MISSIONE 4
ISTRUZIONE
RICERCA

ITINERIS

D4.7.3 Documentation of fully developed integrated polar data repository

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

Deliverable number:	D4.7.3
Work package:	WP4 – Atmospheric Domain
Intermediate Objective:	IO4.5
Deliverable type:	X Document, report
	☐ Websites, patent filings, videos, etc.
	☐ Other: please specify .....
Dissemination level:	X Public
	☐ Restricted
Estimated delivery (bimester):	B15
Actual delivery date:	12/05/2025
Author(s) (Partner-OU):	Alice Cavaliere, Mauro Mazzola, Chiara Ripa, Giulio Verazzo, Vito Vitale (CNR-ISP-BO)
Reviewed by:	Alberto Salvati (CNR), Antonio Novellino (ETT)
Note:	

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006 (D.D. n. 130/2022)

Funded by EU - Next Generation EU

Mission 4 “Education and Research” - Component 2: “From research to business” -

Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Table of contents

Index of figures and tables

1. INTRODUCTION

The research activities in the Polar Area are promoted and supported by two government funded research programs: the Italian Antarctic National Research Program (PNRA) and the Italian Arctic Research Program (PRA). The PNRA is inherently multidisciplinary, encompassing a wide range of scientific fields interested in Antarctic research, including marine, atmospheric, astrophysical,

biological, geological, glaciological, and geophysical sciences. In parallel, Italy has also engaged in Arctic research through the Svalbard Integrated Earth Observing System (SIOS). This involvement led to the creation of the Italian Arctic Data Center (IADC), which aims to compile and make available data, particularly from the Arctic Station Dirigibile Italia, to the SIOS Data Management System (SDMS). This system also adheres to the principles of distributed networks and interoperability.

The following section presents the architecture of the IADC, with a focus on the implementation of the FAIR principles. This is followed by a description of the metadata catalog, including an overview of the collected entries. A subsequent section provides a panoramic view of the datasets and their management. Finally, the last sections introduce the tools developed to support data analysis and visualization.

2. IADC ARCHITECTURE

The Italian Arctic Data Center (IADC) is a scientific and technological infrastructure designed to gather, handle, publish and provide access to scientific data and metadata regarding Polar regions. The architecture of the IADC infrastructure is built on the System-of-Systems (SoS) paradigm—an approach that connects a network of independently managed systems through mediation and adaptation services, commonly referred to as brokering services (Nativi, 2013). At the core of this distributed infrastructure there is a single central component known as the Common Node (Nativi, 2015), which is the only element that requires shared governance among participating entities, each of which manages a First Level Node. The Common Node serves as the backbone of the system by establishing connections between all the First Level Nodes through specific adapters that facilitate interoperability. It performs continuous harvesting of metadata published by the First Level Nodes, harmonizing this information according to a unified metadata schema and mediating between heterogeneous formats. Furthermore, the Common Node shares a semantic framework maintaining a common vocabulary, supporting metadata consistency and discoverability. It is also responsible for implementing the data policies defined by the PRA, operating and publishing the infrastructure's main web portal, and acting as the bridge to external systems and international polar research initiatives. In addition, the Common Node ensures compliance with the Quality-of-Service (QoS) requirements established by PRA and promotes a cooperative governance model with the first level nodes. Each of the first level nodes, in turn, is responsible for ensuring data curation and preservation through dedicated data servers such as ERDDAP™, aligning with European standards and adhering to the Horizon Europe 2020 guidelines for Open Access. ERDDAP is recommended by the GOOS Data Management OGC as a FAIR enabler tool. Furthermore, they support the FAIR principles by making data and metadata publicly accessible and interoperable. Additionally, they serve as aggregators for data and information contributed by their associated Second Level Nodes, where applicable.

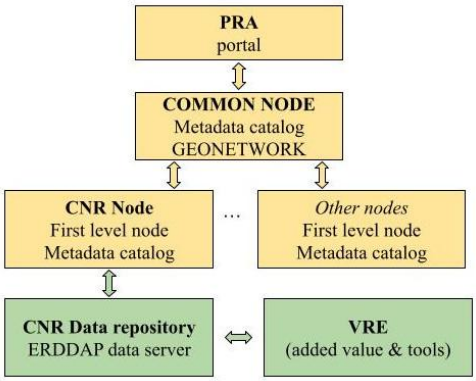


Figure 1: IADC data system architecture

The Italian Arctic Data Center (IADC) utilizes GeoNetwork opens ource, a web-based geographic metadata catalog, to manage and share geospatial data and metadata. GeoNetwork facilitates the organization, discovery, and dissemination of spatial data resources, supporting interoperability and standardization. Key Features of GeoNetwork at IADC:

- Metadata Management: GeoNetwork enables the creation, editing, and storage of metadata records for geospatial datasets, adhering to international standards such as ISO 19115.
- Advanced Search Functionality: Users can perform full-text and faceted searches across metadata records, refining searches based on keywords, resource types, organizations, and more, to efficiently locate relevant geospatial data.
- Interactive Map Viewer: The platform provides an interactive map viewer based on OpenLayers, allowing users to visualize geospatial data and access OGC services like WMS and WMTS.
- Data Sharing and Interoperability: GeoNetwork supports the sharing of geographically referenced thematic information among different organizations, promoting multidisciplinary approaches and enhancing understanding of geographic data benefits.

