# Politica di gestione dei dati PNRA e infrastruttura NADC

<u>Vito Vitale</u> - Institute of Polar Sciences — Italian National Research Council (CNR-ISP) Emilia La Nave - International Relations Officer — Dipartimento per la Trasformazione Digitale (PCM)





## **Data Policy**

# Referimenti a livello Europeo e Nazionale (Agenda Digitale)

IT Gazzetta ufficiale dell'Unione europea 20.1.2023 L 19/43 REGOLAMENTO DI ESECUZIONE (UE) 2023/138 DELLA COMMISSIONE del 21 dicembre 2022 che stabilisce un elenco di specifiche serie di dati di elevato valore e le relative modalità di pubblicazione e riutilizzo (Testo rilevante ai fini del SEE)

2022

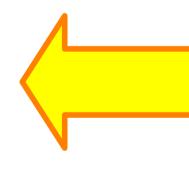




4.3 Serie di dati di elevato valore

LA COMMISSIONE EUROPEA,

4.4 Dati della ricerca



Linee Guida recanti regole tecniche per l'apertura dei dati e il riutilizzo dell'informazione del settore pubblico

Art. 12 D.Lgs. n. 36/2006 e s.m.i.



il piu' aperto possibile – chiuso il tanto necessario

# La Data policy del PNRA (2016)

Indicazioni sulla politica di gestione dei dati prodotti dal PNRA

A cura del GdL 'Raccolta diffusione dati del PNRA'

Stefano Aliani, Lili Cafarella, Guido Di Donfrancesco, Antonio Meloni, Giuseppe Orombelli, Nigel Wardell

- Garantire la stabile disponibilità dei prodotti dal PNRA, rendendoli sempre fruibili per la ricerca e la divulgazione scientifica;
- Agevolare la <u>diffusione dei dati</u> prodotti dal PNRA verso la comunità scientifica nazionale e internazionale;
- Garantire la **qualità e la tracciabilità dei dati** raccolti nell'ambito di progetti finanziati dal PNRA.
- 1. Breve definizione di dato e metadato
- 2. Descrizione del ruolo del National Antarctic Data Center (NADC)
- 3. Vincoli e accordi tra i proponenti di ricerca e PNRA
- Conferimento dei dati al NADC
- 5. Gestione condivisione e pubblicazione dei dati
- 6. Il diritto d'autore
- 7. La politica internazionale dei dati e lo SCAR.

## Data Policy PNRA: licenze di utilizzo

### **Creative Commons**

Simboli	Sigla	Descrizione	
CCC BY	CC BY	Permette di distribuire, modificare, creare opere derivate dall'originale, anche a scopi commerciali, a condizione che venga riconosciuta la paternità dell'opera all'autore.	
CC (1) (S) BY NC	CC BY-SA	Permette di distribuire, modificare, creare opere derivate dall'originale, anche a scopi commerciali, a condizione che venga riconosciuta la paternità dell'opera all'autore e che alla nuova opera vengano attribuite le stesse licenze dell'originale (quindi a ogni derivato verrà consentito l'uso commerciale).	

Ad ogni data set possono essere associate una o più tipologie di licenza in base al livello di protezione richiesto dal tipo di dato.

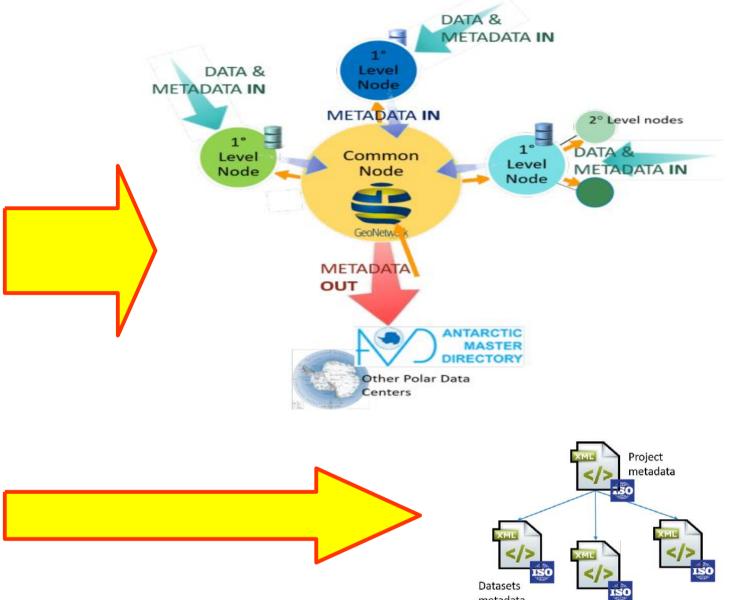
Il Coordinatore/Responsabile scientifico del progetto deve indicare al momento di firmare il contratto il tipo di licenza che vuole attribuire.

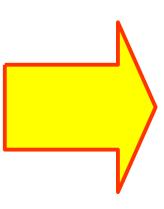
Permette di distribuire l'opera originale senza alcuna modifica, anche a scopi commerciali, a CC BY-ND condizione che venga riconosciuta la paternità dell'opera all'autore. Permette di distribuire, modificare, creare opere derivate dall'originale, a condizione che venga riconosciuta la paternità dell'opera all'autore, ma CC-BY-NC non a scopi commerciali. Chi modifica l'opera originale non è tenuto a utilizzare le stersse licenze per le opere derivate. Permette di distribuire, modificare, creare opere derivate dall'originale, ma non a scopi commerciali, a condizione che venga riconosciuta CC BY-NC-SA la paternità dell'opera all'autore e che alla nuova opera vengano attribbuite le stesse licenze dell'originale (quindi ad ogni opera derivata non sarà permesso l'uso commerciale) Questa licenza è la più restrittiva: consente soltanto di scaricare e condividere i lavori origniali CC BY-NC-ND a condizione che non vengano modificati né utilizzati a scopi commerciali, sempre attribuendo la partnità dell'opera allautore.

