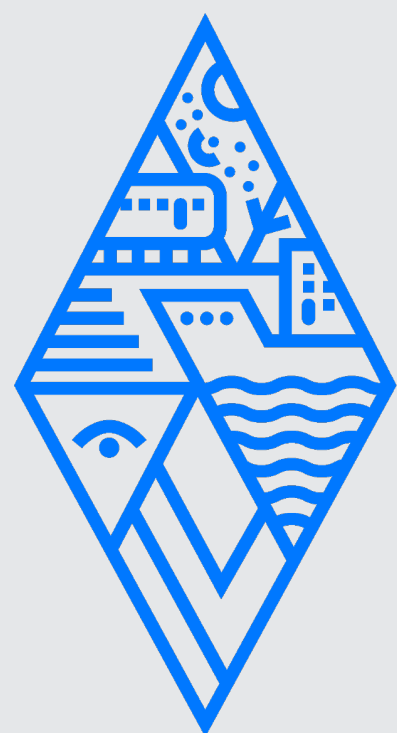




POLARIN

POLAR
RESEARCH
INFRASTRUCTURE
NETWORK

Work Package 4
Improvement of data
services and
customised data
products

Antonio Novellino, ETT
Vito Vitale, CNR



AARHUS UNIVERSITY



Norsk institutt for Luftforskning
Norwegian Institute for Air Research



inkode



FUNDED BY THE
EUROPEAN UNION

1. Objectives



Facilitating the discovery and consumption of data that is already available by (i) ensuring databases, datasets and/or data initiatives from polar regions are identifiable, (ii) facilitating access to them, (iii) improving workflows and reducing data Fragmentation, (iv) making available tools to facilitate the access of data from multiple providers as well as data integration.

➤ Specific WP4 objectives:

- *Establish a web portal with workflow-based access to data for various data consumer communities.*
- *Provide data products and tools integrating data across data repositories.*

And in relation with other WPs:

- *Making available M2M interfaces to include unified VA data catalogue*
- *Making available services for integration of VA data and for the challenge-driven call definition in TA1/WP1.*



2. Tasks

Task 4.1: Improving access to polar data (ETT, CNR)
design/implement the Hub. Focus on general users

Task 4.2: Improving polar data availability and interoperability
(CNR, AU) develop/implement tools services to improve
accessibility, interoperability, data integration.

Task 4.3: Data products, tools and data services (SIOS, NILU)
develop/implement data products and services for
specific identified topics (Glacier tomography, short-lived
atmospheric species, cryosphere)