The IADC's implementation of GeoNetwork ensures that metadata is accessible and discoverable, aligning with the FAIR (Findable, Accessible, Interoperable, Reusable) principles, thereby enhancing the utility and impact of polar region data. F and A the others come with vocabs and standards.

3. FUNCTIONALITY WITH A FAIRNESS-ORIENTED APPROACH

The FAIR Implementation Profile (FIP) is a collection of community-defined choices for implementing the FAIR Principles—Findability, Accessibility, Interoperability, and Reusability. It serves as a structured approach for documenting how an organization or infrastructure applies these principles in practice, highlighting the specific standards, technologies, and policies adopted to ensure that data and metadata are as FAIR as possible. IADC embraces a Fairness-oriented strategy based on a suite of established standards and tools that support the discoverability, accessibility, and long-term usability of scientific data and metadata.

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 (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 Research Infrastructures (RIs), as identified through the survey completed by the Fair Implementation Working Group. Deliverable WP2 D2.7, Nestola et al., State of the Art review of FAIR-enabling best practices.

3.1.Findability (F)

To ensure global uniqueness, persistence, and resolvability (F1), IADC assigns Persistent Uniform Resource Locators (PURLs) to both metadata records and datasets, with ongoing efforts to implement Digital Object Identifiers (DOIs) in collaboration with DataCite. DataCite provides Fabrica, a DOI and metadata management service that enables organizations to register and manage DOIs using the DataCite Metadata Schema. Metadata is structured using the ISO19115 standard (F2) and is indexed by major search engines, including Google and Bing (F4), enhancing visibility. Persistent identifiers (DOI) are connected to metadata descriptions through a metadata-data linking mechanism established between GeoNetwork and ERDDAP™ (F3).

3.2.Accessibility (A)

Metadata records are shared using standardized protocols such as OAI-PMH, GeoNetwork Harvester, GeoPortal REST, and various OGC services (e.g., WMS, WFS, CSW) (A1.1). Datasets are made accessible via OPeNDAP and REST APIs. Access control (A1.2) is enforced through user registration with email, password, and role-based authorization. Dataset-level access is further supported by ORCID integration and CNR SSO authentication. A long-term metadata longevity plan (A2) ensures that metadata remains accessible even if the associated datasets are no longer available.

3.3.Interoperability (I)

Machine-actionable metadata is encoded using XML and validated against XML Schemas (I1), while datasets are provided in formats such as NetCDF/CF, XML, JSON, and other ERDDAP™ -supported standards. Metadata is enriched with structured vocabularies such as Natural Earth, Seavox, GCMD, and GEMET INSPIRE themes (I2), while datasets adopt the Climate and Forecast (CF) conventions for standard names and units. Metadata modelling follows ISO19115, represented in ISO19139 XML, and datasets are structured using the NetCDF/CF model, with ACDD metadata standards applied where appropriate (I3).

3.4.Reusability (R)

Both metadata and datasets default to a CC-BY 4.0 license, while allowing users the flexibility to choose alternative licenses if required (R1.1). Provenance is documented through a dedicated "Lineage" field in ISO19115-compliant metadata and a "History" field in CF/ACDD-compliant dataset metadata (R1.2), detailing all processing steps. This supports transparency, reproducibility, and user trust in the data.

4. IADC METADATA CATALOG

As shown in Fig. 1, the Italian Arctic Data Center (IADC) Geonetwork currently hosts 124 metadata entries, the majority of which are related to activities in Ny-Ålesund. To enhance data discoverability and accessibility in line with the FAIR principles (Findable, Accessible, Interoperable, Reusable), each dataset is assigned one or more ISO topic categories.

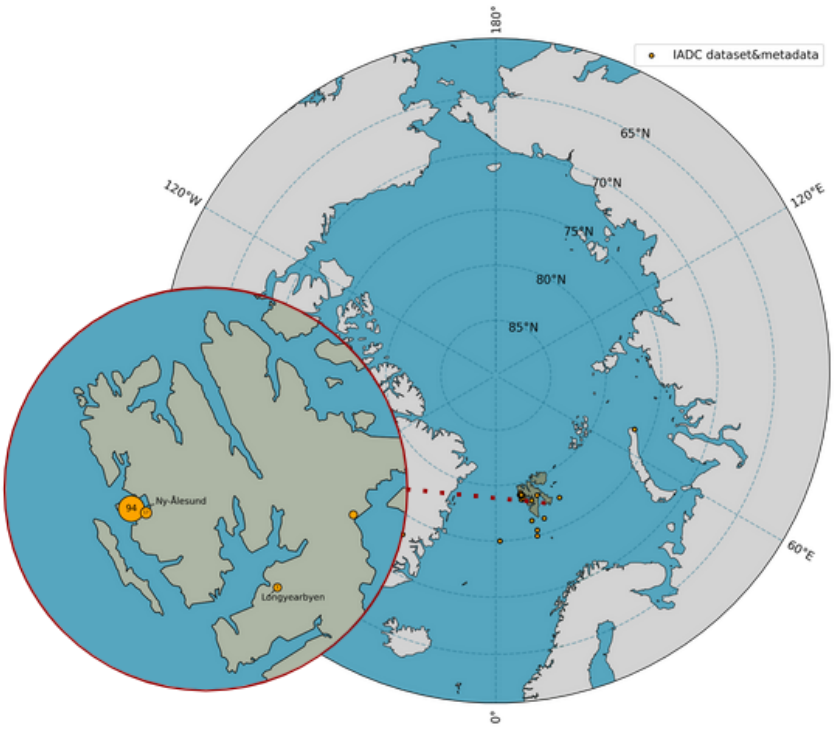


Figure 3: IADC Geonetwork datasets geolocated on a map using the North Polar projection.

