**HORIZON 2020 - Coordination and Support Action**

**Grant Agreement No: 101003766**

EU-PolarNet 2 - Co-ordinating and Co-designing the European Polar Research Area

Deliverable No. D6.1

# Procedure for ongoing collection and collation of European Polar observing capacities and activities

# Submission of Deliverable

| Document information | |
|---|---|
| Work Package | WP 6 – European Polar Coordination Office |
| Deliverable No | D6.1 |
| Deliverable title | Procedure for ongoing collection and collation of European Polar observing capacities and activities |
| Version | Draft version 30th January 2022 |
| Dissemination level | ☒ PU - Public<br>☐ PP - Restricted to programme partners<br>☐ RE - Restricted to a group specified by the consortium<br>☐ CO - Confidential, only for members of the consortium |
| Lead Beneficiary | AMAP (partner 16) |
| Contributors | ☐ 1 – AWI, ☐ 2 – MICINN, ☒ 3 – UOULU, ☐ 4 – ISP-CNR, ☐ 5 – RCN, ☒ 6 – EPB, ☐ 7 - NWO, ☐ 8 – DAFSHE, ☐ 9 - CNRS, ☐ 10 – UoS-CPS, ☐ 11 – BAI, ☐ 12 – UNIVIE, ☐ 13 –IG-TUT, ☐ 14 – WOC Europe, ☒15 – BELSPO, ☒ 16 – AMAP, ☐ 17 – IGOT UL, ☐ 18 - SPRS, ☐ 19 – UKRI-BAS, ☐ 20 – ITU, ☐ 21 – USB, ☐22 – RANNIS, ☐ 23 – FAMRI, ☐ 24 – ICR, ☐ 25 – SPI |
| Coordinating author | Jan Rene Larsen, AMAP Secretariat |
| Contributing authors | Renuka R. Badhe, Nicole Biebow, Joseph Nolan, Anton Van de Putte, Hannele Savela, Annette Scheepstra, Serge Scory, Anneli Strobel |
| Due date | 30 September 2021 |
| Delivery date | 07 February 2022 |

| Document history | |
|---|---|
| Creation Date | 2021 |
| Revision | V1 |
| Revision Date | 31 January 2022 |
| Author | Jan Rene Larsen, AMAP Secretariat |
| Status | ☐ Draft<br>☒ WP lead approved<br>☒ Coordinator approved<br>☒ Executive Board approved |
| Status date | 07 February 2022 |

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003766

# TABLE OF CONTENTS

## Summary

A diverse and distributed range of existing polar observing facilities and activities could be better utilised to address societally relevant scientific problems. The fundamental challenge related to the utilisation of polar observations is that observing networks and observing systems that produce observations are diverse and distributed of nature. There are many networks of polar observing facilities and activities and portals that organises information about observing facilities and activities, but often they do not share the information. If they do, it is done in a non-consistent way with custom data structures, custom vocabularies, and typically with limited access. There are only a few standards for organising the information within this field.

The hindrances for an efficient organisation of the information are technical as well as political, social, and psychological. This document formulates recommendations that can address these hindrances.

The technical recommendations have focus on *federated search*, a concept that is already well developed within the polar data management community. Achieving a basic level of technical, syntactical, and semantic interoperability (applying the so-called *FAIR* principles), would make it easier to aggregate information across sources.

The recommendations to overcome political, social, and psychological hindrances have focus on tools that have worked in the data management world: Data policies should be formulated to also cover organising information about observing facilities and activities, as well as funding and training within them.

## 1. Background

A diverse and distributed range of existing polar observing facilities and activities could be better utilized to address societally relevant scientific problems. For example, observations and associated datasets could be integrated through synthesis in new ways, focusing on a single societal benefit area (SBAs) of interest using one or more indicators or essential variables (EVs).

The fundamental challenge related to the assessment, planning, integration, and synthesis of polar observations is that observing networks and observing systems that produce observations are diverse and distributed of nature. Fundamental questions that should be addressed include: Which research activities, networks and monitoring sites are capturing measurements necessary for any given analysis? Are there disciplinary or geographic gaps or overlaps? What are the current capabilities? How can we better plan, coordinate, and achieve objectives that rely on observing?

Users of this information could be

- An individual researcher could become more familiar with observational activities in the discipline or area of interest; be able to find space on a vessel surveying an area of interest; be able to find support staff to conduct observations at a specific location; and find community-based observer networks to partner with to conduct observations in remote areas.
- A logistics coordinator can co-locate infrastructure for cost savings.
- Indigenous organisation, local community members and governments can see which activities and types of organisations are active in the area or may have information to share.
- A science planner will be able to assess which networks are most active, as a step towards optimising and planning for polar observing.
- A funding agency will see and track which activities have been funded through their resources, find other networks with which to coordinate to share the costs or logistics for deploying or recovering autonomous sensors like moored instruments, and otherwise implement and optimise funding resources.

Box 1: High-level political statements on organising information about *observing assets*

- *A properly resourced, comprehensive effort is needed to identify strengths and gaps in the current set of systems, sensors, networks, and surveys used to observe the Arctic* (2nd Arctic Science Ministerial, from the theme on *Strengthening, Integrating and Sustaining Arctic Observations, Facilitating Access To Arctic Data, and Sharing Arctic Research Infrastructure*).
- *A knowledge map connecting these observations to societal benefits can then guide new observations, data management needs, and development of products and services, leading to a much-needed roadmap for Arctic system observing (*Arctic Observing Summit, 2018)
- *Advance system models of U.S. observing inventories and data centers to further understanding of these capacities so that informed, optimal, strategic decisions and design, and spending plans can be made* (IARPC's Arctic Research Plan: FY2017-2021, from the research goal to *Enhance Frameworks for Environmental Intelligence Gathering, Interpretation, and Application toward Decision Support*)
- *Encourage SAON to update a gap analysis of where Arctic observations are missing and recommend strategies to address priority gaps* (Joint statement of Ministers: *On the occasion of the Third Arctic Science Ministerial*)

There are many portals for research projects (like Arctic LCC, ARMAP, ASDI, BOEM, ELOKA, INTERACT GIS, ISAAFFIK, NASA ABoVE, NPDC, PDC, RiS, and USGS ScienceBase), and there are numerous portals with details about individual observing sites (e.g., AOOS Research Assets, ArcticConnect, BAID, GTN-P, JCOMMOPS, NCEI, NEON, NIPR, SIOS, and WMO WIGOS) (See appendix B). These resources exist to distribute information that goes beyond the level of data sets.

It is currently not possible to make strategic assessments across these many different sources of information. Recent efforts have been relevant (e.g., by AOV, EU-PolarNet, INTAROS, Polardex and SAON CON), but most existing inventories and portals are limited in scope, and almost none share information in a way that can be harmonised and aggregated for a necessarily comprehensive perspective.

The need for organising this information has been recognised at the highest political level (see Box 1). To summarise, there is a recognized need for a more integrated polar observing system with a means of identifying strengths and gaps, a “knowledge map” to clarify directions, and ways to build capacity to better meet observing needs.

In the context of EU-PolarNet 2, this document formulates recommendations and priorities of relevance for the European Polar Coordination Office (EPCO). The EPCO might need to organise this information so that relevant audiences can easily access it.

# 2. Scope

## 2.1 Current activities and their challenges

A number of activities already exist that organise or have organised information about polar observing facilities and activities. This include (but is not limited to) the *Arctic Observing Viewer* (AOV) and *Arctic Research Mapping Application* (ARMAP) (see Box 2), EU-PolarNet (see Box 3), *INTERACT-GIS* (see Box 4) and *Polardex* (see Box 5). Each of these have or have had challenges in accessing information sources and bringing the information into a structured framework.

Many observing facilities have been established in the polar region (research stations, moorings, flux towers, boreholes, etc.) and there are regular surveys by ships and aircrafts to the area, in addition to drifting buoys, etc. There are many networks and portals to describe these, but what is often lacking is sharing of the information with the public and with other networks and portals. If they do share the information, it is done in a non-consistent way with custom data structures, custom vocabularies, and typically with limited access. There are only a few standards for organising the information within this area.