3. Partners roles and responsibilities

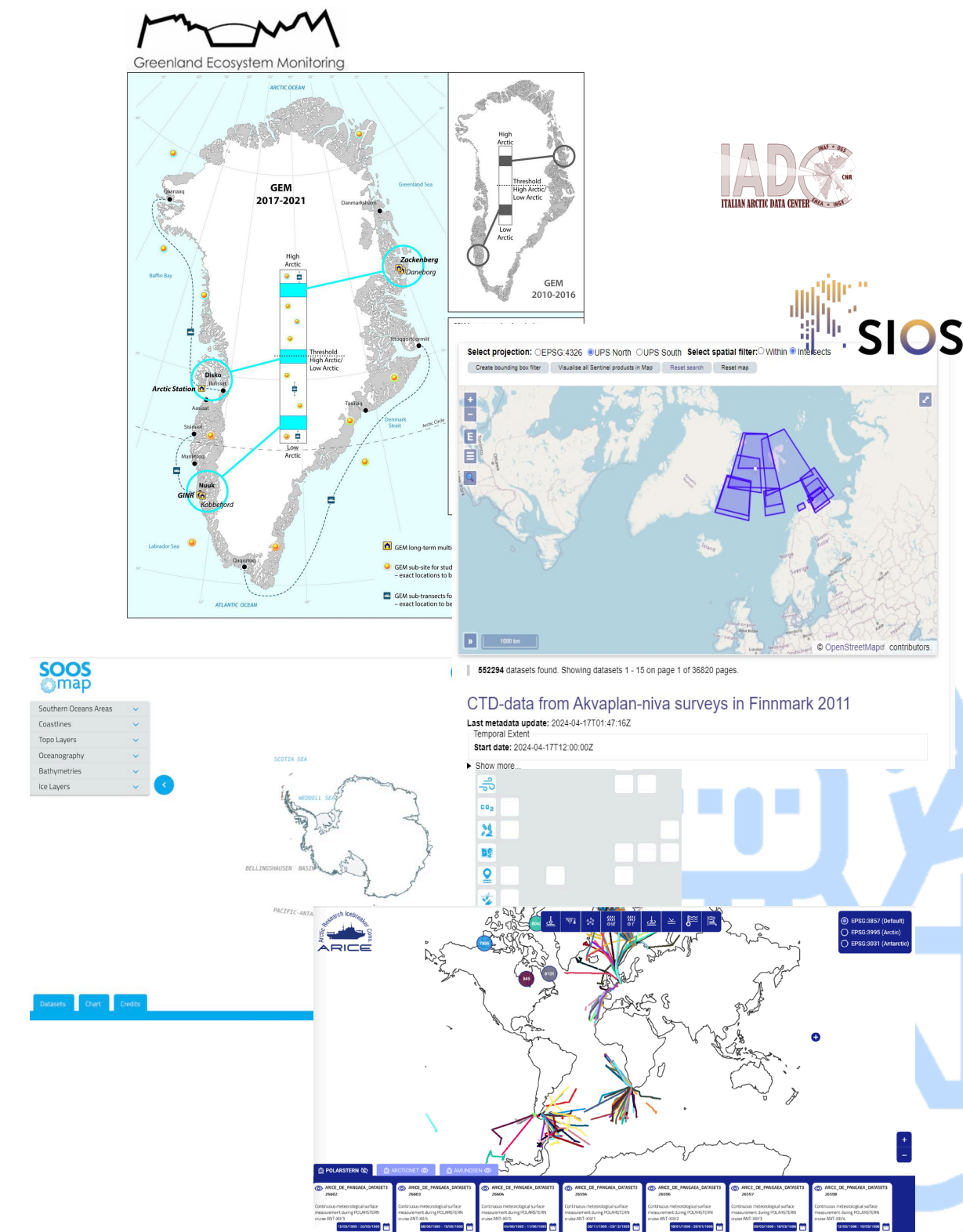


Task 4.1: Improving access to polar data (CNR, ETT)

- reduce data fragmentation and data search complexity by interoperating with SAON/IASC, SCADM, SOOS and IARPC, Arctic PASSION, and Ris (SIOS, INTERACT, ISAFFIK, ACTRIS)
- set up a not for expert web-tool to discover data

Task 4.2: Improving polar data availability, interoperability, integration (CNR, AU)

- to assess and document the full framework of these interoperability approaches, a second goal is to provide the partners with tools and guidelines to improve interoperability in practice
- cross-platforms serving diverse disciplines (catalogue)
- training/deploying/working on FAIR compliant community standards
- Provide/develop tools integrating data across data repositories



3. Partners roles and responsibilities



Task 4.3: Data products, tools and data services (SIOS, NILU):

- selected data products (maps with key climate and ecosystem variables from existing raw data) will be developed to contribute to the identification of gaps in geographical coverage and gaps in process understanding
- with POSEDA network: Glacier tomography in Greenland and Antarctica derived from seismic networks, providing information on glacier dynamics
- with ESFRI Landmark ACTRIS: Footprint products of short-lived atmospheric species (e.g., black carbon) for analysis of measurement data
- with SIOS, regional cryosphere datasets covering snow, glacier and permafrost aspects will be established.



4. Approach

- Data Management Plan
- Gap and needs analysis (broad audience)
- Focus on a large spectrum of stakeholders
- Open science developments
 - standards, open tools, open sw, ...
- sharing tools and tools for sharing developments
 - ERDDAP, GeoNetwork, ...
 - drive, github, ...
- Agile approach (develop, deploy, test, update, ...)

The collage illustrates the POLARIN project's approach to data management and open science. It features the POLARIN logo, a Google Drive folder named 'Polarin PUBLIC' containing various files, the GitHub repository 'EMODnet-Physics' with a 'Let's build from here' banner, and a Jupyter Notebook titled 'gem_api_example.ipynb' showing code for using the GEM API to get a list of datasets and query for data from a specific dataset.

https://github.com/EMODnet-Physics
you can get your docker (ready to use installation package)

```
[ ] # imports required to perform https requests to the api, and handle the json formatted responses
# pandas is for plotting csv
import requests
import json
import pandas
import sys
from io import StringIO

[ ] headers = {'X-API-KEY': 'test'} # the api key
```

5. Next steps for implementation



- Data Management Plan (M6)
 - “as open as possible, as closed as necessary” → CC-BY/CC0
- Mapping of standards and technologies
 - common baseground, best practice transfer,
 - common metadata/vocabularies and semantics
- Mapping of resources (unified POLARIN catalogue)
- Mockups of web tools and catalogue
- Build services and tools for integrating resources/infrastructure and make added value services and products



6. Collaboration with other WPs/programmes



- daily interaction with WP5 for desining and developing backend tools and services (webportal, catalogue, interoperability tools ...)
- Support to WP1/WP2 for gap analysis
- Support to WP1/WP2 for implementing TA DMPs
- Interaction with WP2 to integrate TA outcomes in the POLARIN catalogue
- Interaction with WP7 (contents on DM for training)
- SOOS, EMODnet, CMEMS, WMO, ..., SCADM, Polar Data Forum, ...