Specifically, the datasets are currently organized into the following ISO topics:

- Climatology, Meteorology, Atmosphere: 50 entries
- Oceans: 30 entries
- Environment: 13 entries
- Geoscientific Information: 11 entries
- Inland Waters: 4 entries
- Biota: 3 entries

In addition to assigning these macro-categories, Geonetwork allows assigning keywords to each metadata record from a thesaurus. In a metadata catalog, concepts from a thesaurus can be assigned to a metadata record (as keywords) to associate it with one or more concepts from a specific field of knowledge. Out of the IADC collected metadata, 107 records have keywords assigned from the GMDC Science Keyword thesaurus. The Global Change Master Directory (GCMD) Keywords are a hierarchical set of controlled Earth Science vocabularies that help ensure Earth science data, services, and variables are described in a consistent and comprehensive manner and allow for the precise searching of metadata and the subsequent retrieval of data, services, and variables. Table 1 shows the 20 most frequently used entries of GMDC Science Keywords along with their occurrences.

GMDC Science Keyword	Total occurrences
Aerosol Particle Properties	17
Aerosols	16
Carbonaceous Aerosols	12
Snow/Ice	8
Salinity	8
Atmospheric Chemistry	8
Water Temperature	7
Atmosphere	7
Oxygen	5
Snow Cover	5
Ocean Temperature	5
Snow/Ice Temperature	4
Snow	4
Aerosol Extinction	4
Particle Flux	4
Ionosphere/Magnetosphere Dynamics	4
Water Quality	4
Biogeochemical Processes	4
Carbon	4

Table 1: First 20 Entries of GMDC Science Keywords with their occurrences.

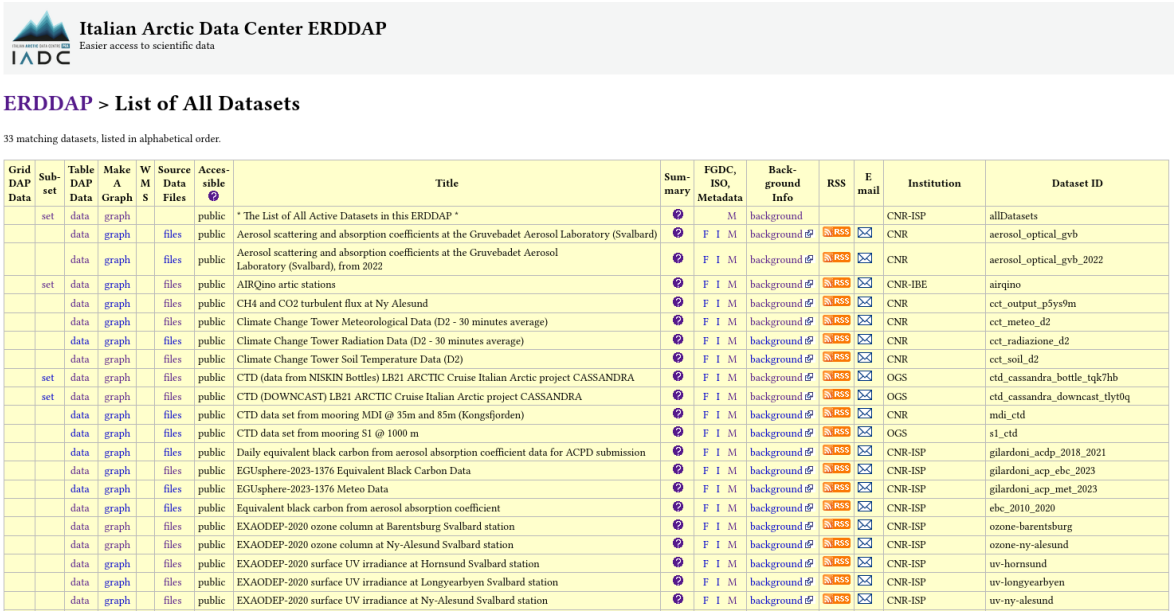
Finally, from the entire collected dataset, the entries associated with the SIOS category have been harvested into the SIOS Data Portal. The SIOS Data Management Service (SDMS) integrates information from SIOS partner data repositories into a unified virtual data centre, the SIOS Data Access Portal. The current focus of the portal is dataset discovery through standardised metadata, as well as the retrieval, visualisation, and transformation of data. A total of 34 IADC dataset labels have

been harvested under the SIOS label category. Among these, the following categories have been identified:

- Oceans: 15 entries,
- Climatology, meteorology, atmosphere: 11 entries,
- Geoscientific information: 7 entries,
- Environment: 2 entries.

