HORIZON 2020
Coordination and Support Action
Grant Agreement No: 652641

EU-PolarNet
CONNECTING SCIENCE WITH SOCIETY

Deliverable No. D3.8
White paper on European polar data accessibility
## Submission of Deliverable

<table>
<thead>
<tr>
<th>Work Package</th>
<th>WP 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliverable no. &amp; title</td>
<td>D3.8 White paper on European polar data accessibility</td>
</tr>
<tr>
<td>Version</td>
<td>Final</td>
</tr>
<tr>
<td>Creation Date</td>
<td>21.11.2019</td>
</tr>
<tr>
<td>Last change</td>
<td>16.03.2020</td>
</tr>
<tr>
<td>Status</td>
<td>WP lead accepted, Executive Board accepted</td>
</tr>
<tr>
<td>Dissemination level</td>
<td>PU-Public</td>
</tr>
<tr>
<td>Lead Beneficiaries</td>
<td>AMAP (partner 20) and RBINS (partner 17)</td>
</tr>
<tr>
<td>Due date</td>
<td>30.11.2019</td>
</tr>
<tr>
<td>Delivery date</td>
<td>18.03.2020</td>
</tr>
</tbody>
</table>
White Paper on European polar data accessibility

Authors: Jan-René Larsen, Serge Scory and Maxime Sweetlove

Contributors:
Sabrina Delgado Arias, David Arthurs, Taco De Bruin, Bente Lilja Bye, Ranjan Datta, Piotr Glowacki, Øystein Godoy, Peter Pulsifer, Anton Van de Putte, Vito Vitale.

DoW Task Description:
A white paper on European polar data accessibility, including proposals for capacity building methodology to interconnect the existing and upcoming polar research data systems and infrastructures and thus ensure data are made more widely available and in a timely manner.

Table of content

Motivation and Background 4
Societal relevance 5
The Needs 6
Topics and definitions 7
Subtopic 1: Being FAIR and taking CARE 7
Subtopic 2: The need for proper data management 7
Subtopic 3: Data system interconnection 7
Subtopic 4: Regulation 8
Enabling Capacities and Resources 9
Relevant Cooperation Partners 9
The way forward and key action areas 10
Motivation and Background

In recent decades, scientific and traditional observations of polar environmental and socio-economic systems have revealed a pace, magnitude, and extent of change that is unprecedented by many measures. These changes include rapid depletion of the cryosphere, shifts in ecological communities that threaten biodiversity while precipitating challenges to food security and resilience across northern communities.

Climate and environmental impacts have effects not only within the Polar Regions but - because of the connections between the poles and lower latitudes - affect the entire global society. It is, for instance, widely accepted that the Southern Ocean is a major driver of the Antarctic climate with the circumpolar currents that isolate and insulate the continent, while through the flow of cold Antarctic bottom water currents it links the Atlantic, Pacific and Indian oceans.

However, in the face of increased human presence and local pressure from climate change and pollution, high quality current and historic data from Polar Regions are becoming increasingly valuable by contributing to a global framework of scientific knowledge. Such a reliable framework is indispensable to effectively address pressing issues, such as the impact of increasing economic activities in a sustainable manner. Moreover, due to the harsh environmental conditions that characterize Polar Regions, such as extreme cold, limited accessibility of remote places, or polar night, the collection of data is often logistically difficult and expensive. But while often great care is taken in collecting and archiving data from specific public funded expeditions or scientific programs, the management of these data is still largely lacking coordination. As a consequence, the data that has been collected is fragmented across different institutions or data centers and is often very difficult to access and use. Europe has a long tradition of scientific interest in the Arctic and the Antarctic, and records can sometimes date back to the 19th century. These can provide irreplaceable benchmarks that are indispensable to understand the current rates of environmental change.

An operational template of coordinated stewardship of data and information within different national and international programmes and its accessibility to all can be found in the Antarctic. There, The Antarctic Treaty (1) (section III.1.c) requires, "Scientific observations and results from Antarctica shall be exchanged and made freely available." Therefore, the Scientific Committee on Antarctic Research (SCAR) and the Council of Managers of National Antarctic Programmes (COMNAP) established the Standing Committee on Antarctic Data Management (SCADM) to facilitate cooperation between scientists and nations and ensure the scientific user community has access to data.

Although the Northern hemisphere is politically and sociologically more complex than Antarctica, a similar movement within the Arctic data community is taking place. There, the Sustaining Arctic Observing Networks program (SAON) together with International Arctic Science Committee (IASC) created the Arctic Data Committee (ADC), in order to promote and facilitate international collaboration towards the goal of free, ethically open, sustained and timely access to Arctic data through useful, usable, and interoperable systems.