7. Risks and challenges for implementation



WP4/WP5 workshop shows very good starting point and alignment on strategy and tools (low risk)

good understanding of DMP and related commitments

WP1/WP2 needs feedback on DMP by October to implement TA properly

(wider) stakeholders panel to assess/test developments and fit-for-use outcomes



8. Upcoming deliverables (Year 1)



Deliverables:

- D4.1 Data management plan, M6, (ETT)
- D4.2 POLARIN graphic products package for multiple audiences and gap analysis, M10, (AU)
- D4.3 POLARIN web data portal, M12, (ETT)
- D4.4 Guidance on dataset granularity, M12, (CNR)

Milestones:

- M4.1 POLARIN web data portal implemented (M12)



8. Upcoming deliverables (Year 1)



Data Management Plan

- The Data Management Plan (DMP) outlines the ways in which data is collected, generated and/or processed throughout the lifespan of a research project
- DMP is compulsory in Horizon Europe as pillar of European Commission goals to advance Open Science policy and practices
- Although the name may suggest that the purpose of the DMP is to technically describe the way data will be handled, its actual purpose is to ensure the availability and utility of the project's research data
- The plan outlines the measures that will be taken in order to maximize access and re-use of the data for further purposes and applications.

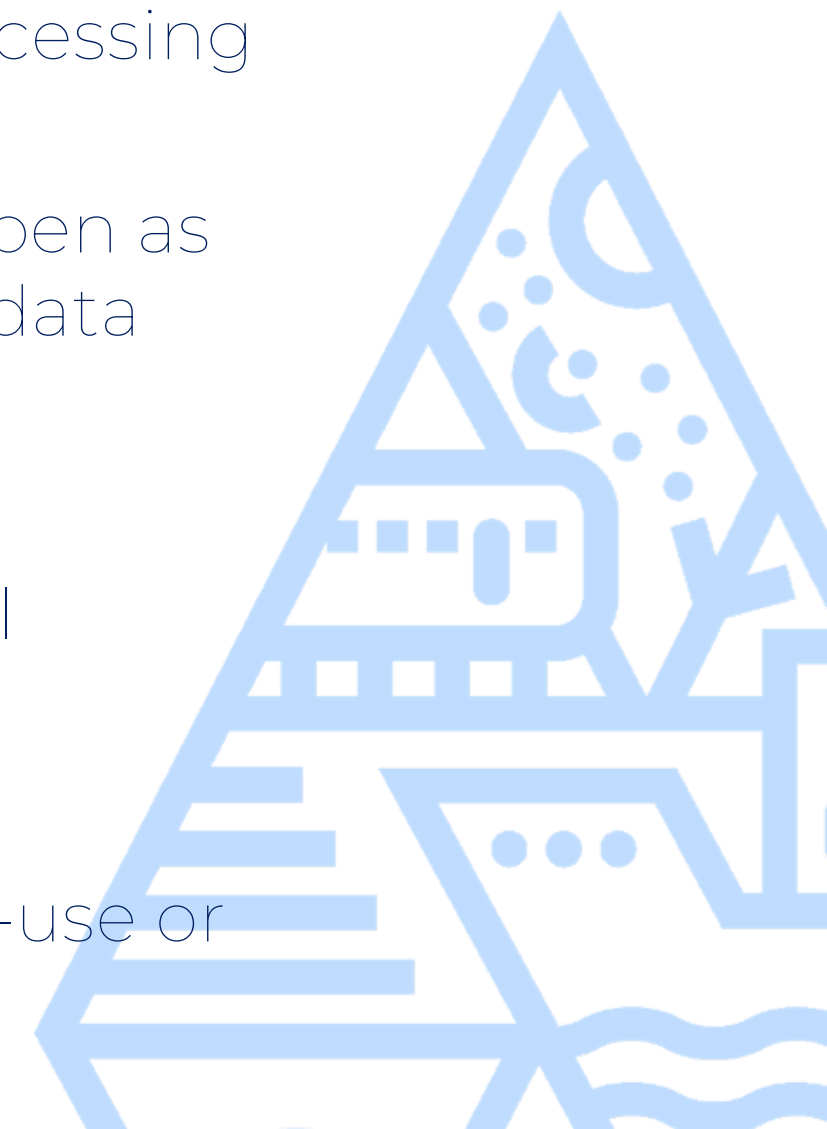


8. Upcoming deliverables (Year 1)



Data Management Plan

- Project must manage the digital research data in line with the FAIR principles
- Data Management Plan (DMP) is required by M6; updated periodically (midproject and at end of project)
- Deposit (meta)data as soon as possible after production/generation or after processing and quality controls
- Make them open as soon as possible (deadlines set in DMP), following the “as open as possible, as closed as necessary” (open by default) principles – in any case metadata must be FAIR and under CC0
- Open licence, preferentially CC-BY or CC0 licence
- Deposit data in a trusted repository (federated in the EOSC if required in the call conditions)
- Detailed information about research outputs or tools/instruments needed to re-use or validate the data (e.g. data, software, algorithms, protocols, models, workflows, electronic notebooks)



9. Expected results



- Facilitate the discovery and consumption of data that are already available
- Improve workflows among data initiatives by reducing data fragmentation
- Develop tools (stepwise/wizard-based approach) to help user to access to fit-for-purpose data
- Develop web-based tools to facilitate data discovery
- Develop tools to help integration of data across data repositories