Some sources provide machine readable end-points, but when these are not standardised, accessing the information necessitates custom programming, harmonisation and backfilling. When the sources are not machine-readable, the harvesting of information involves manual processes, often called “data wrangling” or “screen scraping”. Such processes have the risk that information is misunderstood or misinterpreted. Moreover, a database populated in this way will always be outdated since the underlying sources are dynamic. Finally, it can only be maintained by a continuous and manual effort.

In the long term and for the collation of information to be comprehensive, the envisaged procedure should be designed to automatically compile information from existing, accessible sources. The concept is already well developed within the polar data community (see section 3.1) and is based on

Box 2: Case: Applications for Arctic science planning: ARMAP and AOV.

The Arctic Research Mapping Application (ARMAP, armap.org) tracks over 3200 field-based research projects funded by the US National Science Foundation, as well as 17 other US agencies and organizations. Users can answer questions such as: “Who is doing what, when and where? How do we plan for logistics? Where are medical facilities, field research stations, ship tracks, airports, etc? How can we assess status and clarify directions to better meet objectives?”

The Arctic Observing Viewer (AOV, arcticobservingviewer.org) encompasses over 35,000 observing sites across 39 observing networks, with details and precise locations for individual monitoring assets, with links to more information. Users can answer questions like: “Where are existing observing sites? Where are more sites needed? Who operates and manages existing sites? Which sites can I use? Is there overlap?” Networks in AOV include Arctic-GRO, BGEP, BSSN, BTF, CALM, CALON, CBMP, DBO, EcoFOCI, FluxNet, GLISN, GNET, GTN-P, IABP, IASOA, ICECAPS, IMB, INTERACT, IPY, IRIS, ISAAFFIK, ITEX, LTER, NASA ABoVE, NDACC, NDBC, NEON, OASIS, O-BUOY, RUSALCA, SIZONET, Snow-Net, USCRN, WHOI-PS, YOPP. Data from these sources have been ingested and harmonised. Only a few of these networks have machine-readable endpoints and most of the information was harvested through laborious and manual processes, i.e., copied and pasted from source web pages.

*federated search* through a *federated discovery portal*. This is a point of access that allows users to perform basic searches (text, time, space) on records held in otherwise independent catalogues.

Box 3: Case: EU-PolarNet: Inventory of existing monitoring and modelling programmes (D2.3)

The inventory gave an overview of current polar region monitoring and modelling programmes. It was meant as input to a subsequent *Strategic analysis of monitoring and modelling programmes* (D2.5).

Organising the inventory documented a number of issues, including

- There was not firm definition of a 'programme' and the inclusion of activities in the inventory was to some extent subjective.
- Several countries and organisations reported that overviews did not exist or were not maintained.
- It was known that there were additional sources, but these were not accessible.

Organising the inventory meant extracting data from a number of very different sources: emails, MS Excel files, MS Word files, PDF files and 'screen scraping' from web pages.

In the short term and for the so-called 'long tail' (see section 3.4) there may be a need to develop centralised portals with specific thematic or geographic scopes that allow information providers to enter or upload information.

The hindrances for an efficient organisation of the information are technical as well as political, social, and psychological. This document focuses on formulating recommendations that can address these hindrances.

## 2.2 Technical scope

In order to address the technical hindrances for the efficient organisation of information about polar observing assets, it should be noted that the information to be collected and collated is data and that the challenges and solutions are not basically different from those addressed in the data management community. Consequently, this document recommends that the organisation of information should follow the *FAIR* principles (*Findability, Accessibility, Interoperability,* and *Reusability*) and formulates the technical recommendations aligned with them.

Achieving a basic level of technical, syntactical, and semantic interoperability, would make it easier for users and developers to aggregate and compare information across sources. It would allow the development of new applications, such as establishing a search interface that would be able to index and search contributing sources. As noted earlier, the recommendations in the document are meant to enable this so-called *federated search*.

Recommendations for addressing technical hindrances are found in chapter 3.

Box 4: Case: INTERACT GIS

INTERACT is a network of 89 terrestrial research stations across the Arctic. The INTERACT GIS system organises descriptive metadata for station operations, facilities, logistics, climate, environment, etc.

The system also provides information about variables monitored at the stations and includes a research project repository. The project repository is fed by an application module made available to INTERACT stations (currently only used by three stations) or through a project metadata template and an upload function for stations that use their own application system. The system also links to the INTERACT Data Portal where data set searches can be made across 30+ station repositories.

INTERACT GIS also maintains a registry of app. 30+ global and circum-arctic scientific networks, including information about which INTERACT stations contribute to these.

Stations are displayed on a scalable map with the opportunity to select a number of thematic background maps (e.g., climate zone, permafrost zone, vegetation, etc.). The system has filtering functions that allow scientists to search for stations complying with specific research needs (e.g., location, natural features, monitored variables etc.).

INTERACT GIS work towards the FAIR principles and has developed an endpoint that allows external systems to harvest the information.

The project description metadata model is proprietary of INTERACT.

### 2.3 Political, social, psychological scope

The hindrances for the efficient organisation of information about polar observing assets are not just technical, but also political, social, and psychological in nature. This became clear in a survey that was conducted in the context of this document, which demonstrates several concerns among the potential providers of information to the information organisation procedure. Recommendations for addressing these are found in chapter 4.

<u>Box 5: Case: Polardex</u>

Polardex is an online application for the discovery of polar research infrastructures and observing assets, as well as logistical and project plans in the Arctic and Antarctic. With Polardex, researchers and other stakeholders can explore existing infrastructures and their capabilities/facilities, supporting research planning and design of future projects. Polardex includes infrastructures such as research stations and other field facilities, vessels and aircraft in both polar regions. Integrating and further developing SOOS's DueSouth system, Polardex includes details of planned routes (scheduled cruises by research vessels, terrestrial campaigns, flights, etc.) with the intention of reducing duplication logistics and allowing researchers to identify opportunities for collaboration.

Polardex combines *assets metadata* of polar infrastructures and logistics from a range of sources, including COMNAP, SOOS's DueSouth, EUFAR, EuroFleets, INTERACT and SIOS. Such *assets metadata* is highly varied and heterogeneous, with stark differences between relevant information to significantly different infrastructures and assets. However, there are essential aspects to infrastructures, assets and logistical plans that are common to all.

Challenges in compiling theses varied *assets metadata include:*

- Each source has their own way of organising their information (database structure).
- Aligning ontologies and dictionaries (for humans and machines)
- Populating the database (in an automated way)
- Different structures and organisation for Arctic and Antarctic and different infrastructure types
- Community buy-in (users and information providers)

As described, there is currently no universally agreed standard for metadata or information describing such resources. Polardex is a clear example of an application that will greatly benefit from the development of such a standard and improved interoperability of *assets metadata*.

# 3. Recommendations to overcome Technical Hindrances.

## 3.1 Background

As described above, many sources of information exist, but only a few of these resources are “aggregators”, compiling observing-related information across multiple sources for a comprehensive perspective. This document will promote the solution that existing sources share basic information about their observing assets. What is needed are the same types of steps toward interoperability that have been taken by the broader polar data community. Existing sources should come together to agree on adoption of community-based standards for metadata, transfer protocols and other best practises towards a more comprehensive perspective of observing activities and facilities (cf. Manley *et al.*, 2015).

As was discovered in the data management world, the solution is not just one master portal or “one stop shop”, but many portals that have enabled *federated search* for inclusivity and comprehensiveness. Each will operate with their own funding, mandate, thematic or geographic scope, governance, or user audience, but will provide information that meet agreed requirements as formulated through the *FAIR* guiding principles (see section 3.3).

The proposed *federated search* mechanism still requires some technological developments, but these are expected to make considerable progress within the next few years. *Federated search* portals currently exist for research and observation data as roadmaps and demonstration projects, see for instance the World Data System (WDS) initiative[1], the SAON *Data Search Facility*[2] and the *Arctic Data Federation*[3].