In conclusion: given the importance of polar data for evidence-based decision making, and the financial and carbon cost of collecting this data, there is a strong and growing imperative from science and society to effectively manage, disseminate and archive European polar data in a formalized way, allowing this information to be as openly available as possible for a wide range of users. Moreover, there is also a growing need to link up with datasets held by other polar nations, where there is a similar long tradition of collecting records in Polar Regions.
Societal relevance

There is an increasing understanding of the societal relevance of Earth Observations and to have effective information systems to manage the data that is the outcome. The mandate of the Group on Earth Observations (GEO) is to improve the availability, access and use of Earth observations for the benefit of society (2). For the Polar Regions, this has been described in detail in the EU-PolarNet White Paper *Advancing operational informatics for Polar Regions* (3). In the document, it is noted that such an efficient system, and hence the appropriate data management practices and means, will improve interoperability and exploitation of distributed datasets allowing enhanced services and information systems for society, industry and science. The White Paper lists the specific business and society sectors that will benefit substantially from the development of such a system. The White Paper also underlines the fact that building an effective data and information system in the Polar Regions is aligned with the seventeen UN Sustainable Development Goals (SDGs) in a variety of ways, as shown in the table. For the Arctic region, the societal benefits of observing and the services that are based on the associated data are described in the *International Arctic Observations Assessment Framework* (4).

<table>
<thead>
<tr>
<th>Sustainable Development Goal</th>
<th>Explanation</th>
<th>Relevance to the Polar Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation</td>
<td>The Polar Regions should be a major recipient of progress in SDG9, and much of this White Paper is a consequence of the present-day lack in infrastructure.</td>
<td></td>
</tr>
<tr>
<td>11 Make cities and human settlements inclusive, safe, resilient and sustainable</td>
<td>There are numerous cities in the Polar Regions, and because of the harsh environments surrounding them, there is an urgent need to consider ways to make them safer and more resilient, as well as sustainable.</td>
<td></td>
</tr>
<tr>
<td>13 Take urgent action to combat climate change and its impacts</td>
<td>The Polar Regions are seeing some of the greatest impacts of climate change on the planet – through polar amplification of atmospheric warming and through the melting of ice.</td>
<td></td>
</tr>
<tr>
<td>14 Conserve and sustainably use the oceans, seas and marine resources for sustainable development</td>
<td>The oceans are a major source of food and income for the Polar regions, and their sustainable use and management is key to future prosperity.</td>
<td></td>
</tr>
</tbody>
</table>
Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss.

Polar regions land use change needs to be managed sustainably, if we are to maintain ecosystems and their natural services.

Strengthen the means of implementation and revitalize the global partnership for sustainable development.

The Polar Regions are very much an international space, where collaboration is both natural and essential or tackling the major problems that exist, such as access to data and information.

**Table caption:** Alignment between research in infrastructures for informatics in the Polar Regions and the seventeen UN Sustainable Development Goals (SDGs). From the EU-PolarNet White Paper *Advancing operational informatics for Polar Regions* (3).

**The Needs**

The societal relevance of observing in the Polar Regions is well established and for the Arctic the return on investment is well-documented (5, 6). While there is growing understanding that the funding of observing in the Polar Region is not just a cost, but an investment, almost all spheres in the Polar Region are still under-observed (7). Many nations and institutions are, however, currently investing in observing the Polar Regions, and for satellite observations, there are more data from the Polar Regions than from other regions; the issue is this context is that *in situ* data necessary for the interpretation are missing.

Many nations and institutions have also made significant investments in basic data structures (like data centers), and there are also initial attempts to establish connections between these.

Initiatives like the Global Observing System for Climate (GCOS) work on *Essential Climate Variables* (8), the *International Arctic Observations Assessment Framework* (4), and the *SAON Roadmap for Arctic Observing and Data Systems* (ROADS) (9) describe how Polar (Arctic) observing data should be evaluated in the light of their societal relevance. It is also described how the proper management of data from the Polar Regions should be seen as only one component in the ‘value chain’ from observing to services. And while there is still a need for expanding polar observing, there is an even stronger need for developing and sustaining infrastructures that link data to services.
Topics and definitions

The discussion of ‘polar data accessibility’ should be framed by the application of the FAIR principles (10) to the metadata and data: They must Findable, Accessible, Interoperable and Reusable (Subtopic 1). An underlying assumption for applying this principle is that adequate data management and stewardship procedures are in place (Subtopic 2). The principle of interoperability (as demonstrated for instance through the GEOSS Portal) should be applied to all data systems (Subtopic 3). Finally, relevant regulation should be in place to ensure accessibility (Subtopic 4).