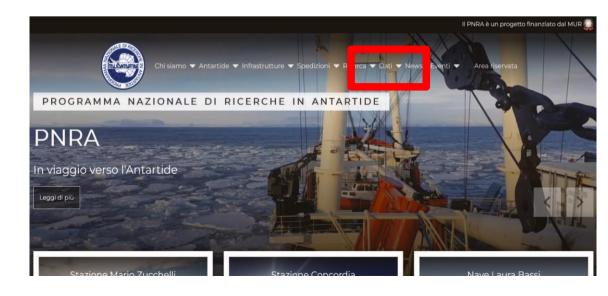
Tale scelta và ribadita al momento di compilare il metadato nel sistema NADC

### **ANTARCTIC NATIONAL DATA CENTER (NADC)**

Since 2018 developed thanks a dedicated support from PNRA on the basis of borakering and FAIRNESS principles Designed as distributed system with a common node manage by CNR and first level nodes manage by CNR, ENEA, OGS, INGV and MNA Based on open source software/platforms largely used at international level (GEONETWORK, ERDDAP) Metadata model developed using the ISO 19115 standard, and following a a hierarchical structure that allows a detailed descri-ption of different datasets referring to the same project context. Access through PNRA web portal WWW.PNRA.AQ CNR will secure continuity to NADC, leveraging the human resources recruited so far (supported by the CNR IT services) Synergy with similar actions devoted to the Arctic (IADC) and PNRR-ITINERIS project will provide cost-effective way to move forward.

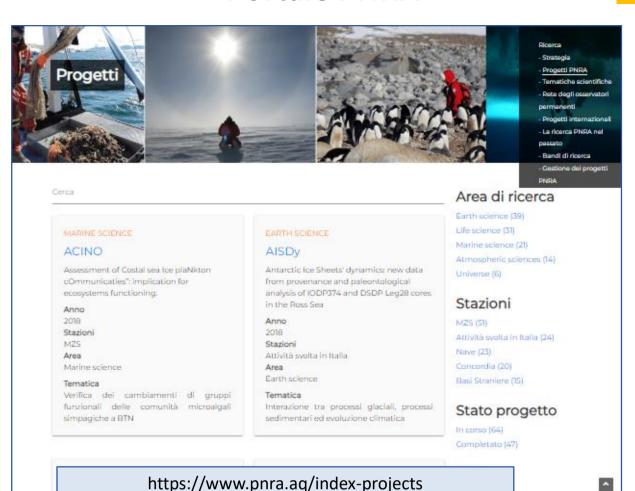


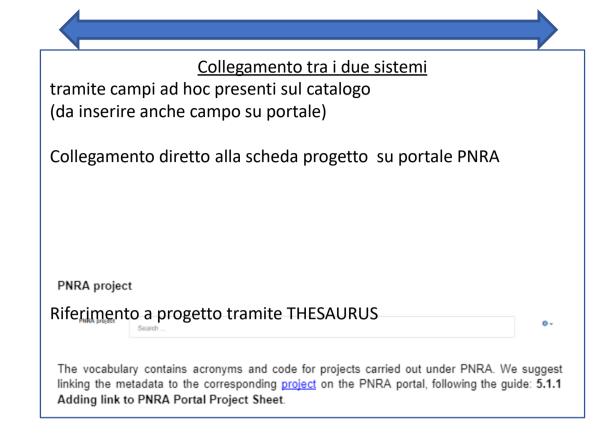




#### **Portale PNRA**

## I sistemi di catalogazione del PNRA





#### 111 schede progetto

33 CNR

3 CONsiMA

2 ENEA

1 CMCC

3 INAF 11 INGV

1 ISS

10 OGS

1 Osservatorio Astronomico Valle d'Aosta

2 Stazione Zoologica Anton Dohrn

42 università

#### Area di ricerca

Earth science (39) Life science (31) Marine science (21) Atmospheric sciences (14) Universe (6)



Scaricare e collegamenti

RESTORE - PNRA project

monitoring-ross-sea

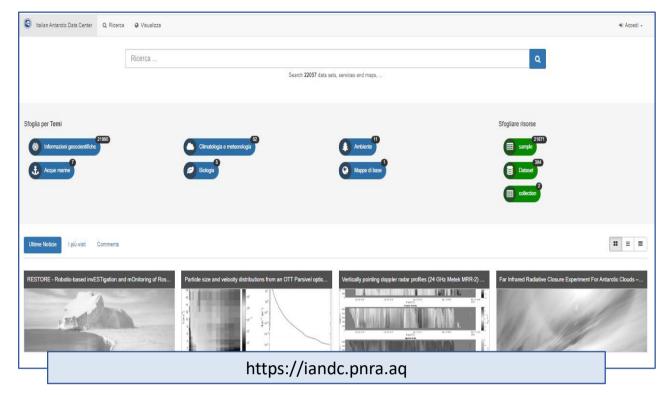
https://www.pnra.aq/it/project/333/robotic-based-investigation-and-

