve digital objects and the associated metadata. The use of DOIs has a number of advantages that would be beneficial to consider within the ITINERIS context as well. It 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 DOs, 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. While the selection of DOI is the most common strategy adopted, it can nonetheless be expensive or challenging to maintain in the long term without an adequate financial management plan. On the other hand, long-term maintenance of in-house produced PIDs, such as URIs based on UUIDs, and others, is entirely dependent on the long-term persistence of the infrastructure that generates them. Nevertheless, the discussion if UUID can be properly adopted as PID is still opened, as reported by the FDO Forum[218]. Additionally, some RIs use a number of PIDs for non-data entities (e.g., ORCID and ROR) but there is no common adoption across the RIs. In general, the financial aspect and the long-term maintenance of the PIDs are particularly relevant for RIs, thus deserving careful consideration in the development of the long-term strategy for the ITINERIS HUB.

The various RIs adopt distinct metadata schemas because the schema is selected according to the specific type of DOs and the standard used by each community of reference. However, a certain uniformity of choice has been found within each subdomain, such as in the marine or atmosphere sub-domains, where the trend is towards adopting NetCDF CF. ISO19115/19139 for geo-spatial data are also common metadata schemas used across the different subdomains. In addition, the marine components of SIOS and ICOS participating in the BlueCloud project are developing a metadata schema aligned with the project's guidelines. Furthermore, within the ENVRI-FAIR project, agreement was reached on adopting CERIF as the underlying schema and ingesting metadata records via EPOS-DCAT-AP[219]. While there might be some convergence among schemas, a concerted effort to map and align the varying terminologies - or at least a core set of terms - across these schemas is needed, ensuring their interoperability in terms of structure and semantics (Ulrich et al., 2022).

[218] https://zenodo.org/records/7824714
[219] ENVRI-FAIR D5.6 Synthesis and future strategy https://zenodo.org/records/10363036

Most of the (meta)data management systems analysed are based on open-source software, e.g., CKAN, GeoNetwork, and FAIRDOM-SEEK. The open-source foundation of these platforms ensures transparency in their functionalities and development, making them reliable choices for managing and sharing research data and other products. Anyway, some RIs do not rely on a single management system for data retrieving. It usually occurs when an RI integrates various pre-existing networks (see for example, SMINO).

Besides this, it is also important to highlight the diversity in terms of the management level of the different systems, which depends on the governance of the considered research infrastructure. We found that, for example, the gathering of data and related services can be either managed by the national nodes (e.g. LifeWatch Italy) or have a centralised/federated approach (e.g., ACTRIS, ICOS, Euro-Argo, and EMSO). Specifically, ICOS employs a centralised approach in which all the data and their management fall under European jurisdiction through the creation of a *Thematic Centre*. ACTRIS instead has a system of federated thematic data centres, each one managing data for all the European sites related to that ACTRIS component. One of these thematic nodes (ARES) is led by the Italian DC node at CNR-IMAA, managing data for all European stations. On the other hand, other infrastructures, such as LifeWatch manage data at the national node level (LifeWatch Italy) and expose their metadata at the European level. Furthermore, RIs often serve specific scientific subdomains or communities with unique data requirements. Building dedicated portals enables the customisation of features, functionalities, and interfaces that align precisely with the needs and workflows of their user communities. This specialised approach ensures that researchers can access and interact with data in ways that are highly relevant and valuable to their particular fields of study.

Euro-Argo stands out for its widespread data distribution, publishing on its own portal, European portals (EMODnet and SeaDataNet), repository (SEANOE), and global data portals (GDAC). The same approach is followed by eLTER, using in addition to its own data portal and metadata catalogue, a B2SHARE community and a ZENODO collection to store and centralise different types of DO (datasets, softwares, reports, etc) produced by LTER-Italy. This strategy ensures extensive findability and long-term accessibility to their products. Within the geosphere landsurface, ECORD does not have its own specific data repository but relies on PANGAEA, which has a 30-year history as an open-access library for publishing data from the earth, environmental and biodiversity sciences.