Within this paper ‘polar data’ are defined as ‘data that are measured or generated by monitoring, research or operational activities in or around the polar regions and that have a strong relation to the processes or activities taking place in these regions.” Additional qualifications are that these data often are associated with the cryosphere (water in its solid form), impacted by extreme cold or polar nights/days, and have been obtained from remote places under difficult conditions.

Subtopic 1: Being FAIR and taking CARE

The H2020 ENVRI-FAIR project developed methodology to diagnose the level of “FAIRness” of the major European environmental research infrastructures. This principle should be applied wherever possible to the polar observing systems and other data collection mechanisms.

Where indigenous people are involved, the emerging concept of the CARE principles (Collective benefit, Authority to control, Responsibility and Ethics (11)) should be a reference for appropriate ethical data collection and handling.

Subtopic 2: The need for proper data management

Observational data are unique by essence, especially when related to highly varying processes. This alone gives them an intrinsic value that should be adequately preserved. But looking at the efforts and resources usually needed to collect them, and the samples that later will be used to generate data in the laboratories, it is obvious that this value should be not only preserved but maximised.

Every project programme should have an efficient Data Management Plan (DMP), focusing on the documentation of the data, preferably using standards and community good practices. There should be well-described and effectively applied Quality Control procedures and provision for adequate dissemination of the data, according to the FAIR principles, and their long-term preservation. Current efforts to identify and understand the consequences of climate change demonstrate the fact that data often get more valuable over time.

Proper data management, leading to an increased accessibility to high quality data also helps in avoiding the duplication of efforts and in coordinating research programmes by optimizing their complementarity and their use of resources (including computer infrastructures), hence minimizing carbon footprint. More details on these needs and recommendations to address these are found in (12).

Subtopic 3: Data system interconnection

A way of making polar data more easily accessible is by interconnecting as much as possible the various data holding infrastructures. One of the most advanced services striving toward this objective is the GEOSS portal:

© EU-PolarNet Consortium 18/03/2020
"The GEOSS portal is an online [...] user interface which allows users to discover and access Earth observation data and resources from different providers from all over the world. [It] provides a single internet discovery and access point to the ever-growing quantities of heterogeneous collections of Earth observations from satellites, airplanes, drones and in-situ sensors at global, regional and local scales through the Global Earth Observation System of Systems (GEOSS). The GEOSS is a social and software ecosystem connecting a large array of observing systems, data systems and processing services to strengthen monitoring of the state of the Earth. It facilitates data and information accessibility and interoperability. [It] is the “glueware” that enables the connection and coordination of the many autonomous and multi-organizational systems and services contributing to GEOSS.” (adapted from 13).

Such an interconnection service is made possible by ensuring the interoperability at the level of the metadata description, ensuring that all datasets dealing with the same discipline (or even up to the same parameters, depending on how precise the interoperability of the metadata profiles is) are findable. Datasets themselves are made available by the data infrastructures interconnected by means of the portal. Future efforts should concentrate on making sure that the interoperability also exists at the level of the data themselves.

The GEOSS model is pragmatic in the sense that it doesn’t want to interconnect everything with everything but instead, to increase the discoverability of data sets, leaving to the user to interconnect the relevant datasets.

Subtopic 4: Regulation

Polar data are collected from several spheres and by many nations, institutions and initiatives. In order to maximise the access to these, relevant regulations should be in place.

The European ‘data landscape’ is to some extent already structured, and relevant regulations exists. When managed by public bodies; data relevant to European member states needs to be compliant with several European Directives. The most relevant ones are, on the one hand, the INSPIRE Directive and on the other hand the new Open Data Directive. The first one aims at creating a European Union spatial data infrastructure for the purposes of EU environmental policies and addresses 34 spatial data themes needed for environmental applications. Through an extended set of specifications for the metadata, the data and the related services, it creates a technical framework for implementing the FAIR principles. The latter one doesn’t provide technical specifications at this stage but is a lever for disseminating the data in a wider way.

Most non-European countries have similar regulations that seek to ensure that data are made accessible, especially when they have been obtained through public funding. At the Arctic Council level, the Agreement on Enhancing International Arctic Scientific Cooperation (14) commits the Parties to “…support full and open access to scientific metadata and shall encourage open access to scientific data and data products…” Similarly, The Antarctic Treaty (1) stipulated that “…scientific observations and results from Antarctica shall be exchanged and made freely available….”