Le **schede Progetto** forniscono

informazioni generali sui progetti

I **metadati** descrivono Progetti o dataset o campioni

#### Catalogo metadati NADC



**CNR 29** 

**ENEA 39** 

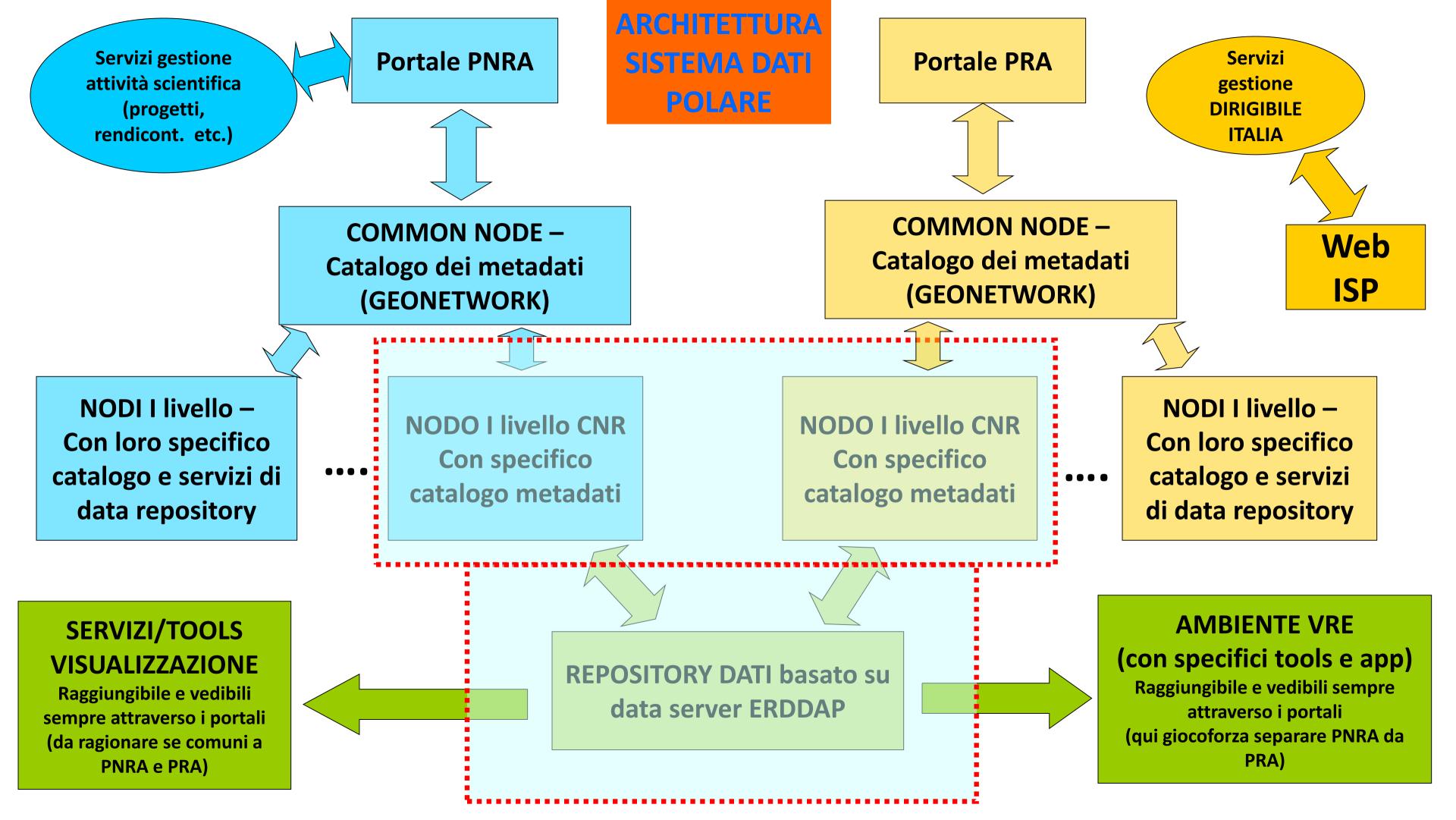
INGV 13

#### **22057** record

Sample 21671: campioni museo Antartide

Dataset 384: CNR ENEA INGV OGS

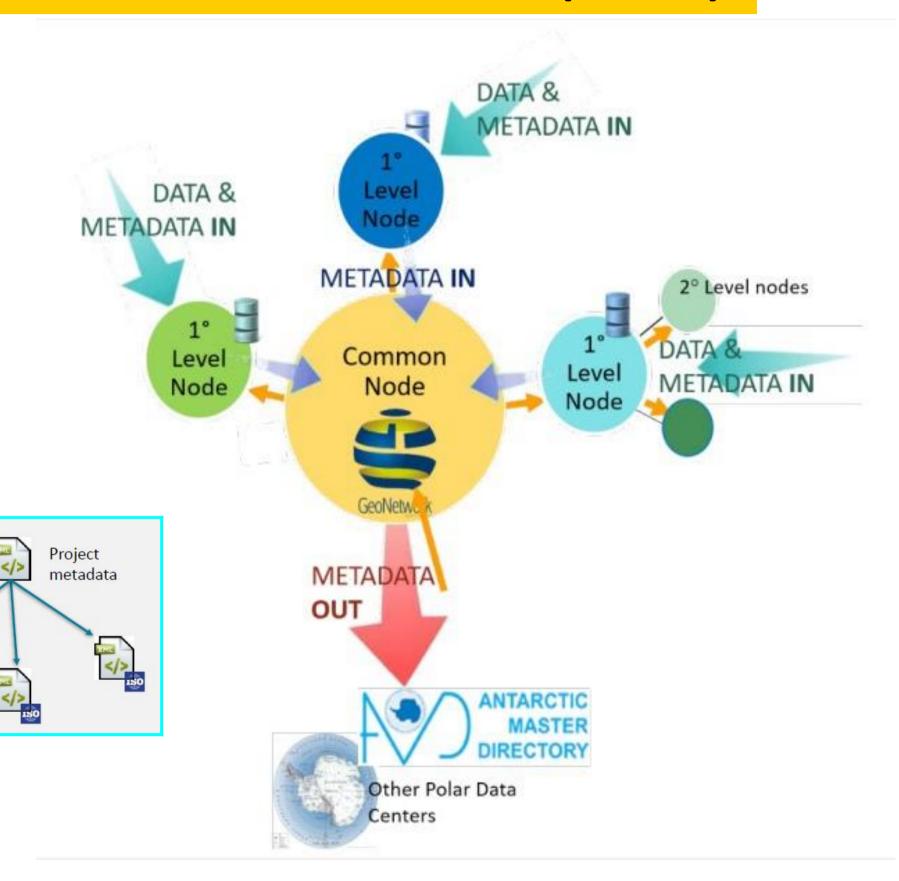




# Data & Metadata flow at Italian National Antarctic Data Centre (NADC)

Datasets metadata

- Metadata & Data are managed remotely by 1st level nodes
- Metadata are harvested by the Common Node (central infrastructure)
- Metadata are exported to other external systems from the Central node
- Metadata records:
  - Standard format ISO19115
  - Hierarchical organization of records
  - Thesaurus: GCMD and NASA
  - Record link to external resources (Data repositories, Project websites, Dataset direct download)