5. IADC DATA REPOSITORY

While Common Node and First Level Nodes manage metadata as mentioned in section 3, data oversee second level nodes. Partners of PRA are opting to use ERDDAP™ (Figure 4). ERDDAP™ data server is open-source software made by NOAA written in Java that builds upon the open-source ideals of the OPeNDAP, WCS, SOS and OBIS standards. ERDDAP™ supports both human interaction (e.g. OPeNDAP requests) and machine-to-machine interoperability. ERDDAP™ data server supports several common data file formats (html table, netcdf, csv, txt, mat, json, etc.) and output files are created on-the-fly in any of these formats. ERDDAP™ implements FGDC Web Accessible Folder (WAF) with FGDC-STD-001-1998 and ISO 19115 WAF with ISO 19115-2/19139. ERDDAP™ addresses the challenge of managing different data formats and sharing protocols by offering an easy-to-use, consistent way to request data through the OPeNDAP standard. Thanks to this feature, the IADC ERDDAP is linked to the IADC GeoNetwork: metadata records reference ERDDAP datasets as online resources using the OPeNDAP protocol, enabling seamless access to data directly from the metadata catalog.



Italian Arctic Data Center ERDDAP
Easier access to scientific data

ERDDAP > List of All Datasets

33 matching datasets, listed in alphabetical order.

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	WMS	Source Data Files	Acces-sible	Title	Sum-mary	FGDC, ISO, Metadata	Back-ground Info	RSS	E mail	Institution	Dataset ID
	set	data	graph			public	* The List of All Active Datasets in this ERDDAP *		M	background			CNR-ISP	allDatasets
		data	graph		files	public	Aerosol scattering and absorption coefficients at the Gruvebadet Aerosol Laboratory (Svalbard)		F I M	background	RSS		CNR	aerosol_optical_gvb
		data	graph		files	public	Aerosol scattering and absorption coefficients at the Gruvebadet Aerosol Laboratory (Svalbard), from 2022		F I M	background	RSS		CNR	aerosol_optical_gvb_2022
	set	data	graph		files	public	AIRQino artic stations		F I M	background	RSS		CNR-IBE	airqino
		data	graph		files	public	CH4 and CO2 turbulent flux at Ny Alesund		F I M	background	RSS		CNR	cct_output_p5ys9m
		data	graph		files	public	Climate Change Tower Meteorological Data (D2 - 30 minutes average)		F I M	background	RSS		CNR	cct_meteo_d2
		data	graph		files	public	Climate Change Tower Radiation Data (D2 - 30 minutes average)		F I M	background	RSS		CNR	cct_radiazione_d2
		data	graph		files	public	Climate Change Tower Soil Temperature Data (D2)		F I M	background	RSS		CNR	cct_soil_d2
	set	data	graph		files	public	CTD (data from NISKIN Bottles) LB21 ARCTIC Cruise Italian Arctic project CASSANDRA		F I M	background	RSS		OGS	ctd_cassandra_bottle_tqk7hb
	set	data	graph		files	public	CTD (DOWNCAST) LB21 ARCTIC Cruise Italian Arctic project CASSANDRA		F I M	background	RSS		OGS	ctd_cassandra_downcast_tlyt0q
		data	graph		files	public	CTD data set from mooring MDI @ 35m and 85m (Kongsfjorden)		F I M	background	RSS		CNR	mdi_ctd
		data	graph		files	public	CTD data set from mooring S1 @ 1000 m		F I M	background	RSS		OGS	s1_ctd
		data	graph		files	public	Daily equivalent black carbon from aerosol absorption coefficient data for ACPD submission		F I M	background	RSS		CNR-ISP	gilardoni_acdp_2018_2021
		data	graph		files	public	EGUsphere-2023-1376 Equivalent Black Carbon Data		F I M	background	RSS		CNR-ISP	gilardoni_acp_ebc_2023
		data	graph		files	public	EGUsphere-2023-1376 Meteo Data		F I M	background	RSS		CNR-ISP	gilardoni_acp_met_2023
		data	graph		files	public	Equivalent black carbon from aerosol absorption coefficient		F I M	background	RSS		CNR-ISP	ebc_2010_2020
		data	graph		files	public	EXAODEP-2020 ozone column at Barentsburg Svalbard station		F I M	background	RSS		CNR-ISP	ozone-barentsburg
		data	graph		files	public	EXAODEP-2020 ozone column at Ny-Alesund Svalbard station		F I M	background	RSS		CNR-ISP	ozone-ny-alesund
		data	graph		files	public	EXAODEP-2020 surface UV irradiance at Hornsund Svalbard station		F I M	background	RSS		CNR-ISP	uv-hornsund
		data	graph		files	public	EXAODEP-2020 surface UV irradiance at Longyearbyen Svalbard station		F I M	background	RSS		CNR-ISP	uv-longyearbyen
		data	graph		files	public	EXAODEP-2020 surface UV irradiance at Ny-Alesund Svalbard station		F I M	background	RSS		CNR-ISP	uv-ny-alesund

Figure 4: ERDDAP™ IADC web page

The Italian Arctic Data Center (IADC) ERDDAP™ currently hosts 61 datasets. These datasets are provided by institutions such as the Institute of Polar Sciences (CNR-ISP), the Institute of Atmospheric Sciences and Climate (CNR-ISAC), and the National Institute of Oceanography and Applied Geophysics (OGS). The datasets cover a wide range of Arctic research topics, including aerosol scattering and absorption coefficients, meteorological measurements from the Climate Change Tower, soil temperature data, CTD (Conductivity, Temperature, Depth) data from various moorings, and measurements of equivalent black carbon.