This diverse landscape, with varied operational models, underscores the need for streamlined solutions to promote efficient data utilisation and findability. All these aspects make clear the importance and usefulness of a single access point for DOs in the environmental domain at a national level, which represents the main goal of the ITINERIS project with the establishment of the Italian Hub of Research Infrastructures. This implementation goes in the direction of avoiding and overcoming the fragmentation of information. In fact, ITINERIS will harmonise the way to access the existing research outputs that are made available by the distributed national RIs, which will be reinforced by the project activities. ITINERIS HUB will allow access to DOs, with a proper access management system and a complete catalogue of resources.

6.2 *Accessibility*

For the access to the DOs, a variety of modalities has been delineated during the analysis of the RIs. Speaking about communication and transmission protocols, the analysed RIs

mainly use HTTPS and its derivatives or implementations (i.e., DAP protocols), while only some of them use HTTP (not secured). Few RIs also offer FTP for (meta)data access. Moreover, to facilitate accessibility several web services (such as SPARQL service) and tools (web-tools or software packages) have been made available through web portals.
Regarding machine-to-machine accessibility service, traditional methods, like downloading collections of files via hyperlinks on web pages or via FTP links, are unable to adequately meet the demands of globally distributed scientific and operational requirements. Instead, machine-to-machine APIs provide more sophisticated capabilities such as querying, subsetting, and offer a host of advantages over simply bulk downloading files. For this reason some RIs developed specific APIs for data and metadata harvesting. Among servers and registries, RESTful seems to be the mainly adopted approach while in some cases SOAP is also utilised. Data exchange within a distributed node system necessitates APIs specifically engineered. Indeed, a wide range of mature options exists to facilitate web-based APIs for data and metadata access. A notable example of this is the OPeNDAP[220] protocol. As Hankin et al. (2010) described, OPeNDAP has proven extremely useful for data serving. Over time, it has become a standard in the Earth science community, with several software packages implementing this protocol. Data servers like THREDDS[221], Hyrax, and ERDDAP[222] embody such examples. These servers can ingest data files in a variety of formats and publish data to the web using the OPeNDAP protocol. In doing so, they offer a more scalable and efficient solution for data exchange in globally distributed systems. Remarkably, both eLTER and ECORD use the OAI-PMH as the protocol to allow the harvesting of metadata.
Regarding the authorization and authentication procedures, most of the RIs offer full open access to their data and metadata products. A minority of RIs required a mandatory registration step (e.g. ACTRIS, Eurofleets, JERICO) to download the data. The upload of data and the use of eventual additional analytical tools are always allowed only after a registration step. This addresses the requirement of RIs to manage user access and privileges effectively, distinguishing between novice and proficient users.
Preserving the accessibility of DOs for future generations is a key objective in advancing scientific research in the years to come. For this reason, ensuring the sustainability and longevity of metadata is crucial to sustain the enduring accessibility of DOs over the long term (Li and Sugimoto, 2017; Jeffery, 2020). On the contrary, metadata preservation remains unsatisfactory without a cohesive and clearly articulated financial and long-term plan. In our analysis, 8 out of the 13 RIs analysed give information about the strategies adopted for the management and preservation of metadata. In general, the institutional commitment to metadata preservation is growing. This is supported by the fact that all the RIs are long-term sustained by the European Commission[223] and/or National funding resources[224]. In most cases, their facilities are physically hosted by public funded institutes, ensuring their continued operation in the long term. ITINERIS project is built upon a consortium of 22 infrastructures, all located within high-profile research and academic centres. This represents a solid foundation in terms of long-term (meta)data preservation.

[220] https://www.opendap.org/
[221] https://www.unidata.ucar.edu/software/tds/
[222] https://upwell.pfeg.noaa.gov/erddap/information.html
[223] European Commission, Directorate-General for Research and Innovation, *Horizon Europe – Strategic plan 2021-2024*, Publications Office of the European Union, 2021, https://data.europa.eu/doi/10.2777/083753
[224] https://www.mur.gov.it/sites/default/files/2021-10/Decreto%20Ministeriale%20n.1082%20del%2010-09-2021%20-%20PNIR%202021%20-%202027.pdf

Nevertheless, it will be necessary to delineate the terms and methods for long-term maintenance when developing the national HUB. This includes addressing storage, disaster recovery, and economic sustainability, among other critical aspects.