## National Data Repositories e contesto internazionale

### **Polar Data Forum**

The Polar Data Forum (PDF) is a place where polar data holders get together and make more use of data. The Forum has two main components: the Conference, where the border between funding, policy and data is explored through presentations and posters; and Workshop Sessions & Hackathons, where the Polar Data Community opens the dialogue to make polar data forum. Cit.

## **SCADM**

ambridge



# STANDING COMMITTEE ON ANTARCTIC DATA MANAGEMENT (SCADM)





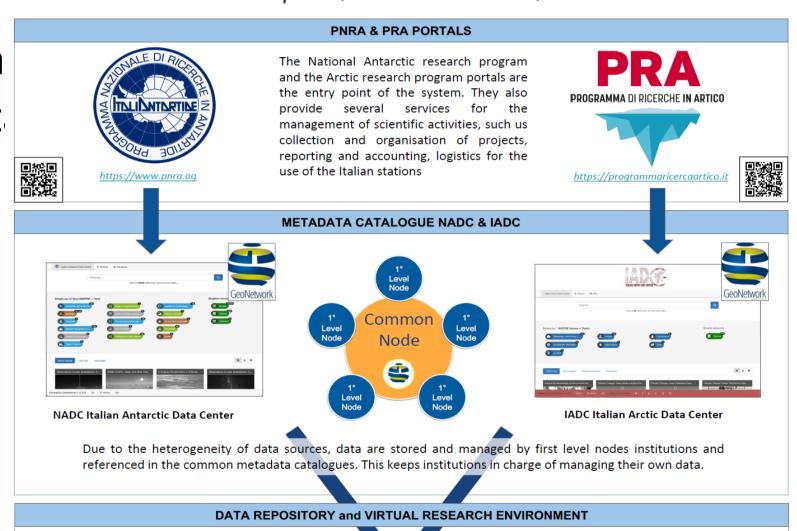




Rappresentante Italiano: ALBERTO SALVATI

# Status and Development of the Italian Polar Repository

Chiara Ripa<sup>1</sup>, Giulio Verazzo<sup>1</sup>, Alice Cavaliere<sup>1</sup>, Alberto Salvati<sup>2</sup>, Antonino Principato<sup>2</sup>, Mauro Mazzola<sup>1</sup>, Vito Vitale<sup>1</sup>