This principle would mean that information from existing portals can be stored and maintained appropriately by the source organisations, but also made available to harvesting procedures from other facilities. All can have their own internal procedures for organising the information and still be valid sources for federated search. The procedure also means that regional initiatives like AOV, Polardex, SAON, SCADM and SOOS can organise information in the way they find relevant for their purposes.

While the perspective in this document in principle is European/Polar, the recommendations formulated are universal. The European perspective is addressed by noting that the described compilation and collation procedure could define a ‘window’, focusing only on European/Polar oriented sources. This will allow the definition of other ‘windows’ that can utilise the same underlying principles/structures, for instance a pan-Arctic or a pan-Antarctic ‘window’.

## 3.2 Definitions

In order to consistently organise the information on polar observing capacities and activities, some basic definitions will have to be formulated. The scope of this document is *observing capacities and activities*. No authoritative definitions exist for these, but some candidates are Cruises, Expeditions, Infrastructures, Networks, Observatories, Observing systems, Platforms, Research activities, Programs, Projects, Sensors, Sites, Stations, Transect, etc. The term <u>*observing assets*</u> is used in this context to cover observing capacities and activities.

[1] https://docs.google.com/document/d/1J2lduxDMC-Kwvj_HSMQK3e7qjuLXjpO7QQSoAoBs8CY/edit#
[2] https://saon.met.no
[3] https://search.dataone.org/portals/polderdemo and https://api.test.dataone.org/polder

As a simple example, a *station* can be several buildings including residential buildings. But it can also simply be a weather station, which usually is 'only' a collection of instruments in a simple shelter. A *site* could be instrument installations, where repeat measurements have been made, like bore holes, moorings, flux towers, vegetation plots (called *observation facilities* under WMO/WIGOS).

It should also be decided if individual *observing assets* types can relate to each other in a hierarchical, network or otherwise relational model. Some definitions of these relationships exist but should be further developed. One perspective should be that it would eventually be possible to couple the compiled information about *observing assets* to the actual data, and thereby ease the access to these.

Some work on these definitions have already been done by ADIwg (see also section 3.3.3.1) and through INTAROS[4].

In this document, information (data) about *observing assets* will be termed *assets metadata*. The term *metadata* is often used to characterise scientific or observational data as such, and therefore another terminology is used here to avoid confusion.

Recommendations:

- A definition of each of the *observing assets* types (Cruises, Expeditions, Infrastructures, Networks, etc.) should be developed.
- The nature of their relationship should be defined.
- It would be useful to prioritise among the mentioned candidates when it comes to focusing efforts to organise *assets metadata*.

## 3.3 Applying the FAIR principles

The added value of applying the *FAIR* principles in the polar community is thoroughly described by Tronstad *et al.* (2021). The principles were presented by Wilkinson *et al.* (2016) and have significantly influenced data sharing and data policy developments. The paper was motivated by a need to define 'good data management' in a sense that would facilitate discovery and use of data by assisting humans and machines in their discovery of, access to, integration, and analysis of scientific data. The *FAIR* principles put specific emphasis on enhancing the ability of machines to automatically find and use data.

The polar data community has come a long way in the last decade to better achieve interoperability. Many data repositories now publicly share metadata records (the documentation that makes individual scientific datasets findable and accessible), using interoperable metadata standards and web service protocols. This means sharing information across a fragmented landscape of entities in such a way that data discovery and integration are possible. This has documented that the *FAIR* principles are viable and could also be applied to the *assets metadata*.

The *FAIR* principles assert that data collections should be *Findable*, *Accessible*, *Interoperable*, and *Reusable*, and each of the four are translated into specific requirements to the data management system. *Findability* refers to the capacity to search for and discover data collections. *Accessibility* is a measure of the ease with which information can be directly obtained or accessed once discovered.

[4] INTAROS: Synthesis of gap analysis and exploitation of the existing Arctic observing systems. URL: https://intaros.nersc.no/sites/intaros.nersc.no/files/D2.10_INTAROS_Synthesis_v9.0.pdf

*Interoperability* is the degree to which independent datasets can be combined and integrated with one another (especially by computers), which can be facilitated by using consistent standards and vocabularies. *Reusability* means that the data can be put to multiple uses beyond its original purpose.

A side benefit of *FAIR* data is that data released from individual data centres can increasingly be fed into federated data sharing facilities, allowing for searching and aggregation, regardless of the origin of particular data. These benefits are the basis for the recommendation also to apply the *FAIR* principles for *assets metadata*.

By adopting *FAIR*, it would be easier to assess status, identify gaps, and clarify directions across the diverse, distributed, and fragmented range of existing collection of information about polar *observing assets*. Such an approach also allows for information and data to be maintained by those with the greatest domain knowledge, which are the data originators and owners.

For data to stay authentic and *FAIR* over time requires "trustworthy" repositories that will actively support and maintain the *FAIR*ness of these data. The *TRUST* principles (Lin *et al.*, 2020) (*Transparency, Responsibility, User community, Sustainability,* and *Technology*) should also apply to repositories that organise this information. Repositories must earn the trust of the communities they intend to serve and demonstrate that they are reliable and capable of appropriately managing the data they hold.

### *3.3.1 Making information Findable*

The information should be easy to find for both humans and computers, and *Findability* refers to the capacity to search for and discover data collections. This is an area that is much less mature and developed in the *observing assets* world than in the data management world. Making the information *Findable* could mean that one or more registries of existing catalogues is established. There are many actors within this area (like AOV, EPB, SAON, SCADM, and SOOS) and the registries could be jointly owned and maintained by these actors. Such systems would register those polar observing systems and other source catalogues that provide assets metadata. The result would be to improve discovery, clarify best practices, and establish a basis for integration.

<u>Recommendation</u>:

- A registry of existing observing assets catalogues should be established. It should be jointly owned and maintained by the major regional actors within the area. The registry should include specification for data transfer protocols.

### *3.3.2 Making information Accessible*

*Accessibility* is a measure of the ease with which information can be directly obtained or accessed once discovered. It is known that information about *observing assets* is already compiled and organised through many existing initiatives. Some sources make the information available through the web, while many do not. The sources that make the information available through the web often do this in a non-standardised way, and this hampers or prevents machine-access to the information. Many simply present the information as a table on a webpage or a downloadable PDF-file.

Sources should have endpoints for accessing/querying the information programmatically, as illustrated for instance by DEIMS-SDR. Such an endpoint should as a minimum implement defined *core fields* (see section 3.3.3). This is to ensure that the information is accessible and interpretable. The access to and transfer of data should be through standardised protocols.

Schema.org and the associated *Science On Schema.Org* (SOSO) are emerging frameworks, but there are other options equally viable for observing assets (see Box 6).

Recommendations:

- Sources that already organise information about observing assets must make this information web-accessible through a public-facing endpoint. In the simplest case, this could be a web folder with a spreadsheet.
- The public-facing endpoints should be using standardised transfer protocols.

### *3.3.3 Making information Interoperable*

Interoperability is the degree to which information from independent data sources can be combined and integrated with one another, which can be facilitated by using consistent data models and vocabularies. In this context interoperability is the ability of information from otherwise non-cooperating sources to be integrated in an automated way and with minimal effort. The definition ideally also includes machine interpretable information.

Only a few sources share their information in a manner that permits such integration. Many share *assets metadata* details ("fields") through custom structures, which often prevents harmonisation and integration of information. Integration efforts will often have to focus on 'mapping' elements against each other, and this often means that the integration is inefficient.

What is needed is to develop relevant metadata standards, and for some of the fields, standardised vocabularies should also be developed. Asset-level standards have been developed at the global and regional level (like CERIF, INSPIRE, ISO, WMO/WIGOS), but these are not widely known or adopted.

Various aspects of interoperability are broken down into subsections below.