In ERDDAP™, datasets can be categorized in different ways based on the values of various metadata attributes, with two key elements being the common data type and the ERDDAP™ class. The common data type refers to the standardized format in which the data is stored, ensuring consistency across various datasets. This data type defines the structure and organization of the information, such as whether the dataset consists of time series, geospatial data, or other scientific measurements. The ERDDAP™ class instead defines protocols that specify the standards for requesting data. Different protocols are suited to various types of data and are tailored to different client applications. Datasets in ERDDAP fall into two main categories: EDDGrid, for gridded data structures such as satellite and model data, and EDDTable, for tabular data structures such as in-situ instruments and station data. There are many data types available for both EDDGrid and EDDTable. Table 2 presents the dataset classes used in the IADC ERDDAP.

ERDDAP™ Class	ERDDAP™ data types	Total occurrences
EDDTable	EDDTableFromAsciiFiles	28
	EDDTableFromMultidimNcFiles	24
	EDDTableFromNcCFFiles	4
	EDDTableFromDatabase	2
	EDDTableFromHttpGet	1
	EDDTableFromFileNames	1
EDDGrid	EDDGridFromNcFiles	1

Table 2: ERDDAP™ Class protocol distribution

EDDTableFromAsciiFiles aggregates data from tabular ASCII files that are comma-, tab-, semicolon-, or space-separated. *EDDTableFromMultidimNcFiles* and *EDDTableFromNcFiles* aggregates data from NetCDF (v3 or v4) .nc files containing several variables with shared dimensions. *EDDTableFromDatabase* manages tabular data from a single database table. *EDDTableFromHttpGet supports* both data import and export via HTTP GET requests. Lastly, *EDDTableFromFileNames* creates a dataset from information about a group of files in the server's file system, but it does not serve data from within the files themselves. Each of these datasets is associated with a data model. The common data type (*cdm_data_type*, from the ACDD metadata standard) is a global attribute that indicates the Unidata Common Data Model data type for the dataset. ERDDAP complies with the related and more detailed Discrete Sampling Geometries (DSG) chapter of the CF 1.6 metadata conventions. The same data model can be shared across multiple dataset classes. As described in Table 3, the common data types used in the IADC ERDDAP are predominantly *Time Series*, followed by *Other*, *Time Series Profile*, and finally *Point* and *Grid*.

Common Data Model Type	Total occurrences
TimeSeries	53
Other	4
TimeSeriesProfile	2

Point	1
Grid	1

Table 3: ERDDAP™ Common data model type distribution.

TimeSeries refers to a sequence of measurements (e.g., air temperature) taken at a single, fixed location, defined by latitude, longitude, and depth (or altitude). It can be thought of as a station that collects data over time. The dataset must include the global attribute *cdm_timeseries_variables*, where the value is a list of the variables that provide information about each station, such as latitude, longitude, and a unique identifier. *TimeSeriesProfile* represents a sequence of profiles taken at a fixed latitude and longitude. Each profile consists of measurements collected at multiple depths or altitudes. *Point* datasets consist of individual measurements taken at unrelated times and locations, while *Grid* datasets are designed to handle gridded data. *Other* datasets have no specific requirements; for example, they may not include latitude, longitude, or time variables.

5.1.ERDDAP™ Content management system

One of the critical aspects of ERDDAP™ is the creation of dataset XML schemas, which contain essential metadata and links to dataset files. However, this process can be challenging due to its technical complexity and the requirement for specialized skills. To address the challenge of managing dataset XML schemas, an ERDDAP™ Content Management System (CMS) was implemented, as detailed in the following section. ERDDAP™ (Environmental Research Division's Data Access Program) Content Management System (CMS) is a versatile tool developed by CNR and designed to efficiently manage environmental data access. It is designed as a Docker container, ensuring easy deployment and scalability. The system supports seamless login through ORCID and CNR Single Sign-On (SSO), providing a secure authentication process for a wide range of users.