6.3 *Interoperability*

As regards the adoption of metadata format, a general consensus has been found across the analysed RIs, with JSON, XML and RDF being the most widely adopted standards in the semantic Web (Batista et al., 2022; Decker et al., 2000). These formats provide structured and interoperable ways to manage metadata, ensuring consistency and ease of integration. As for data, the most adopted solutions are NetCDF, CSV and ASCII. However, compared to metadata, a wider range of formats is employed when it comes to data. This variability arises from the nature of the data itself, with different types of data requiring specific formats to accurately capture their complexity. Considerable progress has been made in the field of interoperability in the marine domain. However, it appears that true international interoperability has only been realised for a minor fraction of the diverse types of data being gathered in the world's oceans. There has been great progress regarding numerous physical and meteorological variables that are crucial for the operational ocean community. Additionally, significant advancements have been made in validated archives of marine data via metadata standards and semantics. On the other hand, the biogeochemical and biological data communities are still striving for improvement and need increased and sustained effort to meet observing systems' needs. There are multiple reasons for the slower progress of biogeochemical and biological communities when it comes to data interoperability. These communities are largely operating more in a mode where they lean on traditional research activities that inherently rely on an individual's specialised skills and knowledge, with a low capacity for fast and interoperable data exchange and with a large and complex set of variables being measured.

In the past decade, there has been significant growth and development in the field of environmental sciences regarding ontologies, thesauri, and controlled vocabularies in general as demonstrated by the extensive research conducted for the deliverable 2.11, activity 2.4 within the ITINERIS project. Known as *semantic artefacts*, these resources have become crucial tools for formally representing and semantically organising aspects of the real world. They enable the application of computational techniques that extract useful information from available data and metadata (Diepenbroek et al., 2017). Most of the RIs use semantic artefacts included mainly in three semantic artefact repositories, i.e., the NVS, EcoPortal, and EMBL-EBI Ontology Lookup Service. Some RIs use subdomain specific semantic artefacts such as the Climate and Forecast standard names (atmosphere, marine), the SeaDataNet Common Vocabularies (marine), the Traits thesauri (biosphere). In addition, ACTRIS, ICOS, LifeWatch, eLTER, and AnaEE manage semantic artefacts created in-house. LifeWatch is also the only RI within ITINERIS that manages a repository of semantic artefacts, namely EcoPortal, that provides the possibility to create, manage, and publish FAIR semantic resources. EcoPortal includes: an editing tool for the creation and management of semantic resources (VocBench; Stellato et al., 2020), a publishing tool for the assignment of resolvable URIs to terms/concepts (ShowVoc[225]), a tool for assessing the FAIRness of semantic artefacts included in the repository (O'FAIRe), and the possibility to assign DOI to the published resources. Moreover EcoPortal, for any semantic

[225] https://showvoc.uniroma2.it/

artefact hosted, offers different widgets (java script code) that can be incorporated in third-party websites.

6.4 *Reusability*

The most common licences adopted by RIs operating in the environmental domain were open Creative Commons licences for DOs and metadata. However, there is considerable variability in the specific licences and their application, with some RIs offering more transparent and consistent licensing policies than others. In particular, the FAIR principles state that both data and associated metadata should be released with a clear and accessible data usage licence (Wilkinson et al., 2016). Among the RIs we analysed, this was not always clear, especially for metadata. It is worth noting that, at least for metadata, being the RIs funded mainly by EC projects, it should be expected that all are licenced under CC0[226] or equivalent public licences.

7 Conclusions and future challenges

The ITINERIS project aims to establish the Italian Hub of Research Infrastructures in the environmental domain facilitating access to both pre-existing and newly generated research products. Within this framework, WP 2 collaborates closely with all other WPs, overseeing targeted initiatives for the organisation and categorization of diverse resources. This document serves as an initial guide, based on the expertise of established RIs, to propose effective strategies and protocols for the FAIRNESS implementation. In fact, the strategies adopted from well-established RIs that have been identified through the analysis can provide valuable examples of FAIR implementation actions for those RIs that are in the emerging and evolving stages. Finally, it will provide essential insights for the design and development of the ITINERIS HUB, ensuring the management of research outputs adheres to the FAIR principles.