#### 3.3.3.1 Asset metadata model

A set of *core fields* that describe observing assets should be identified and defined. It is acknowledged that there are some fields that are applicable for all types of *observing assets* and also some that are unique to particular kinds of assets. An *observing asset* information data model can be established on the basis of so-called *crosswalks*. In a *crosswalk,* different existing metadata models are compared and fields that are identical or almost identical are identified.

Work is ongoing within the Polar Observing Assets WG[5] (POAwg, under the SAON Committee on Observations and Networks (CON)) to create a crosswalk for *observing systems*[6] and *observing sites*[7]. This work documents, that established systems often only share a few fields, but also that crosswalks are important for establishing joint standards (See figure 1). The outcome is a selection of a few, common fields; these are often text-based, like title and URL. Similarly, and within the 'Polar to Global Online Interoperability and Data Sharing' initiative (P2G), crosswalks are created mainly across inventories of logistical information in the polar regions[8].

The Alaska Data Integration Working Group (ADIwg) has developed a Project Metadata Standard. The US Fish and Wildlife Service's Science Applications Program, the US Geological Survey and the US Bureau of Ocean Energy Management have continued to work in partnership to utilise this work to describe projects and datasets. The USGS Alaska Science Center has also been working to include descriptions of collections of sampling sites. The Arctic Research Mapping Application (ARMAP) and Arctic Observing Viewer (AOV) have adopted the ADIwg specification and provide access to information for NSF projects.

Work is ongoing also outside the polar region to standardise within this area. Within ENVRI-FAIR, a use case was "Site Documentation Interoperability". It investigated ways to increase interoperability of site documentation in the ecosystem domain by:

- Identifying fields that are common across (ecosystem) research infrastructures (RI)
- Agreeing on a set of fields to be harvested from each RI
- Collecting and making that information available through a centralised search interface.

Box 6: Schema.org (SDO)

Is a collaborative community activity with a mission to create, maintain and promote schemas for structured data on the internet, on web pages, in email messages and beyond.

A schema is a set of 'types', each associated with a set of properties. The types are arranged in a hierarchy. As of January 2021, the vocabulary currently consists of '778 types',

SDO is a 'discovery language' and allows a community to collectively agree on some core principles of best practice before too many data centres have implemented it.

Within the scientific community, a collaborative process through *Earth Science Information Partners* (ESIP) – Science on Schema.Org (SOSO) – has provided guidance on implementation of schema.org for scientific datasets, data repositories, and more. URL: https://github.com/ESIPFed/science-on-schema.org

Specifically, within the polar science community, additional guidance has been formulated by the Working Group on federated search (POLDER). URL: docs.google.com/document/d/1r4OSRuVBfdJpMbyhjkghHSeckraFEhxs0f1ld4aGkg/edit#

The work is documented by Wohner *et al.* (2020), who investigated the interoperability across six major sources (mainly with a terrestrial theme: DEIMS-SDR, ICOS, INTERACT, NEMSR, NEON, and SIOS) that organise information about research sites. The work conclude that they only share five elements (Figure 2, upper table) and that only these few can be integrated with few semantic issues. The selected five fields (name, identification, contact, textual description, location, observed properties) can be mapped to global standards (Figure 2, lower table). In a separate study (Wohner *et al*., 2021), that builds on existing site information, it has been possible to evaluate the value of adding a new observation site to a geographical region, which is one of the described benefits of organising this information (section 1).

Recommendations:

- Where these not already exists, asset metadata standards should be developed and applied to existing sources. New sources should organise information in agreement with these. Existing standards should be used to the extent possible.
- Adopting *core fields* would guarantee a basic semantic level of interoperability between catalogues. It is also important to emphasise that metadata models should not be restricted to these core fields as the fundamental idea is to allow extensions based on needs.
- Crosswalks for harmonisation across asset-level metadata standards (and other well-documented implementations) should be developed and shared.

#### 3.3.3.2 Ontologies and code lists

An ontology defines a common vocabulary and is useful when there is a need to share information in a particular domain.

The creation of controlled domain-specific vocabularies is labour intensive and requires collaboration among stakeholders to reach consensus on which vocabularies should be included, i.e., which terms to actually use. For *observing assets*, there is a large diversity among the entities that should be organised. Inventing new vocabularies if they already exist should be avoided.

Best practise is to re-use terms from well-established and well-served vocabularies. For example, the World Meteorological Organization (WMO) has a controlled vocabulary that the ocean and polar communities typically use called the *International Meteorological Vocabulary*. The WIGOS metadata standard should be used whenever possible, e.g., the site type or the site operation status, or mappings to WIGOS code lists should be provided.

The INSPIRE Directive aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. Under the Directive, a *Data Specification on Environmental Monitoring Facilities* has been developed[9]

Other vocabularies that could be used for this are the GEneral Multilingual Environmental Thesaurus (GEMET) developed by the European Environment Agency (EEA) and the European Environment Information and Observation Network (Eionet).

[9] https://inspire.ec.europa.eu/Themes/120/2892

Recommendations:
- Best practice is to re-use terms from well-established and well-served vocabularies.
- Controlled vocabularies should be developed (only) where necessary and mapped to existing sources. New sources should organise information in agreement with these.

### *3.3.4 Making information Reusable*

One of the main goals of the *FAIR* principle is to optimise the reuse of information with the basic underlying principle, that data should be recorded once, but be able to be used many times. Reusability means that the data can be put to multiple uses beyond its original purpose.

One aspect of this principle is that the organised information should be associated with relevant international licenses. Such licenses describe how users are allowed to copy, modify and distribute the information in any format for any purpose. Users are obligated to give appropriate credit (attribution) and indicate if they have made any changes, including translations. An example of such a licence widely used in the data management world is *The Creative Commons Attribution International* license[10].

Another requirement (also related to the *Findable* principle) is to unambiguously identify each *observing asset*. In this context, where many existing sources already organise the information, there is a risk that different sources store the same information, and an integration effort will result in duplicates. One way of solving this would be to define and apply *persistent unique identifiers* to each *observing asset*. The use of identifiers will allow 'clean-up' of duplicates during the integration process. It should be noted that the integration process will still work, even if identifiers are not assigned, but the outcome will be less satisfactory.

DEIMS-SDR issues identifiers for sites (the DEIMS.ID), and these have already been used in other applications and for EU reporting. There will be many questions associated with the governance of such identifiers, especially about how to assign them. There are global systems like DOI or ORCID, originally developed for other purposes, but they may be applicable to *observing assets*.

Recommendations:
- The facilities that organise *metadata assets* should associate their holdings with relevant license statements.
- Persistent unique identifiers should be defined and applied to *observing assets*.

### *3.4 Organising assets metadata for 'the long tail'*

The *observing assets* sources described are usually hosted by large institutions or initiatives. Institutions or countries where polar research and observations only has a minor focus ('the long tail') may need a 'home' that will store *observing metadata* on their behalf. In the short term it may be necessary to establish relevant web-entry or data upload services with specific geographical or thematic scope. It should be noted that data facilities populated in this way will always be outdated since the underlying information is dynamic. They can only be maintained by a continuous and manual effort.

[10] https://creativecommons.org/licenses/by/4.0/

Recommendations:

- The major regional institutions (like AOV, EPCO, SAON, SCADM, and SOOS) could develop and maintain relevant web-entry or data upload services within their geographical and thematic scope. These services should observe the recommendations formulated in this document.
- These institutions should also guide the small initiatives in organising the information through for instance spreadsheets in web-accessible folders (see above).

## 4. Political, social, psychological hindrances. Recommendations to overcome these.

### 4.1 Survey

In summer 2021, a survey was organised among a broad group of institutions and individuals engaged in polar observing. The purpose of the survey was 'to seek to create an overview of existing sources of information on polar observing facilities, systems and activities'. The scope of the survey was also to identify non-technical obstacles to the organisation of information about *observing assets*.

There was a total of 89 respondents, mainly from Europe (63%) and North America (29%), but also a few from Asia (7%) and Oceania (1%). Of the 89 respondents, 52 were managing observation data, 53 were producing observation data, and 55 were using/consuming observation data (more than one reply was possible). The geographical interest/scope of the respondents' work were Antarctic (7%), Arctic (43%) or polar (49%). The outcome of the survey can be found in appendix D, including tables referenced below.