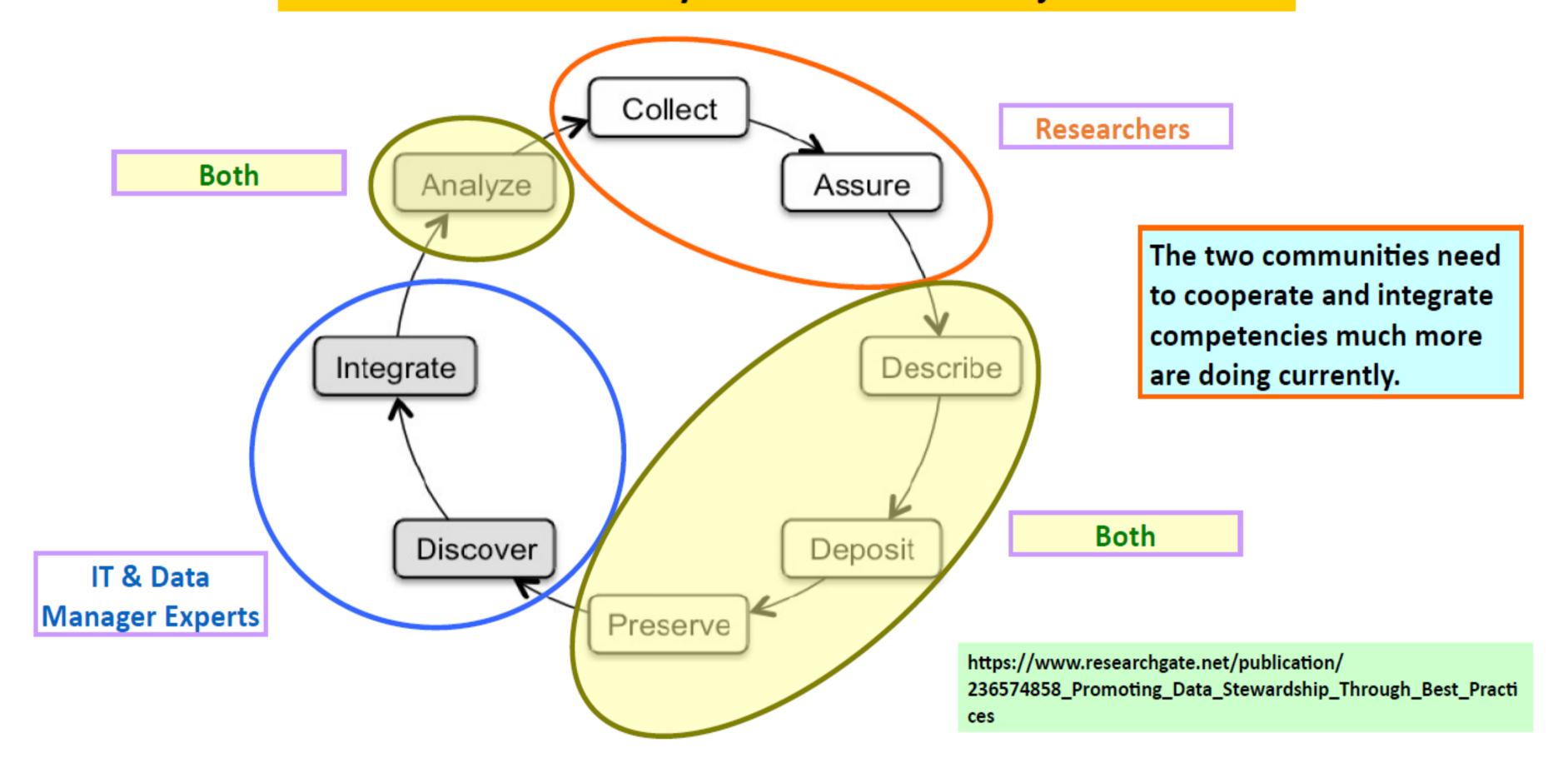


The Italian Institute of Polar Sciences manages an ERDDAP data server where

data is uniformed to a common format where possible. This allows to extract added value from data using a Jupyter notebook

based virtual research environment

# Data Life Cycle: a collaborative job



# DESCIBE: the importance to document all steps ... and the challenge

The Journey of our data from the measurement site (or laboratory) to the correct insertion in a data center or data repository (DEPOSIT/INGEST, PRESERVE) includes a fair number of steps and the use of different devices/tools/software.

#### Collect/Store Row Data

Locally and/or remotely in a computer

#### From Row to usable data

Using calibration costants, removing instrumental errors, averaging when useful/necessary

#### **Data Qualification and Assurance**

Application of QA procedure, data filtering, when necessary/useful recovery of hole in the data series