8 Acknowledgement

We would like to thank the FAIR representatives from different RIs who completed the survey and made themselves available for one-to-one meetings. We are also thankful to the reviewers of the deliverable for their valuable contributions and insights.

[226] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf

9 REFERENCES

Admin, I. B. I. S. B. A., & Goble, C. (2022). D8.2 Preliminary architecture for the IBISBA digital infrastructure. IBISBA. https://doi.org/10.34701/IBISBA.1.DOCUMENT.39.1

Andrabi, M. (2023). D8.5 FAIR research infrastructure services for supporting IBISBA-ERIC. IBISBAHub. https://doi.org/10.34701/IBISBA.1.DOCUMENT.66.1

Batista, D., Gonzalez-Beltran, A., Sansone, SA. et al. Machine actionable metadata models. Sci Data 9, 592 (2022). https://doi.org/10.1038/s41597-022-01707-6

Babady, N. E., Burckhardt, R. M., Krammer, F., Moore, P. L., & Enquist, L. W. (2022). Building a resilient scientific network for COVID-19 and beyond. Mbio, 13(5), e02223-22. https://doi.org/10.1128/mbio.02223-22

Benson, T., Grieve, G. (2016). Why Interoperability Is Hard. In: Principles of Health Interoperability. Health Information Technology Standards. Springer, Cham. https://doi.org/10.1007/978-3-319-30370-3_2

Boeckhout, M., Zielhuis, G.A. & Bredenoord, A.L. The FAIR guiding principles for data stewardship: fair enough?. Eur J Hum Genet 26, 931–936 (2018). https://doi.org/10.1038/s41431-018-0160-0

Bragato, P. L., P. Comelli, A. Saraò, D. Zuliani, L. Moratto, V. Poggi, G. Rossi, C. Scaini, M. Sugan, C. Barnaba, et al. (2021). The OGS– Northeastern Italy Seismic and Deformation Network: Current Status and Outlook, Seismol. Res. Lett., 92, 1704 -1716. https://doi.org/10.1785/0220200372

Conze, R., Wallrabe-Adams, H. J., Graham, C., & Krysiak, F. (2007). Joint data management on ICDP projects and IODP Mission Specific Platform Expeditions. Scientific drilling, 4, 32-34.

Corcho, O., Eriksson, M., Kurowski, K., Choirat, C.,Van de Sanden, M., & Coppens, F. (2021).EOSC Interoperability Framework. Report from the EOSC Executive Board Working Group (WG) FAIR and Architecture. Brussels. https://doi.org/10.2777/620649

David, R., Baumann, K., Le Franc, Y., Magagna, B., Vogt, L., Widmann, H., Jouneau, T., Koivula, H., Madon, B., ̊Akerstr ̈om, W.N., et al., 2023a. Converging on a semantic interoperability framework for the european data space for science, research and innovation (eosc), in: 2nd Workshop on Ontologies for FAIR and FAIR Ontologies (Onto4FAIR), 9th Joint Ontology Workshops (JOWO 2023).

de Castro, Pablo, Herb, Ulrich, Rothfritz, Laura, & Schöpfel, Joachim. (2023). Building the plane as we fly it: the promise of Persistent Identifiers. Zenodo. https://doi.org/10.5281/zenodo.7258286

Decker S. Melnik; F. van Harmelen; D. Fensel; M. Klein; J. Broekstra; M. Erdmann; I. Horrocks, The Semantic Web: the roles of XML and RDF, in IEEE Internet Computing, vol. 4, no. 5, pp. 63-73, Sept.-Oct. 2000, doi: 10.1109/4236.877487.

de Natale, F., Bossi, M., Chanzy, A., de Nart, D., de Pascalis, F., Ferrighi, L., Fiore, N., Gasner, P., García Rodríguez, J. M., González-Aranda, J. M., Islam, S., Ignatiuk, D., L'Abate, G., Parisse, B., Peterseil, J., Pichot, C., Rosati, I., Sáenz-Albanés, A. J., Sánchez Cano, F. M., … Basset, A. (2021). ENVRI-FAIR D11.2: Report on FAIRness implementation activities in the Biodiversity and Ecosystem subdomain (Version 1). Zenodo. https://doi.org/10.5281/zenodo.4682826