### 4.2 Restrictions on access. Regulation

In the survey, 17 out of 71 respondents indicated that there are restrictions on the access to the information about *observing assets* in their facilities (table D.1). A summary of the reasons for restrictions is (table D.2):

- In the simplest cases, information is available upon registration.
- The information is embargoed for a certain period. This is either because there is a processing time or because there is a wish to publish the information elsewhere before it is made publicly available.
- The information can be made available upon request. In some cases, a permission has to be given; it is not clear what a permission entails.
- The information is available as summaries, while detailed information is not.
- In one case, access can be granted after "contacting the PI to assess possibilities for collaboration".

Approximately half of 15 respondents that report on restrictions on access respond that they consider relaxing these now or in the future (table D.3)

### 4.3 Restrictions on access. Resource limitations

For the respondents that limit access to information about *observing assets*, the survey explores potential motivations that could change this. Most are related to resources (table D.4):

- Capacity seems to be an issue, both in terms of the lack of funding but also from the perspective of the researcher, who finds that it is not the responsibility of the individual researcher.
- Related to the first item is the lack of interaction with relevant IT resources. These resources seem to be scarce, and there is a lack funding to support their effort.
- Finally, one respondent notes that the polar research is a minor part of the research effort at the institution.

### 4.4 Recommendations

In the data management world, there is currently intensive work ongoing to make polar observational data *FAIR* and to establish the infrastructures that allow *federated search* access to these. The organisation of *assets metadata* is lacking behind this development, and the recommendations to address this are in many ways similar to recommendations from the data world.

#### *4.4.1 Regulation*

Different levels of regulation exist within this area and are an important factor. In the survey, one of the respondents proposed that mandates are developed within this area (table D.5). In the Antarctica, the organisation of this information is to some extent regulated by the *Antarctic Treaty System* (ATS), and organisations like SCADM and SOOS advocate for the importance of organising this. In the Arctic, the Arctic Council (AC) recently (2021) '(…) welcome progress on implementing guiding principles on management of and access to data (…)', and AC associated initiatives like IASC, ASDI, and SAON promote the principles of free and open access to data. Recently (2021), the mentioned Antarctic and Arctic initiatives in cooperation with others developed principles for the formulation of polar data policies (Tronstad *et al.*, 2021). While such policies traditionally have focused on observational data and their metadata, new policies should be formulated on the organisation of *assets metadata*.

<u>Recommendations</u>:

- Relevant data policies should be formulated at national and regional level that also regulate the organisation of *assets metadata*. This could also be formulated in terms of nations meeting international obligations.

#### *4.4.2 Funding and capacity*

In the data management world, progress toward public release of datasets and associated metadata has been mandated via the funding of data repositories, the requirement of data management plans and in announcements of opportunities. No such incentives or guidance has been applied for the organisation of *assets metadata*.

When it comes to funding and capacity, the management of *assets metadata* is often not prioritised, either because it is not seen as a core part of a particular program/project or because of lack of capacity. It may not necessarily be the scientist/expert itself that does the work to organise the information, but research institutions should have relevant data management resources to do the work.

<u>Recommendations</u>:

- Within the data management world there are requirements that a certain portion of funding should be allocated to data management, and a similar principle should apply to the organisation of assets metadata.
- Researcher often do not have access to relevant IT competency. It should be recognised that the management of assets metadata requires skilled and funded staff at the institutions.

#### *4.4.3 Indirect incentives*

Like in the data management world, there are more indirect reasons for organising information about *assets metadata*. These could include

- Documenting for institutions and their funders the impact of assets, also if it can be linked to the data that have been produced by the assets.
- Documenting that assets are fully utilised and that their use is coordinated with other initiatives. This includes documenting cost savings.

- Raising awareness of working towards common scientific goals in a broader community. If this includes taking advantage of an established facility or approach, this is saving time and effort.
- Nations and institutions can have a stronger voice at the international level if they can document their *observing assets*. Joseph *et al.* (2019) describes how several nations have expressed an interest in this.
- For managers of *observing assets* there can be an interest in recording information in acknowledged systems. An example is that when documenting an observing site in DEIMS, a tool examines the completeness of the information provided, scores this, and this can serve as motivation.

Recommendation:

- The value of organising *observing assets metadata* should be promoted in relevant fora, putting more emphasis on more indirect advantages of this: Saving resources by better coordination and utilisation of existing assets and being able to raise the international profile of nations and institutions.

## 5. Other considerations

As within the data management world there are certain considerations associated with the organisation of *assets metadata*, when these are closely associated with human beings, their health, and their culture. While this is a recurring consideration within the social sciences and humanities in the management of observational data, this is probably less of an issue within this context. An exception is that the work on the basic definitions outlined in section 3.2 should have in mind that terms like 'expedition' and 'exploration' should be avoided in inhabited areas, while it may be more relevant in the Antarctic and the open ocean. Terms like 'field campaign', 'field work' or 'planned route' are more appropriate.

When it comes to indigenous people, the *CARE Principles for Indigenous Data Governance* by Carroll *et al.* (2020), reflect the role of data in advancing innovation, governance, and self-determination among indigenous peoples and should be taken into consideration, when relevant. For the Arctic area there are specific documents that guide engagement like the *National Inuit Strategy on Research* (2018) and *ICC Ethical and Equitable Engagement Synthesis Report* (2021). As for the social sciences and humanities, the organisation of *assets metadata* is probably less sensitive than in the data management world.

In some cases, like for the USA, this is a regulated area: The Geospatial Data Act (2018) on open access to data does not include "geospatial data and activities of an Indian tribe (…), as determined by the tribal government". Similarly, sensitive information could be about, for example, the geographical position and naming of locations of cultural importance.

Recommendations:

- In the development of the basic definitions, care should be taken to engage social scientists and Arctic community members to help enhance the terminology.
- When it comes to organisation of assets metadata related to human beings, their health, and their culture, the legislation formulated nationally, and by the indigenous people and their organisations, should be considered.

## 6. Acknowledgements

The EU-PolarNet D6.1 team would like to thank William Manley (University of Colorado, USA), Pip Bricher (Southern Ocean Observing System (SOOS)) and Øystein Godøy (Norwegian Meteorological Institute) for valuable discussions and input. The *Polar Observing Assets Group* (POAwg) and participants at the *Interoperability for polar observing networks and logistical resources - Hackathon* at the *Polar Data Forum IV* are also thanked for their help and input. Roberta Pirazzini (FMI, Finland), Elmer Topp-Jørgensen (University of Aarhus, Denmark), and Christoph Wohner (Environment Agency, Austria) are thanked for their valuable review comments.