#### Structure of data

Organize data in files and directories

#### Semantic annotation of data

Conventions for naming variables in a consistent manner

Physical encoding of data in a file format/RDMS

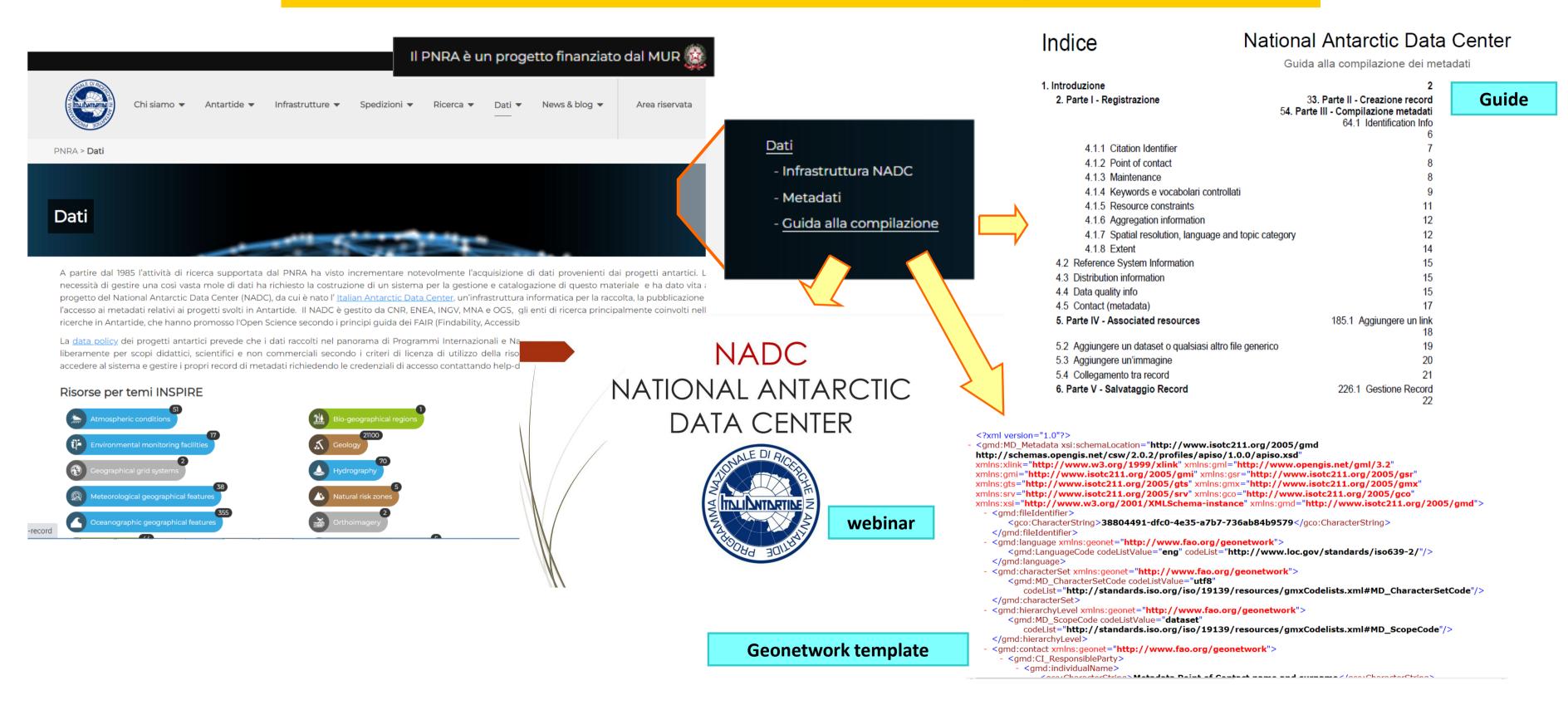
Ensure FAIR guiding principles are met

#### THE CHALLENGE

plan collect assure/ qualify analyse publish/ present

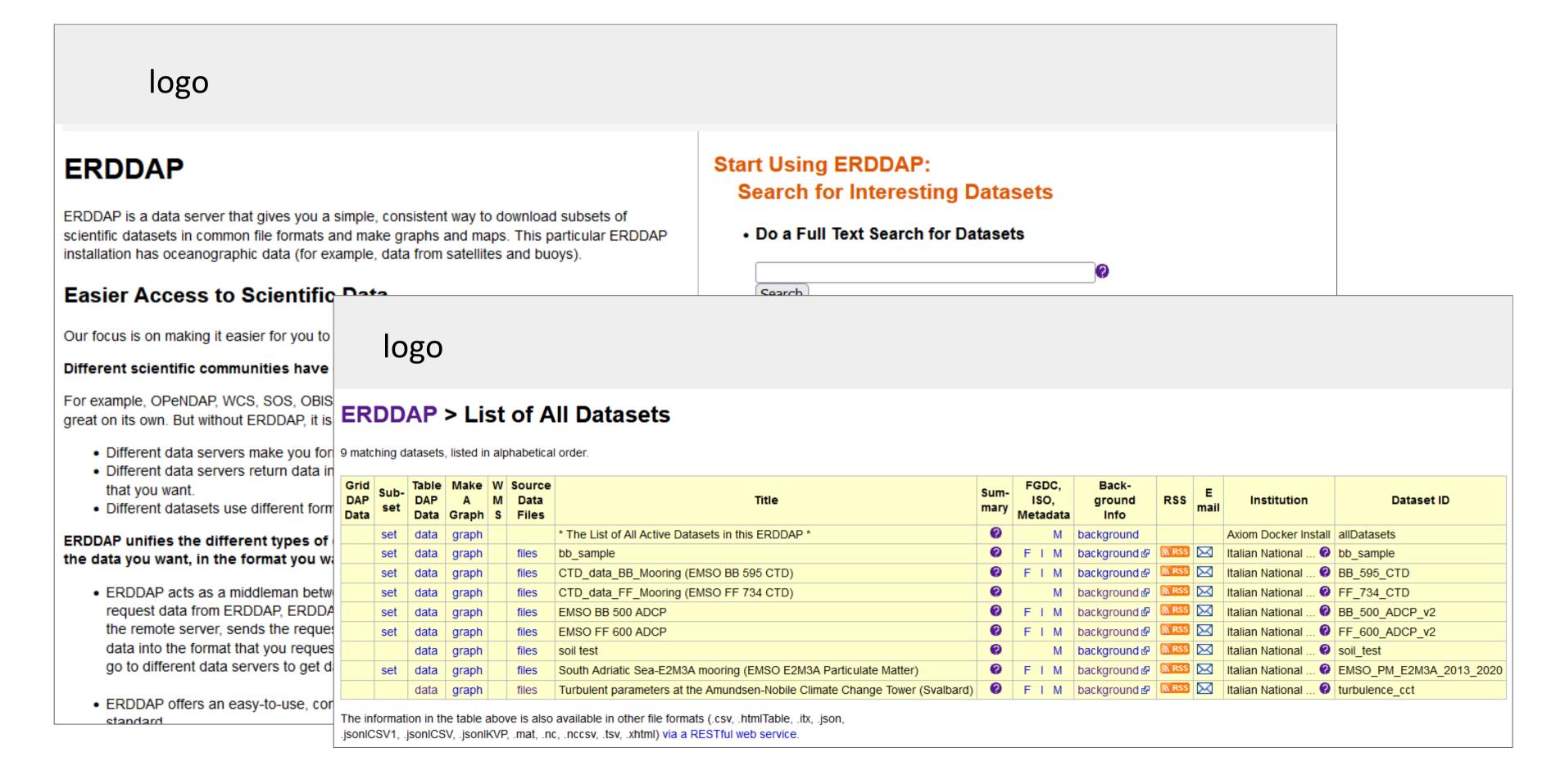
Data lyfe Cycle for a Researcher

## **DESCIBE, DEPOSIT, PRESERVE: work togehter (1)**



Tra gennaio e marzo verranno programmati