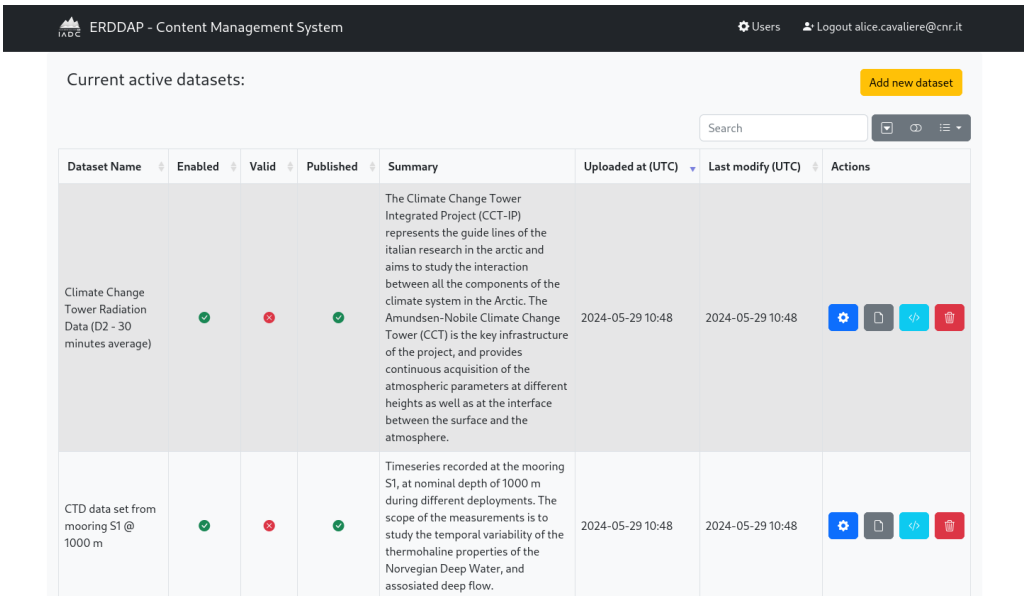


Figure 5: ERDDAP™ CMS linked to ERDDAP™ IADC datasets.

The primary functionality of the ERDDAP™ CMS is to simplify the creation of XML schemas that are compliant with the Geonetwork standard, directly generating metadata in the required format. By providing a user-friendly interface and automating certain tasks, the CMS enhances the efficiency

and effectiveness of managing dataset metadata within ERDDAP™ empowering members of the polar research community to independently contribute polar datasets to the system. This functionality is developed to enhance data accessibility within the system by leveraging common file formats compatible with the OPeNDAP standard. Initially centred on managing tabular data, the implementation conforms to the NetCDF Climate and Forecast (CF) Metadata Conventions Tabular. Additionally, Quality Assurance (QA) metadata is incorporated into the system to ensure the accuracy and reliability of the datasets, further enhancing their usability and integrity. The system's architecture is grounded in an ERDDAP™ tabular dataset type capable of ingesting data from various sources, including local repositories like NetCDF files and ASCII files, as well as virtual resources such as databases and HTTP sources. Each dataset, whether continuous or discontinuous, employed a CF discrete sampling geometry, defined by the relationships among its spatiotemporal coordinates, referred to as its feature type. Notably, the implementation included time series and time series profile datasets, with the latter being more complex due to the presence of two element dimensions (profile and depth). Various analyses were conducted to develop effective methods for handling and integrating trajectory-based and grid datasets into the system. This effort formed part of a broader study on new dataset typologies beyond tabular data, aimed at enhancing the system's capacity to accommodate diverse data formats commonly used in research environments.

6. DATA INTEGRATION AND ADDED VALUE

To enhance dataset usability and accessibility, tools have been developed to streamline data entry and visualization processes. These tools are designed to simplify dataset submission and enhance data visualization and quality assessment. To simplify dataset submission, a content management system has been implemented that enables users - regardless of their experience level- to effortlessly create dataset entries for ERDDAP™. This process includes associating comprehensive metadata with each dataset and assigning a Digital Object Identifier (DOI), ensuring persistent and unique referencing. DOIs act as permanent links to data products, promoting reliable citation and access. To enhance data visualization, a dedicated dashboard offers graphical tools that assist users in visualizing datasets, conducting preliminary data quality checks, and performing more in-depth quantitative analyses. These tools make data exploration intuitive and efficient, enabling users to derive meaningful insights from complex datasets.

6.1. IADC data Dashboard

The goal was to create an intuitive and user-friendly dashboard for exploring datasets within the IADC data centers. This tool is seamlessly integrated with each ERDDAP™ instance, offering direct access to all variables within every dataset (Figure 6). Users can easily define temporal queries by selecting their preferred data range. Built on Streamlit, an open-source Python library, the interactive dashboard enables users to not only explore but also visualize the data in various formats.

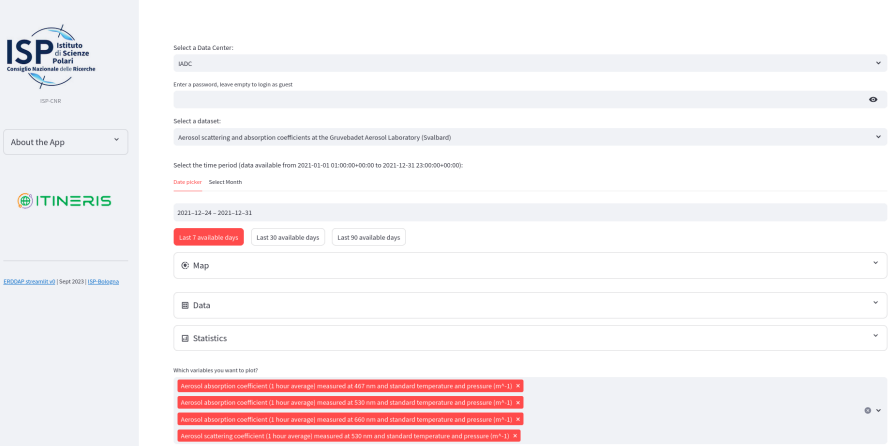


Figure 6: Streamlit dashboard dataset selection.

Supported visualizations include time series trends, histograms, and correlation matrices, which help users gain insights into variable relationships and detect potential data issues (Figure 7).