## 7. References


- Arctic Council. 2021. Reykjavik Declaration. https://oaarchive.arctic-council.org/handle/11374/2600
- Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. 2020. The CARE Principles for Indigenous Data Governance. Data Science Journal, 19(1), p.43. DOI: http://doi.org/10.5334/dsj-2020-043
- ENVRI-FAIR: Report On Fairness Implementation Activities In The Biodiversity And Ecosystem Subdomain. 2021. https://envri.eu/wp-content/uploads/2021/04/ENVRI-FAIR_D_11-2.pdf
- EU-PolarNet. 2016. Inventory of existing monitoring and modelling programmes (D2.3). https://eu-polarnet.eu/wp-content/uploads/2020/11/D2_3_Inventory_of_existing_monitoring_and_modelling_programmes.pdf
- Federal Geographic Data Committee (FGDC) Geospatial Data Act of 2018. https://fgdc.gov/gda
- ICC Ethical and Equitable Engagement Synthesis Report. 2021. https://www.inuitcircumpolar.com/project/icc-ethical-and-equitable-engagement-synthesis-report/
- Joseph, H., Loseto, L. and Larsen, J. R. 2020. SAON – Committee on Observations and Networks: Summary Report of State of Readiness on Arctic Observations. http://arcticobservingsummit.org/wp-content/uploads/2021/06/AOS2020_white_paper_short_statement_036.pdf
- Lin, D., Crabtree, J., Dillo, I. et al. 2020. The TRUST Principles for digital repositories. Sci Data 7, 144. https://doi.org/10.1038/s41597-020-0486-7
- Manley, W.F., Gaylord, A.G., Kassin, A., Cody, R., Copenhaver, W.A., Dover, M., Escarzaga, S.M., Font, R., Garcia, A.E., Lin, D.H., Score, R., Villarreal, S., and Tweedie, C.E. 2015. The U.S. Arctic Observing Viewer: A Web-Mapping Application for Enhancing Environmental Observation of the Changing Arctic: Arctic, v. 69, , no. 5, 11 pp. http://dx.doi.org/10.14430/arctic4477
- Manley, W.F., William Gaylord, A., G., and Tweedie, C. E. 2020. Optimizing Arctic Observing Through Interoperable Information Sharing Across Networks. http://arcticobservingsummit.org/wp-content/uploads/2021/06/AOS2020_white_paper_short_statement_022.pdf
- National Inuit Strategy on Research. 2018. https://www.itk.ca/wp-content/uploads/2018/04/ITK_NISR-Report_English_low_res.pdf
- Tronstad, S., Bricher, P., Kool, J., Pulsifer, P., Van de Putte, A., Larsen, J. R., Peat, H., Rayner, D., De Bruin, T., Tacoma, M., Thomas, J., Nitsche, F., Samy, VS, Ignatiuk, D., Buch, E., Treasure, A., Persäter, F., Riopel, S., Pouplier, P., and Eicken, H. 2021. Alignment of Polar Data Policies - Recommended Principles. https://zenodo.org/record/5734900#.YaTXi7qUlhE
- Wilkinson, M. D, Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., B. Mons, B. 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data, 3 (2016), Article 160018, 10.1038/sdata.2016.18
- Wohner, C., Peterseil, J., Genazzio, M. A., Guru, S., Hugo, W., and Klug, H. 2020. Towards interoperable research site documentation – Recommendations for information models and data provision. Ecological Informatics, 60, 101158. https://doi.org/10.1016/j.ecoinf.2020.101158
- Wohner, C., Ohnemus, T., Zacharias, S., Mollenhauer, H., Ellis, E. C., Klug, H., Shibata, H., & Mirtl, M. 2021. Assessing the biogeographical and socio-ecological representativeness of the ILTER site network. In Ecological Indicators (Vol. 127, p. 107785). Elsevier BV. https://doi.org/10.1016/j.ecolind.2021.107785

## Appendix A. Relevant web pages

- Alaska Data Integration Working Group (ADIWG): https://www.adiwg.org
- Alaska Ocean Observing System (AOOS) Research Assets Map: https://portal.aoos.org/old/research-assets.php
- ArcticConnect: https://arcticconnect.org/services
- Arctic Inventory & Monitoring Network: https://www.nps.gov/im/arcn/index.htm
- Arctic Landscape Conservation Cooperative (Arctic LCC): http://arcticlcc.org
- Arctic Observing Viewer (AOV): https://arcticobservingviewer.org
- Arctic Research Mapping Application (ARMAP): https://armap.org
- Arctic Spatial Data Infrastructure (ASDI): https://arctic-sdi.org
- Atlas of Community-Based Monitoring in a Changing Arctic: http://www.arcticcbm.org
- Common European Research Information Format (CERIF): https://eurocris.org/services/main-features-cerif
- Dynamic Ecological Information Management System - Site and dataset registry (DEIMS-SDR) https://deims.org/
- ENVRI-FAIR: https://envri.eu/home-envri-fair/
- GCMD Keywords, 2020: https://earthdata.nasa.gov/earth-observation-data/find-data
- /gcmd/gcmd-keywords
- ISAAFFIK: https://www.isaaffik.org/
- INTERACT GIS: https://interact-gis.org/
- Local Environmental Observer (LEO) Network: www.leonetwork.org
- NASA ABoVE: https://above.nasa.gov
- Netherlands Polar Data Center: https://npdc.nl
- Polar Data Catalogue: https://polardata.ca
- Polar Observing Assets Working Group: https://www.polarobservingassets.org
- Research Data Alliance (RDA): https://www.rd-alliance.org
- Standing Committee on Antarctic Data Management (SCADM): https://www.scar.org/resources/scadm/overview/
- Scientific Committee on Antarctic Research (SCAR): https://www.scar.org/resources/scadm/overview/
- Southern Ocean Observing System (SOOS): http://www.soos.aq
- Svalbard Integrated Arctic Earth Observing System (SIOS): https://sios-svalbard.org
- Sustaining Arctic Observing Network (SAON): www.arcticobserving.org
- The Polar Citizen Science Collective: https://www.polarcollective.org

## Appendix B. Existing polar observing asset portals

Each of these are defined by their thematic or geographical scope or their funding structure. Some are not solely polar but has polar contents. Some have a machine-readable endpoint (API); this is indicated in the URL-column. The list is not comprehensive.

| | Metadata Standards | Transfer Protocols | URL | Theme |
|---|---|---|---|---|
| Arctic Landscape Conservation Cooperative (Arctic LCC) | ISO 19115-1, FGDC CSDGM | OAI-PMH, OGC CSW, REST, custom protocols | http://arcticlcc.org | Projects |
| AMAP Project Directory | | | http://projects.amap.no/directory/amap | Projects |
| Mapping Arctic in situ observing systems (ARCMAP) | | | https://arcmap.nersc.no/#ac_3575/2/90.0/0.0 | |
| Arctic Research Mapping Application (ARMAP) | ISO 19115-1, ISO 19115-2, FGDC, ADIwg | OGC WMS, OGC WFS, WAF, TXT, KMZ, ArcGIS Feature Service, REST | https://armap.org/web-services/ (API)<br>https://api.battellearcticgateway.org/v1#/ARMAP/get_projects_v1_armap_projects_get (API) | Projects |
| Arctic Spatial Data Infrastructure (ASDI) | ISO 19115 | OGC CSW and other OGC and ISO standards | https://arctic-sdi.org | Projects |
| Atlas of Community-Based Monitoring in a Changing Arctic | | | http://www.arcticcbm.org/index.html | Projects |
| AOOS | | | https://portal.aoos.org/old/research-assets.php | Observing sites |

| | | | | |
|---|---|---|---|---|
| Arctic Observing Viewer (AOV) | | | https://arcticobservingviewer.org/web-services (API) | Observing sites |
| Barrow Area Information Database (BAID) Research Sites Viewer | | | www.barrowmapped.org | Observing sites |
| The Council of Managers of National Antarctic Programs (COMNAP) | | | https://www.comnap.aq/antarctic-facilities-information | Stations, facilities, including vessels |
| Dynamic Ecological Information Management System: site & dataset registry (DEIMS-SDR) | | | https://deims.org | Observing sites |
| DueSouth | | | https://data.aad.gov.au/duesouth/ | |
| ELOKA | | | https://eloka-arctic.org/products | Projects |
| ENVINET Activities Catalogue | | | http://projects.amap.no/directory/envinet/ | |
| GTN-P | | | https://gtnp.arcticportal.org/ | Observing sites |
| ICES Cruise Summary Report (CSR) | | | https://ocean.ices.dk/csr/ | |
| INTAROS | | | http://intaros.eu/ | Observing sites, Projects |
| INTERACT | | | https://dataportal.eu-interact.org/api/3 (API) | Projects |
| ISAAFFIK | | | https://isaaffik.org | Observing sites, Projects |