In general, the Polar ‘data landscape’ is less well regulated, but the community has formulated ‘data policies’ (15, 16, 17, 18, 19, 20) that encourage nations, institutions and initiatives to make data openly accessible. The rationale behind these is that they provide a framework for data to be handled in a consistent manner but tries to establish a balance between the rights of investigators, the rights of indigenous peoples, and the public. Such initiatives should be promoted, guided by the principle “As open as possible - as closed as necessary”.

© EU-PolarNet Consortium 18/03/2020
Enabling Capacities and Resources

In recent years, the polar data community has made significant progress in many different areas of development. The foundation of this progress has been in the area of community building through activities such as meetings organized by the ADC and partners (e.g. the Polar Data Planning Summit and the Polar Data and Systems Architecture Workshop), the Polar Data Forum series, and engagement in the Arctic Observing Summit, coordination of efforts with GEO, the Research Data Alliance (RDA), CODATA and other global efforts. Through these activities, community building priorities have been developed:

- Capacities already exist within institutions engaged in organising polar data; in many cases, it is a matter of mobilising resources within these, bringing all to the same level. This is about sharing knowledge, including learning to use systems;
- Capacities should be developed for strengthening links between social science data and national science data and their owners (21);
- For the Arctic, indigenous peoples’ partnership and funding for their active participation is critical;
- Resources should be identified for supporting ad hoc actions from within the RDA.

At the technical level, there is a need for supporting development within these areas:

- Establishment of a federated search framework that supports polar communities, researchers, decision makers and others in achieving their goals with respect to finding polar data. This is reflected through the joint development of the POLDER (Polar Data Discovery Enhancement Research) group through ADC, SCADM and the Southern Ocean Observing System (SOOS).
- Support the work of the ADC-IARPC-SCADM Vocabularies and Semantics Working group (22)

The Polar Data Forum, the ADC, SCADM and SOOS have issued a series of observations and recommendations in order to meet the expectations concerning accessibility and reusability of data.

Data stewardship and governance needs to be sustained and sustainable over long-time spans. This implies a commitment of governing agencies to provide continued and stable funding for data repositories, and their infrastructures. For data to be of high quality, ethically open and preserved over time, requires governance and sustainability.

In the current ‘funding landscape’, a nationally based funding system is insufficient to secure the necessary international cooperation outlined above. As described, countries currently are investing in polar observing, but programmes are not aligned between countries, nor are they on the same time and topics, and this prevents efficient cooperation and coordination within the polar data community. It is recommended that international funding mechanisms are developed, like the Belmont Forum. This has had a call for Arctic cooperation but has in general lacked consistency and is not regularly announced.

Relevant Cooperation Partners

The polar data community consists of a wide variety of data producers, managers, and users in government, industry, academia and northern communities that need data for scientific research and to support operations and livelihoods in the Polar Regions. Not surprisingly, there is a large degree of overlap among both the organizations and people involved in these activities. The ADC has initiated efforts to provide a ‘map’ that will document the participants and their interactions for the Arctic (Appendix).
The community has created three organizations for the purpose of promoting and facilitating international collaboration towards the goal of free, ethically open, sustained and timely access to polar data through useful, usable, and interoperable systems. In the Arctic, this is the ADC of SAON and IASC; in the Antarctic, this is SCADM and SOOS.

These organizations:
- Advise their communities on matters related to data management and data sharing.
- Contribute to the understanding of the nature and structure of the polar data system in the context of the global data system.
- Promote and enable:
  - Ethically open access to data
  - Norms of fair attribution and use of data
  - Long term preservation of data
- Facilitate the adoption, implementation and development (where necessary) of standards that will enable free, open and timely access to data.
- Facilitate interoperability of data and systems as needed to support the needs of researchers, (Arctic) residents, decision makers and others.
- Establish expert groups to examine specific questions or coordinate the implementation of data management and sharing solutions. Partnerships with existing or proposed initiatives driven by members of the polar science and data community and Northern communities are encouraged.

ADC, SCADM, SOOS, and the Polar Data Forum series are a few examples of how the polar data community is working together to coordinate efforts. There are also major regional and national initiatives established and emerging, like the EU-PolarNet. In Canada, Polar Knowledge Canada is taking a lead in this area along with the Canadian Consortium on Arctic Data Interoperability (CCADI), the Polar Data Catalogue (PDC), and other initiatives. In the U.S., the Interagency Arctic Research Policy Committee (IARPC) Arctic Data Coordination Team (ADCT) is taking a coordinating lead in partnership with initiatives such as the NSF-funded Antarctic and Arctic Data Consortium (a2dc).

There is a convergence and linkage of these efforts through strong cooperation across the groups. For example, the ADC and SCADM have drafted a Memorandum of Cooperation to enhance collaboration and efficiency between these groups and other global and national initiatives such as the IARPC ADCT, a2dc, SOOS, RDA, World Data System (WDS), GEO, and others.