ulteriori webinar (1 al mese) per portare la comunita' scientifica polare a ben conoscere e poi utilizzare questi strumenti

Per i nostri repository polari abbiasmo scelto il software open source ERDDAP per la gestione dei dataset polari. ERDDAP è un data server sviluppato da NOAA ed è attualmente in uso presso 19 organizzazioni in 14 paesi.



#### **ERDDAP-CMS**

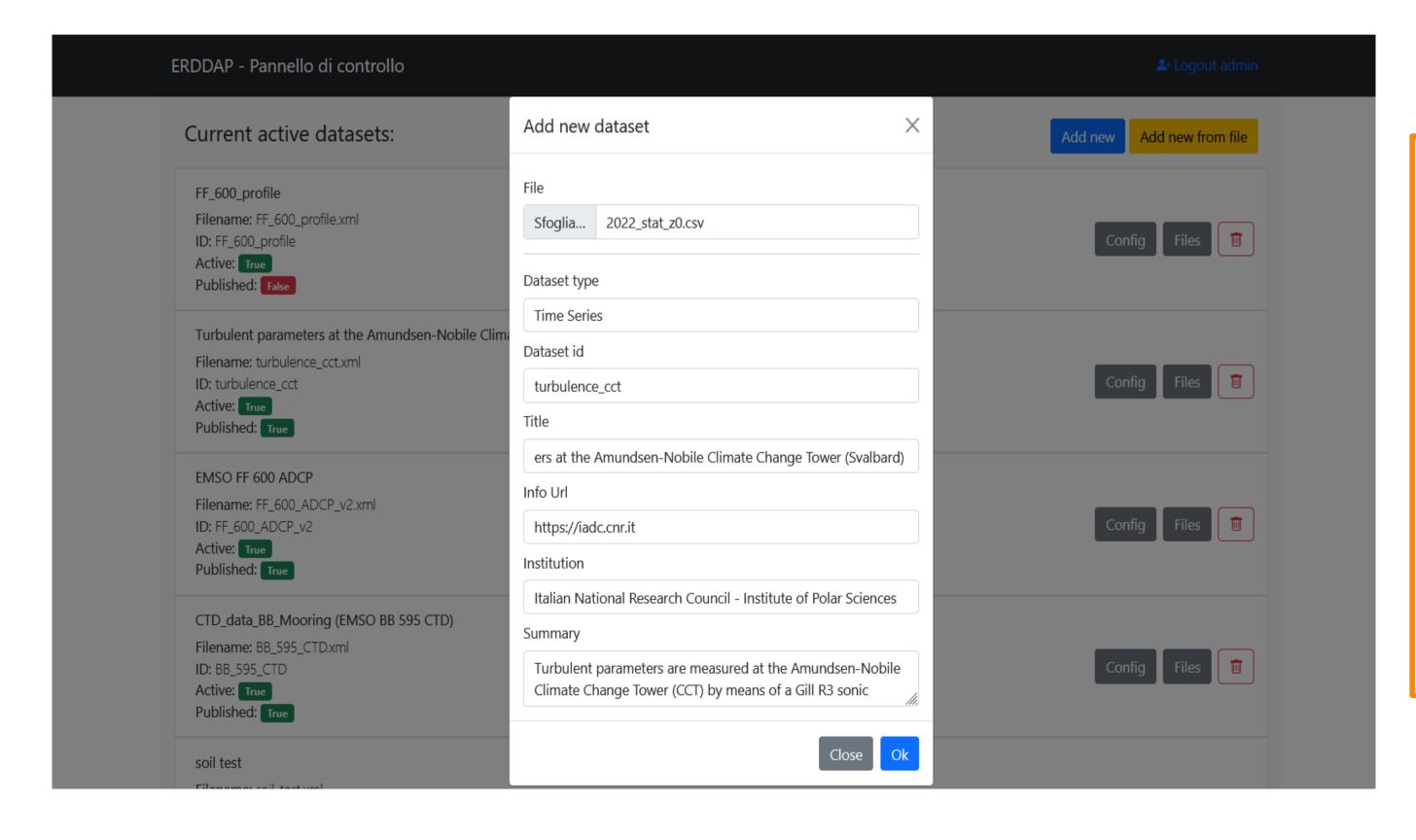
ERDDAP non ha un'interfaccia di gestione dei dataset e prevede l'utilizzo di script da linea di comando e file di configurazione.