| | | | | |
|---|---|---|---|---|
| JCOMMOPS | | | https://ioc.unesco.org/our-work/joint-centre-oceanographic-and-marine-meteorological-observing-programme-support-jcommops | |
| NASA ABoVE | DIF-9, DIF-10, ECHO-10, SERF, ISO 19115 | | https://above.nasa.gov/ | Observing sites, Projects |
| NCEI | | | https://www.ncei.noaa.gov/ | Observing sites |
| NSF NEON - National Ecological Observatory Network | | | https://www.neonscience.org | Observing sites |
| Netherlands Polar Data Center (NPDC) | DIF-10 | WAF | https://npdc.nl/ | Projects |
| NIPR | | | https://www.nipr.ac.jp/english/ | Observing sites |
| OceanOPS | | | https://public.wmo.int/en/our-mandate/how-we-do-it/technical-commissions/joint-wmo-ioc-commission-oceanography-and-marine-meteorology-jcomm | Observing sites |
| Polar Collective | | | https://polarcollective.org/projects/ | Projects |
| Polar Data Catalogue (PDC) | ISO 19115 (North American Profile), FGDC | OAI-PMH, OGC CSW | https://www.polardata.ca | Projects |
| Polardex | | | | Logistics and infrastructures (ships, stations, aircrafts) |
| PDC | | | https://www.polardata.ca/ | Projects |
| RiS | | | https://researchinsvalbard.no | Projects |

| | | | | |
|---|---|---|---|---|
| SAON project inventory | | | http://projects.amap.no/directory/saon/ | Projects |
| SeaDataNet | | | http://seadata.bsh.de/Cgi-csr/retrieve_sdn2/start_sdn2.pl | |
| Study of Environmental Arctic Change (SEARCH) | | | http://projects.amap.no/directory/search/ | |
| SIOS | | | XML: https://sios-svalbard.org/sios-ri-catalogue.xml?attach=page (API)<br>CSV: https://sios-svalbard.org/sios-ri-catalogue.csv?attach=page (API) | Observing sites, Projects |
| USGS ScienceBase | ISO 19139, custom | OGC CSW, DataOne API | https://www.sciencebase.gov/catalog/ | Projects |
| WMO WIGOS OSCAR | | | https://oscar.wmo.int/surface/#/ | Observing sites |

## Appendix C. Abbreviations

| Abbreviation | |
|---|---|
| ADIwg | Alaska Data Integration Working Group |
| AOOS | Alaska Ocean Observing System |
| ARMAP | Arctic Research Mapping Application |
| ASDI | Arctic Spatial Data Infrastructure |
| AOV | Arctic Observing Viewer |
| BAID | Barrow Area Information Database |
| BOEM | Bureau of Ocean Energy Management |
| COMNAP | The Council of Managers of National Antarctic Programs |
| ELOKA | Exchange for Local Observations and Knowledge of the Arctic |
| EPCO | European Polar Coordination Office |
| EUFAR | The European Facility for Airborne Research |
| GTN-P | Global Terrestrial Network for Permafrost |
| ISAAFIIK | Isaaffik is the Greenlandic word for gateway |
| INTAROS | Integrated Arctic Observation System |
| INTERACT | International Network for Terrestrial Research and Monitoring in the Arctic |
| JCOMMOPS | Joint Centre for Oceanographic and Marine Meteorological Observing Programme Support |
| NASA ABoVE | National Aeronautics and Space Administration / Arctic-Boreal Vulnerability Experient |
| NCEI | National Centers for Environmental Information |
| NEON | National Ecological Observatory Network (NSF) |
| NIPR | National Institute of Polar Research (Japan) |
| NPDC | Netherlands Polar Data Center |
| PDC | Polar Data Catalogue |
| OceanOPS | The WMO-IOC Joint Technical Commission for Oceanography and Marine |

| Abbreviation | |
|---|---|
| | Meteorology in situ Ocean Observing System Monitoring and Coordination Centre |
| P2G | Polar to Global Online Interoperability and Data Sharing initiative |
| POAwg | Polar Observing Assets Working Group, a working group under SAON CON |
| POLDER | Polar Federated Search Working Group |
| RiS | Research in Svalbard |
| SAON | Sustaining Arctic Observing Networks |
| SAON CON | Committee on Observing and Networks. A committee under SAON |
| SDO | Schema.org |
| SOSO | Science On Schema.Org |
| SIOS | Svalbard Integrated Arctic Earth Observing System |
| WMO | World Meteorological Organization |
| USGS | US Geological Service |

## Appendix D Survey on polar observing assets

In summer 2021, a survey was organised among a broad group of institutions and individuals engaged in polar observing. There was a total of 89 respondents, and some of the responses are found below. The comprehensive report is available[11].

Even though the guidance to the survey explained that the survey was about *observing assets*, there is reason to believe that some of the responses are more about the organisation of observation data rather than about *assets metadata*.

**D1. Is the information stored in this/these system(s) accessible for people outside your institution?**

| | Frequency |
|---|---|
| Openly | 52 |
| There are restrictions on access | 17 |
| Missing = 20 | |

**D2. If restrictions on the access apply, please specify**

| |
|---|
| Access requires email address |
| Accessible via NATO documents |
| As open as we can without having a general structure |
| Data are restricted only during preparation for public access |
| Data is open after publication |
| Industry permission sometimes required |
| Information not yet accessible to people outside |
| Most open, but can be restrictions due to embargoes etc. |
| Note: I did not select restrictions above, but they are currently in development and not available, yet. |
| Only available internally, but external access is possible through SIOS and WIGOS (OSCAR).. |
| Part of the data can be accessible upon request |
| Raw telemetry |
| Registration |
| Some data sets require contacting the PI to assess possibilities for collaboration |
| Summary and statistical information is open; details (including contact information) is restricted. |
| This is a longer conversation between government in exchange of data |
| Time limited embargo for a few data sets |
| Endpoints are internal only to developers |
| Must partner with or contact our org for data permissions |
| Non commercial and scientific purposes only |
| Simple registration, then open for scientific use |
| Some data are not part of an open licence |

11 https://www.arcticobserving.org/news/451-survey-on-polar-observing-assets-reports

**D3. If restrictions on the access apply, is your institution currently considering or already implementing developments that will relax the restrictions?**

| |
|---|
| Accessible via NATO documents |
| No (x2) |
| Not applicable |
| Not discussed currently, more focus on feeding WIGOS. |
| Not yet |
| Restriction is very relaxed already |
| We aim for data to be as open as possible |
| Yes (x2) |
| Yes, once preparation is complete, data are made public. |
| Yes. Publishing the results faster in data journals |
| In the future |
| Unclear |
| We are revising out open access policy, our restricted data licences and always endeavour to make as much data openly accessible as possible but we are also bound by the data providers to respect their wishes on the use of these data i.e. economic data may be restricted |

**D4. If your institution does not already organise or make the information available, would your institution have the capacity and be interested in making this information available to externals?**

| | Frequency |
|---|---|
| No | 7 |
| Yes | 40 |
| Missing = 42 | |

**D5. If you answered 'No' to the first question in this section, what could motivate you or your institution?**

| |
|---|
| Capacity to do so. |
| More researchers doing polar research at my institution |
| Quicker and less frustrating access to IT help personell at the university! |
| The question should also have a maybe. But once again, this requires more than a simple google form to discuss access to datasets. |
| We would need a mandate from the European Commission. |
| Access to resources to develop end-user applications |
| Maintaining is costly, as a univ. our activities are project based. We cannot sustain suhc service for long wthout third party funds |

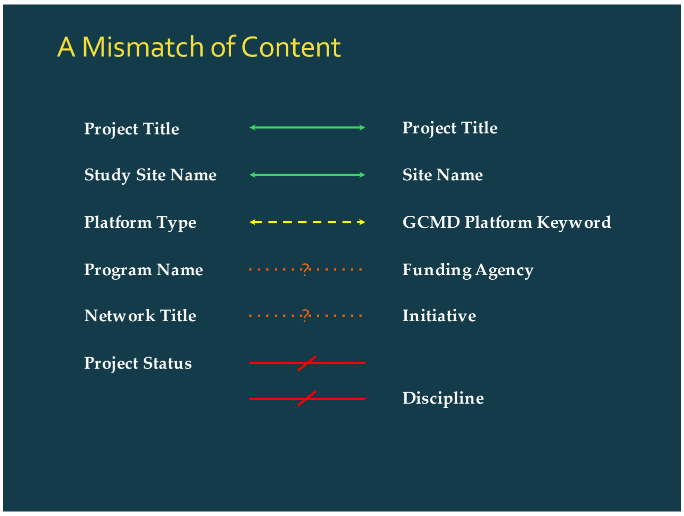
**D6. Would you be willing to update the information provided regularly, say on a yearly basis?**