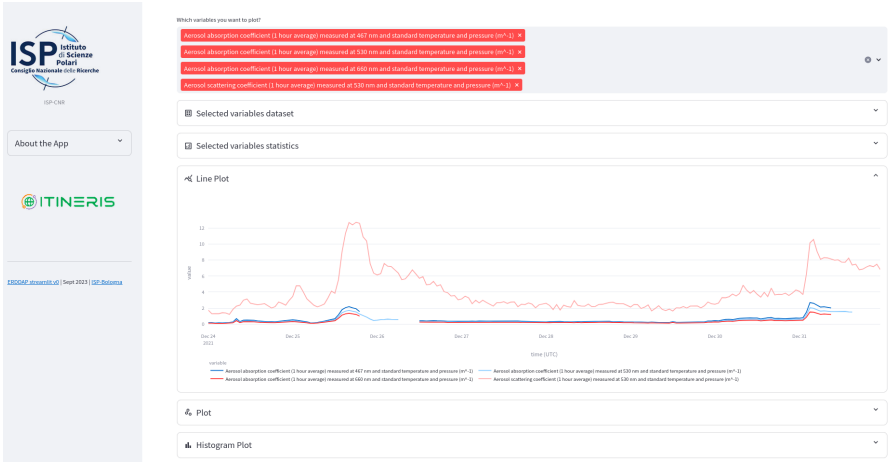


Figure 7: Streamlit dashboard subset variables dataset simple plot

6.2. Virtual research environnement

The objective was to create added value from the data stored in the polar data hubs to support and enhance the work of researchers. The main concept was to develop a Virtual Research Environment (VRE). This environment would contain a collection of web resources and Jupyter notebooks that leverage the data stored IADC and NADC, providing a comprehensive overview of the potential of the available datasets (Figure 8). Each resource within the VRE, whether a web resource or an example of Python code, is characterized by specific keywords to facilitate searchability. Additionally, each resource includes information on the origin of the data, specifying whether it comes from the IADC or NADC node.

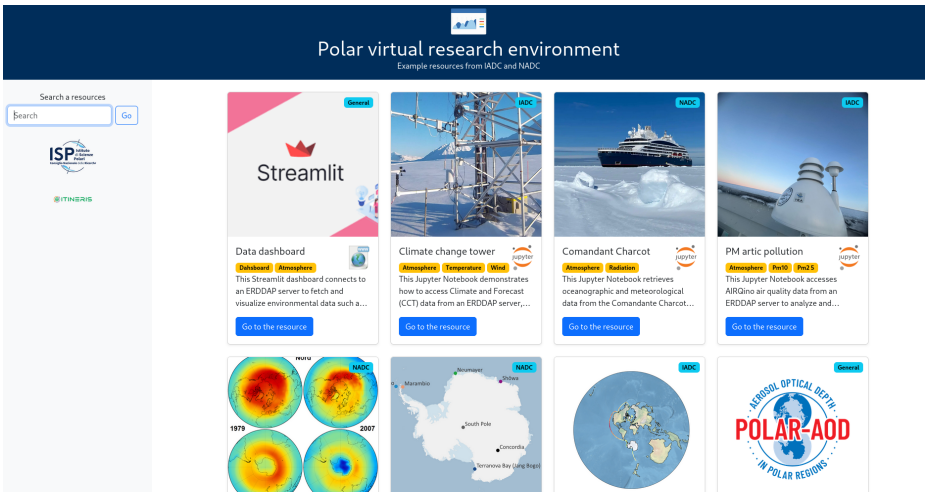


Figure 8: Main web page to access virtual research environment resources.

Currently, several Jupyter notebooks that utilize datasets related to the Climate Change Tower (Figure 9) and various vessels, such as the Commandant Charcot, have been developed and are available for Arctic data within ERDDAP™ integration.