**The way forward and key action areas**

Given the needs and possibilities identified above, it is possible to make some targeted recommendations for adjustments and better use of available infrastructures:
- Promote efforts that evaluates polar observing data in the light of their societal relevance. This should document that data is only one component in the ‘value chain’ from observing to services. An example from the Arctic region is the International Arctic Observations Assessment Framework (4), describing the societal benefits of observing and the services that are based on the associated data.
- Develop societal relevant services based on polar data and develop and sustaining infrastructures that link data to services.
- Proper data management is a strong need for enhancing quality, reducing the need for unnecessary duplication of research, speeding up scientific progress, and it can overall favour economic growth and innovation. Commendable efforts must be made to ensure that data management planning becomes a standard scientific practice. Efforts should support the
dissemination of research data that are Findable, Accessible, Interoperable and Re-usable (the FAIR principles). Where indigenous people are involved, the CARE principles should apply.

- Interconnecting to the extent possible the various data holding infrastructures is a way of making polar data more easily accessible. This would involve
  - Establishment of a federated search framework that supports polar communities, researchers, decision makers and others in achieving their goals with respect to finding polar data.
  - Support the work on developing polar vocabularies and semantics
  - Promote the principles under the GEOSS model and seek inspiration from the model to support coordinated polar data accessibility.

- The polar ‘data landscape’ is not well regulated, but ‘data policies’ have been formulated. They offer a framework for establishing a balance between the rights of investigators, the rights of indigenous peoples, and the public. These should be further developed and promoted, guided by the principle “As open as possible - as closed as necessary”.

- Support the organisations that have been established with the purpose of promoting and facilitating international collaboration towards the goal of free, ethically open, sustained and timely access to polar data through useful, usable, and interoperable systems. In the Arctic, this is the ADC of SAON and IASC; in the Antarctic, this is SCADM and SOOS.
Acronyms

a2dc: Antarctic and Arctic Data Consortium (USA)
ADC: Arctic Data Committee
ASDI: Arctic Spatial Data Infrastructure
CARE: Collective benefit, Authority to control, Responsibility and Ethics
CCADI: Canadian Consortium on Arctic Data Interoperability
CODATA: Committee on Data of the International Science Council
COMNAP: Council of Managers of National Antarctic Programmes
DMP: Data Management Plan
EOSC: European Open Science Cloud
FAIR: Findable, Accessible, Interoperable and Reusable
GBIF: Global Biodiversity Information Facility
GCOS: Global Observing System for Climate
GEO: Group of Earth Observation
GEOSS: Global Earth Observation System of Systems
IARPC: Interagency Arctic Research Policy Committee (USA)
IASC: International Arctic Science Committee
PDC: Polar Data Catalogue
POLDER: Polar Data Discovery Enhancement Research
RDA: Research Data Alliance
SAON: Sustaining Arctic Observing Networks program
SCADM: Standing Committee on Antarctic Data Management
SCAR: Scientific Committee on Antarctic Research
SOOS: Southern Ocean Observing System
STPI: IDA Science and Technology Policy Institute, Washington, DC, U.S.A.
WDS: World Data System

References

5. Value tree for physical atmosphere and ocean observations in the Arctic. https://helda.helsinki.fi/handle/10138/300768

© EU-PolarNet Consortium 18/03/2020
8. GCOS: https://gcos.wmo.int/en/essential-climate-variables


Appendix: The Polar data ‘ecosystem’

Work is ongoing within ADC to map the Antarctic and Arctic data ‘ecosystem’. In alphabetic order, this is the list of particularly important international initiatives that the European community should connect with:

**Antarctic:**
- Scientific Committee for Antarctic Research
- SCAR Standing Committee on Antarctic Data Managers and National Antarctic Data Centres
- Antarctic Master Directory
- Southern Ocean Observation System

**Arctic:**
- Arctic Council and Working Groups
- ArcticNet (Canada)
- Arctic Spatial Data Infrastructure (ASDI)
- Arcus (USA) (Probably best to work through US AON program below)
- Canadian Consortium for Arctic Data Interoperability (Canada)
- IASC-SAON Arctic Data Committee
- Interagency Arctic Research Policy Committee (IARPC) (U.S.)

**Global:**
- Committee on Data for Science and Technology (CODATA)
- Group on Earth Observations (GEO)
- Global Biodiversity Information Facility (GBIF)
- Ocean Biogeographic Information System
- Research Data Alliance (RDA)
- WMO/Global Cryosphere Watch (GCW)
- World Data System (WDS)