Lo sviluppo di un Content Management System (CMS) per ERDDAP nasce dall'esigenza di semplificare l'aggiunta e la gestione dei dataset all'interno del data repository polare tramite l'utilizzo di un'interfaccia web.

```
EDDGridFromEDDTable
                                    EDDTableFromFileNames
EDDGridFromErddap
                                    EDDTableFromHttpGet
EDDGridFromMergeIRFiles
                                    EDDTableFromInPort
EDDGridFromNcFiles
                                    FDDTableFromToosSOS
EDDGridFromNcFilesUnpacked
                                    EDDTableFromJsonlCSVFiles
EDDGridFromThreddsCatalog
                                    EDDTableFromMultidimNcFiles
EDDGridLonPM180FromErddapCatalog
                                    EDDTableFromNcFiles
EDDGridLon0360FromErddapCatalog
                                    EDDTableFromNcCFFiles
EDDTableFromAsciiFiles
                                    EDDTableFromNccsvFiles
EDDTableFromAudioFiles
                                    EDDTableFromOBIS
EDDTableFromAwsXmlFiles
                                    EDDTableFromSOS
EDDTableFromBCODMO
                                    EDDTableFromThreddsFiles
EDDTableFromCassandra
                                    EDDTableFromWFSFiles
EDDTableFromColumnarAsciiFiles
                                    EDDsFromFiles
                                    addFillValueAttributes
EDDTableFromDapSequence
EDDTableFromDatabase
                                    findDuplicateTime
EDDTableFromEDDGrid
                                    ncdump
Which EDDType (default="EDDGridFromDap")
```

```
<dataset type="EDDTableFromAsciiFiles" datasetID="turbulence cct" active="true">
 <reloadEveryNMinutes>10080</reloadEveryNMinutes>
 <updateEveryNMillis>10000</updateEveryNMillis>
 <fileDir>/datasets data/turbulence cct/</fileDir>
 <fileNameRegex>.*</fileNameRegex>
 <recursive>true</recursive>
 <pathRegex>.*</pathRegex>
 <metadataFrom>last</metadataFrom>
 <standardizeWhat>0</standardizeWhat>
 <charset>ascii</charset>
 <columnSeparator>,</columnSeparator>
 <columnNamesRow>1</columnNamesRow>
 <firstDataRow>2</firstDataRow>
 <sortedColumnSourceName>Datetime</sortedColumnSourceName>
 <sortFilesBySourceNames>Datetime</sortFilesBySourceNames>
 <fileTableInMemory>false</fileTableInMemory>
 <addAttributes>
   <att name="cdm_data_type">TimeSeries</att>
   <att name="Conventions">COARDS, CF-1.10, ACDD-1.3</att>
   <att name="creator name">Italian National Research Council - Institute of Polar Sciences</att>
   <att name="creator url">https://iadc.cnr.it</att></att>
   <att name="infoUrl">https://iadc.cnr.it</att>
   <att name="institution">Italian National Research Council - Institute of Polar Sciences</att>
   <att name="keywords">a NaN, amundsen, amundsen-nobile, change, climate, council, data, datetime, dir, direction
   <att name="license">[standard]</att>
   <att name="sourceUrl">(local files)</att>
   <att name="standard name vocabulary">CF Standard Name Table v70</att>
   <att name="summary">Turbulent parameters are measured at the Amundsen-Nobile Climate Change Tower (CCT) by me
   <att name="title">Turbulent parameters at the Amundsen-Nobile Climate Change Tower (Svalbard)</att>
   <att name="cdm timeseries variables">station id,latitude,longitude</att>
 </addAttributes>
 <dataVariable>
   <sourceName>Datetime</sourceName>
   <destinationName>time</destinationName>
   <dataType>String</dataType>
   <addAttributes>
     <att name="long name">Datetime</att>
     <att name="source_name">Datetime</att>
     <att name="standard name">time</att>
```

in fase di aggiunta di un dataset, l'utente carica il file (CSV o NetCDF) e compila la scheda con alcuni campi obbligatori. Il sistema supporta l'aggiunta di dataset di tipo serie temporali (timeSeries), profili di serie temporali (timeSeriesProfile) e generico (other).



Tra marzo e giugno verranno programmati webinar (1 al mese) per portare la comunita' scientifica polare a ben conoscere e poi utilizzare questa interfaccia e iniziare a inserire dati. Si punterà in partenza su alcune categorie specifiche, in particolare a coinvolgere gli osservatori, che non sono gestiti da specifici nodi di primo livello.



#### NATIONAL ANTARCTIC DATA CENTRE









# https://survey.pnra.aq/index.php?r=survey/index&sid=422736&newtest=Y&lang=it

#### LOGOTIPO

COLORE

FONT

OSWALD LIGHT
OSWALD BOLD
AZONIX REGULAR

#### **PITTOGRAMMA**

COLORE



#### **CONCETTI ESPRESSI**









ARTIDE E ANTARTIDE

ICEBERG

DATA CENTER

NORD E SUD



# https://survey.pnra.aq/index.php?r=survey/index&sid=422736&newtest=Y&lang=it

#### NATIONAL **ANTARCTIC** DATA CENTRE









#### LOGOTIPO

COLORE

FONT

DENSE REGULAR
DENSE BOLD
JOST BLACK

#### **PITTOGRAMMA**

COLORE



#### **CONCETTI ESPRESSI**









ARTIDE E ANTARTIDE

ICEBERG

DATA CENTER

NORD E SUD