Devaraju, A., Mokrane, M., Cepinskas, L., Huber, R., Herterich, P., de Vries, J., Akerman, V., L'Hours, H., Davidson, J and Diepenbroek, M. 2021. From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. Data Science Journal, 20: 4, pp. 1–14. DOI: https://doi.org/10.5334/dsj-2021-004

Diepenbroek, M., Schindler, U., Huber, R., Pesant, S., Stocker, M., Felden, J., ... & Weinrebe, M. (2017). Terminology supported archiving and publication of environmental science data in PANGAEA. Journal of biotechnology, 261, 177-186. https://doi.org/10.1016/j.jbiotec.2017.07.016

Dutta, B., Nandini, D., & Shahi, G. K. (2015). MOD: Metadata for Ontology Description and Publication. International Conference on Dublin Core and Metadata Applications, 1–9. Retrieved from https://dcpapers.dublincore.org/pubs/article/view/3758

European Commission, Directorate-General for Research and Innovation, (2022). Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC), Publications Office of the European Union. https://data.europa.eu/doi/10.2777/935288

Felden, J., Möller, L., Schindler, U., Huber, R., Schumacher, S., Koppe, R., Diepenbroek M., Glöckner, F. (2023). PANGAEA - Data Publisher for Earth & Environmental Science. Sci Data 10, 347. https://doi.org/10.1038/s41597-023-02269-x

Gibbs, R. A. The Human Genome Project Changed Everything. Nat. Rev. Genet. 2020, 21 (10), 575–576. https://doi.org/10.1038/s41576-020-0275-3.

Gries, C., San Gil, I., Vanderbilt , C., Garrit, H., 2010. Drupal Developments in the LTER Network. https://lternet.edu/wp-content/uploads/2017/12/2010-spring-lterdatabits.pdf.

Hankin, S. C., Blower, J. D., Carval, T., Casey, K. S., Donlon, C., Lauret, O., et al. (2010). "NetCDF-CF-OPeNDAP: Standards for Ocean Data Interoperability and Object Lessons for Community Data Standards Processes," in Proceedings of OceanObs'09, (Paris: European Space Agency), 450–458

Hellström, M., Heughebaert, A., Kotarski, R., Manghi, P., Matthews, B., Ritz, R., Conrad, A. S., Weigel, T., Wittenburg, P., & Valle, M. (2020). *A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC)*. https://doi. org/10.2777/926037

Hill, T. Turning FAIR into Reality: Review. Learn. Publ. 2019, 32 (3), 283–286. https://doi.org/10.1002/leap.1234.

IEEE. IEEE standard computer dictionary: a compilation of IEEE standard computer glossaries. New York: Institute of Electrical and Electronics Engineers; 1990

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), 10-29.

Jeffery, K. (2020). Data Curation and Preservation. In: Zhao, Z., Hellström, M. (eds) Towards Interoperable Research Infrastructures for Environmental and Earth Sciences. Lecture Notes in Computer Science, vol 12003. Springer, Cham. https://doi.org/10.1007/978-3-030-52829-4_7

Jetz, W., McGeoch, M.A., Guralnick, R. et al. Essential biodiversity variables for mapping and monitoring species populations. Nat Ecol Evol 3, 539–551 (2019). https://doi.org/10.1038/s41559-019-0826-1

Kahn, R., and Wilensky, R. (1995-05-13). *A Framework for Distributed Digital Object Services*. URL: http://www.cnri.reston.va.us/k-w.html

Kahn, R., and Wilensky, R. (2006). A Framework for Distributed Digital Object Services. *International Journal on Digital Libraries* 6.2, pp. 115–123. ISSN: 1432-5012. https://doi.org/10.1007/s00799-005-0128-x. URL: https://www.doi.org/topics/2006 05 02 Kahn Framework.pdf (visited on 2022-05-05).

Kirkpatrick, C. R.; Coakley, K.; Dutra, I. Engaging with Researchers and Raising Awareness of FAIR and Open Science through the FAIR + Implementation Survey Tool (FAIRIST). 2023, 1–11. https://doi.org/10.5334/dsj-2023-032.