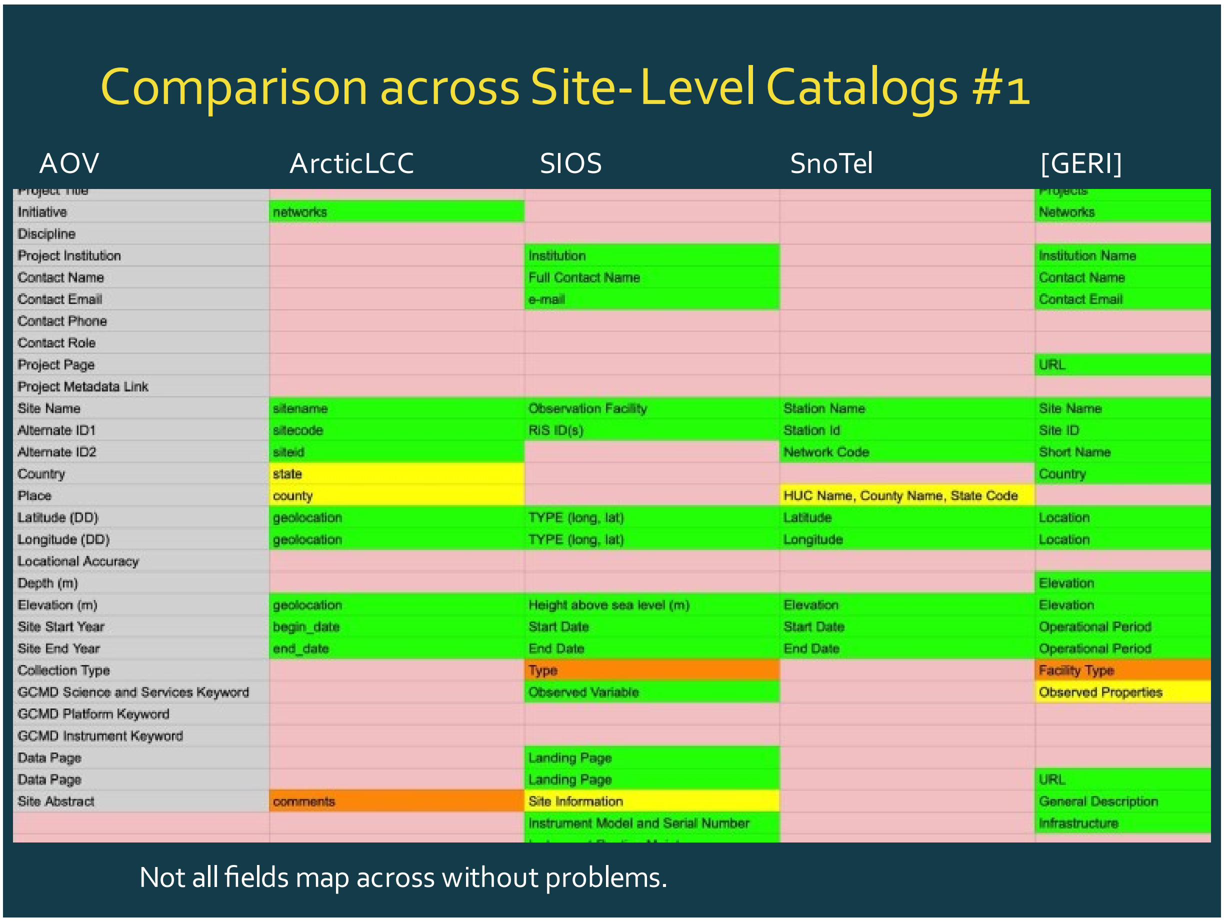
| | Frequency |
|---|---|
| Maybe | 22 |
| No | 1 |
| Yes | 41 |
| Missing = 25 | |

## Figure 1: A comparison of existing observing assets catalogues shows the need for standardisation

Courtesy William Manley (University of Colorado, USA)




| AOV | ArcticLCC | SIOS | SnoTel | [GERI] |
|---|---|---|---|---|
| Project Title | | | | Projects |
| Initiative | networks | | | Networks |
| Discipline | | | | |
| Project Institution | | Institution | | Institution Name |
| Contact Name | | Full Contact Name | | Contact Name |
| Contact Email | | e-mail | | Contact Email |
| Contact Phone | | | | |
| Contact Role | | | | |
| Project Page | | | | URL |
| Project Metadata Link | | | | |
| Site Name | sitename | Observation Facility | Station Name | Site Name |
| Alternate ID1 | sitecode | RiS ID(s) | Station Id | Site ID |
| Alternate ID2 | siteid | | Network Code | Short Name |
| Country | state | | | Country |
| Place | county | | HUC Name, County Name, State Code | |
| Latitude (DD) | geolocation | TYPE (long, lat) | Latitude | Location |
| Longitude (DD) | geolocation | TYPE (long, lat) | Longitude | Location |
| Locational Accuracy | | | | |
| Depth (m) | | | | Elevation |
| Elevation (m) | geolocation | Height above sea level (m) | Elevation | Elevation |
| Site Start Year | begin_date | Start Date | Start Date | Operational Period |
| Site End Year | end_date | End Date | End Date | Operational Period |
| Collection Type | | Type | | Facility Type |
| GCMD Science and Services Keyword | | Observed Variable | | Observed Properties |
| GCMD Platform Keyword | | | | |
| GCMD Instrument Keyword | | | | |
| Data Page | | Landing Page | | |
| Data Page | | Landing Page | | URL |
| Site Abstract | comments | Site Information | | General Description |
| | | Instrument Model and Serial Number | | Infrastructure |

## Figure 2: Investigating the interoperability across six major sources that organise information about research sites.

The work conclude that they only share five elements and that only these few can be integrated (upper). The latter are mapped to existing standards (lower). From Wohner *et al.* (2020) (excerpt).

Existing fields across catalogues.

| Category | Field | DEIMS-SDR | ICOS | INTERACT | NEMSR | NEON | SIOS | Frequency |
|---|---|---|---|---|---|---|---|---|
| Affiliation | Participation in networks/Ris/ projects | ✓ | only ICOS | only INTERACT | ✓ | only NEON | always SIOS | 2 |
| Data | Available data services | ✓ | × | × | × | × | × | 1 |
| Environmental Characteristics | Geology | ✓ | × | × | × | ✓ | × | 2 |
| Environmental Characteristics | Dominant Phenology Species/general vegetation description | ✓ | × | × | × | ✓ | × | 2 |
| Environmental Characteristics | Ecosystem Classification/Biome/ NLCD | ✓ | × | × | × | ✓ | × | 2 |
| Environmental Characteristics | Climate | ✓ | × | ✓ | × | ✓ | × | 3 |
| Environmental | Mean annual temperature | ✓ | × | ✓ | × | ✓ | × | 3 |

Mapping of core fields to metadata standards.

| Core field | Dublin core | ISO 19115/19139 | WIGOS | INSPIRE EF |
|---|---|---|---|---|
| siteName | Title | Dataset title | Station/platform name | Name |
| Siteid | Identifier | fileIdentifier | Station/platform unique identifier | Inspireididentifier |
| Location | Coverage | Geographic location of the dataset (by four coordinates or by geographic identifier) | Geospatial location | Geometry/representativePoint |
| Description | Description | Abstract defining the dataset | | Additionaldescription |
| Contact | Contributor/ Creator | Metadata point of contact | | Responsibleparty |
| Observedproperties | Subject | descriptive Keywords | Observed variable – measurand | ObservingCapability (observedProperty) |