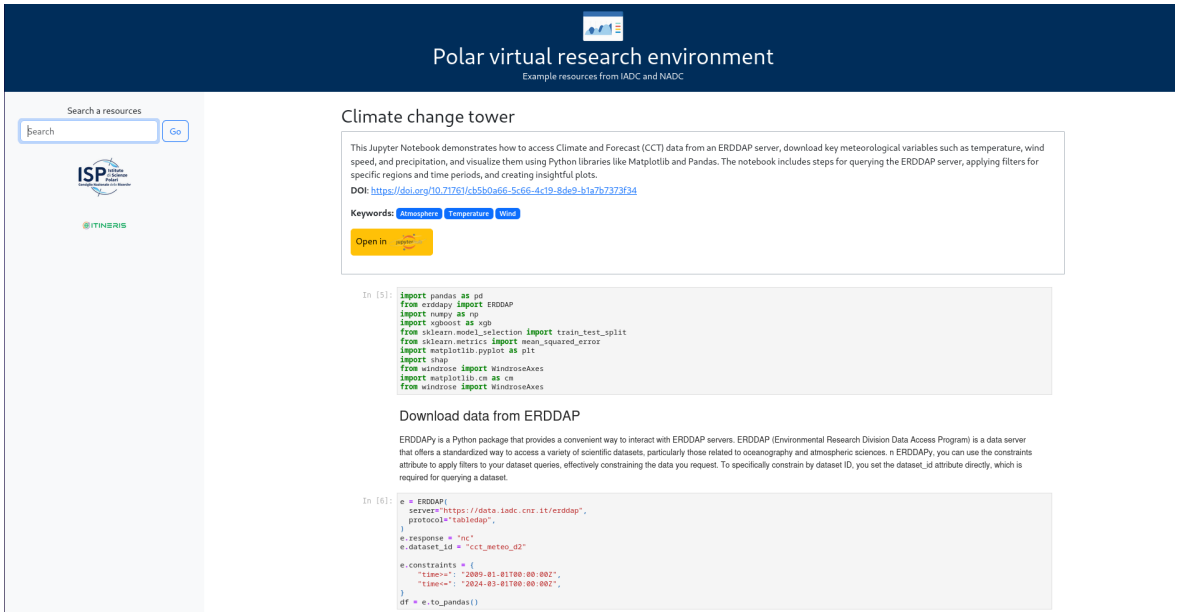


Figure 9: Example page of resource presentation with Climate change tower meteorological dataset detail and code example.

In the future, this virtual space will provide researchers with seamless access to the data, allowing them to analyse, visualize, and collaborate more effectively. The VRE will serve as an integrated platform for managing, sharing, and exploring datasets, offering tools to support various stages of the research process—from initial exploration to in-depth analysis. This initiative will unlock the full potential of the data, enabling researchers to extract meaningful insights, make informed decisions, and drive new discoveries from the extensive polar data collections.

7. ITINERIS HUB CONTRIBUTION

The CNR Institute of Methodologies for Environmental Analysis (CNR-IMAA) leads the development of the ITINERIS HUB, a national platform dedicated to environmental research infrastructures as part of the PNRR ITINERIS project. The institute specializes in cutting-edge research focused on the development and integration of Earth Observation technologies, including satellite, airborne, and ground-based platforms, with the aim of advancing the understanding of geophysical and environmental processes. CNR-IMAA also boasts a strong technological infrastructure and deep expertise in supporting Earth Observation research, particularly in the design and management of complex ICT systems. The CNR Institute of Polar Sciences (CNR-ISP) plays a crucial role in scientific and technological research in the polar regions, contributing to global change studies and supporting Italian and European environmental policies. In addition to managing the Dirigibile Italia Arctic Station, the institute has been instrumental in Italy's involvement in the SIOS infrastructure since its inception. To support the development of the ITINERIS HUB, initial collaborations between CNR-IMAA and CNR-ISP have already been established. According to the agreements made during the planning phase, CNR-IMAA will take the lead in coordinating and executing the technical development of the HUB, including defining standards, provisioning ICT resources, and collaborating with project partners. Meanwhile, CNR-ISP will assist with the development and management of service catalogs (CSW), facilitate the integration of the ITINERIS HUB with the SIOS Data Portal for metadata harvesting, and provide support in configuring templates for dataset creation, DOI assignment, and metadata import from project-designated endpoints.

8. CONCLUSION

An ongoing effort is being made to expand the capabilities of data catalogs and ERDDAP datasets. A key focus is improving the accessibility and usability of both, with the additional benefit of assigning Digital Object Identifiers (DOIs) to datasets. This makes the data more attractive to the scientific community, as DOIs ensure data traceability and enhance the visibility and accessibility of datasets.

Finally, from a future perspective, the integration of polar data with the ITINERIS HUB could enable the continuous collection and aggregation of data from IADC ERDDAP and Geonetwork. This integration will provide users with an even more streamlined and unified platform for accessing both metadata and datasets, allowing for easier data discovery and sharing across scientific communities.

9. REFERENCES

1 Chiarelli C., Longo S., Principato A., Verazzo G. and Vitale V., 2020, NADC – Italian Antarctic Data Centre. 2020, Data Science Journal.
2 Italian Arctic Data Center, URL: https://iadc.cnr.it
3 *ERDDAP*™ documentation, URL: https://github.com/ERDDAP/erddap
4 *ERDDAP*™ *y* documentation, URL: https://ioos.github.io/erddapy/
5 European Commission DG-R&I. (2015, October 01). "Guidelines on Data Management in H2020". Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
6 European Commission, 2016, "Open Science (Open Access)", URL: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access
7 GEO. (2015, October 01). GEOSS Data Management Principles. Retrieved from http://www.earthobservations.org/documents/dswg/201504_data_management_principles_long_final.pdf

8 GEO. (2015, October 01). The GEOSS Data Sharing Principles. Retrieved from https://www.earthobservations.org/geoss_dsp.shtml
9 Jupyter hub project, URL: https://jupyter.org/hub
10 INSPIRE, 2013, "Guidelines for the encoding of spatial data" available at: http://inspire.ec.europa.eu/documents/Data_Specifications/D2.7_v3.3rc3.pdf
11 Nativi S., Craglia M., Pearlman J., 2012, The Brokering Approach for Multidisciplinary Interoperability: A Position Paper. International Journal of Spatial Data Infrastructures Research, Vol.7, 1-15. Available at: http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/281/319
12 Nativi,S., Craglia M., Pearlman J., 2013, "Earth Science Infrastructures Interoperability: The Brokering Approach", Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, vol.6, no.3, pp.1118,1129, June 2013.
13 Nativi S., Mazzetti P., Santoro M., Papeschi F., Craglia M., Ochiai O., 2015, Big Data challenges in building the Global Earth Observation System of Systems, Environmental Modelling & Software, Volume 68, June 2015, Pages 1-26.
14 Stremlit framework: https://streamlit.io/