Klump, J., & Huber, R. (2017). 20 Years of persistent identifiers–Which systems are here to stay?. Data Science Journal, 16, 9-9. https://datascience.codata.org/articles/10.5334/dsj-2017-009

Landi, A., Thompson M., Giannuzzi V., Bonifazi F., Labastida I., Bonino da Silva Santos L.O. & Roos M. The "A" of FAIR – As open as possible, as closed as necessary. Data Intelligence 2(2020), 47–55. doi: 10.1162/dint_a_00027

Labastida, Ignasi, and Thomas Margoni. "Licensing FAIR data for reuse." Data Intelligence 2.1-2 (2020): 199-207.

Le Franc, Y., Bonino, L., Koivula, H., Parland-von Essen, J., Pergl, R., 2022. D2.8 fair semantics recommendations third iteration. doi:10.5281/zenodo.6675295

Li, C., & Sugimoto, S. (2017). Provenance description of metadata vocabularies for the long-term maintenance of metadata. Journal of data and information Science, 2(2), 41-55. https://doi.org/10.1515/jdis-2017-0007

Maxwell, L.; Shreedhar, P.; Dauga, D.; Communication, B.; Mcquilton, P.; World, R. T.; Organization, H.; Denisiuk, A.; Molnar-Gabor, F.; Saxena, A.; Advisor, B.; Sansone, S.-A. FAIR, Ethical, and Coordinated Data Sharing for COVID-19 Response: A Review of COVID-19 Data Sharing Platforms and Registries. Res. Sq. 2023, 5 (10), e712–e736. https://doi.org/10.1016/S2589-7500(23)00129-2.

Melton, J., Buxton, S., 2006. Metadata – an overview. In: Melton, J., Buxton, S. (Eds.), Queyring XML. Morgan Kaufmann, Burlington, USA, pp. 67–84. https://doi.org/ 10.1016/B978-155860711-8/50005-8.

Minić, Vladan & Crnojević, Vladimir & Bodroski, Zarko & Skrbic, Srdjan & Watkins, John & Wohner, Christoph & Peterseil, J. & Oggioni, Alessandro & Lanucara, Simone & Kunkel, Ralf & Sorg, Jürgen. (2018). D8.3 eLTER Data Integration Platform. 10.13140/RG.2.2.18097.84324.

O'Donohue, M (2023).Data manangement plan-IBISBA1.0 (Version3). INRAE.2022. Hal-04136517

Obaton, D., Hassoun, A. E. R., Gomez, B. P., Novellino, A., Carval, T., Mader, J., ... & Coppola, L. (2022, October). Recommendations on data harmonization for ocean observation networks. In 2022 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea) (pp. 01-03). IEEE. DOI: 10.1109/MetroSea55331.2022.9950942

Peterseil, Johannes, Offenthaler, Ivo, Wohner, Christoph, Magagna, Barbara, Schultes, Erik, Lund Myhre, Cathrine, Jeffery, Keith, Bailo, Daniele, Dobler, Delphine, Portier, Marc, Dema, Claudio, Vaira, Lucia, & Rosati, Ilaria. (2023). ENVRI-FAIR D5.6: Synthesis and future strategy (Version 1). Zenodo. https://doi.org/10.5281/zenodo.8118948

Peters-Von Gehlen, K., Ho ¨ck, H., Fast, A., Heydebreck, D., Lammert, A., Thiemann, H., 2022. Recommendations for discipline-specific FAIRness evaluation derived from applying an ensemble of evaluation tools. Data Sci. J. 21 (1), 1–21. https://doi.org/10.5281/zenodo.5879457.

Priolo E., G. Laurenzano, C. Barnaba, P. Bernardi, L. Moratto e A. Spinelli (2015). OASIS: The OGS Archive System of Instrumental Seismology. Seismol. Res. Lett., 86(3), 978-984, doi:10.1785/0220140175.

Quay, A.N., Fiske, P.S., Mauter, M.S., 2022. Recommendations for advancing FAIR and open data standards in the water treatment community. ACS ES&T Eng. 2 (3), 337–346. https://doi.org/10.1021/acsestengg.1c00245

van Reisen, M.; Stokmans, M.; Basajja, M.; Ong'ayo, A. O.; Kirkpatrick, C.; Mons, B. Towards the Tipping Point for Fair Implementation. Data Intell. 2020, 2 (1–2), 264–275. https://doi.org/10.1162/dint_a_00049.

Saraò, A., Bragato P. L., Bernardi P. (2009). Real time seismology of the OGS Seismological Research Centre. https://rts.crs.inogs.it

Singh, M. P., & Huhns, M. N. (2005). Service-oriented computing: semantics, processes, agents. John Wiley & Sons.https://doi.org/10.1002/0470091509

Solle, D. Be FAIR to your data. Anal Bioanal Chem 412, 3961–3965 (2020). https://doi.org/10.1007/s00216-020-02526-7

Stellato, A., Fiorelli, M., Turbati, A., Lorenzetti, T., Van Gemert, W., Dechandon, D., ... & Keizer, J. (2020). VocBench 3: A collaborative Semantic Web editor for ontologies, thesauri and lexicons. Semantic Web, 11(5), 855-881. https://doi.org/10.3233/SW-200370

Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., ... & Goble, C. (2022). Packaging research artefacts with RO-Crate. Data Science, 5(2), 97-138.https://doi.org/10.3233/DS-210053

Strollo, A., Cambaz, D., Clinton, J., Danecek, P., Evangelidis, C. P., Marmureanu, A., ... & Triantafyllis, N. (2021). EIDA: The European integrated data archive and service infrastructure within ORFEUS. Seismological Society of America, 92(3), 1788-1795.

Ulrich H, Kock-Schoppenhauer A, Deppenwiese N, Gött R, Kern J, Lablans M, Majeed RW, Stöhr MR, Stausberg J, Varghese J, Dugas M, Ingenerf J Understanding the Nature of Metadata: Systematic Review, J Med Internet Res 2022;24(1):e25440, doi: 10.2196/25440

Weigel, T., Schwardmann, U., Klump, J., Bendoukha, S., & Quick, R. (2020). Making data and workflows findable for machines. Data Intelligence, 2(1-2), 40-46.

Wilkinson, M. D., Dumontier, M., Sansone, S. A., Bonino da Silva Santos, L. O., Prieto, M., Batista, D., ... & Schultes, E. (2019). Evaluating FAIR maturity through a scalable, automated, community-governed framework. Scientific data, 6(1), 174.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Mons, B., 2016. Comment: the FAIR guiding principles for scientific data management and stewardship. Sci. Data 3, 1–9. https://doi.org/10.1038/ sdata.2016.18.

Wittenburg, P.; Strawn, G.; Mons, B.; Bonino, L.; Schultes, E., 2019 Digital Objects as Drivers towards Convergence in Data Infrastructures; EUDAT: Helsinki, Finland. https://b2share.eudat.eu/records/b605d85809ca45679b110719b6c6cb11

Wittenburg, P., Broeder, D., and Buitelaar, P. 2004. Towards Metadata Interoperability. In Proceeedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology, pages 9–16, Barcelona, Spain. Association for Computational Linguistics.

Wolf, M., Logan, J., Mehta, K., Jacobson, D., Cashman, M., Walker, A. M., Eisenhauer. G., Widener P., Cliff, A., 2021. Reusability first: Toward FAIR workflows. In 2021 IEEE International Conference on Cluster Computing (CLUSTER) (pp. 444-455). IEEE.

Wohner, C., Peterseil, J., Poursanidis, D., Kliment, T., Wilson, M., Mirtl, M., Chrysoulakis, N., DEIMS-SDR – a web portal to document research sites and their associated data, Ecol. Inform., 51 (2019), pp. 15-24, 10.1016/j.ecoinf.2019.01.005

Zuliani D., (2004). The FReDNet database, Centro di Ricerche Sismologiche, OGS 2004, Annual Report, pp 100–101

Zuliani, D., Fabris, P., Rossi, G. (2018). FReDNet: Evolution of a Permanent GNSS Receiver System. In: Cefalo, R., Zieliński, J., Barbarella, M. (eds) New Advanced GNSS and 3D Spatial Techniques. Lecture Notes in Geoinformation and Cartography. Springer, Cham. https://doi.org/10.1007/978-3-319-56218-6_10