Integrated Arctic Observation System

Research and Innovation Action under EC Horizon2020
Grant Agreement no. 727890

Project coordinator:
Nansen Environmental and Remote Sensing Center, Norway

Deliverable 1.8

Data Management Plan V2

Start date of project:	01 December 2016	Duration:	60 months
Due date of deliverable:	31 May 2021	Actual submission date:	31 May 2021
Lead beneficiary for preparing the deliverable:		NERSC	
Person-months used to produce deliverable:		0.5 pm	

Authors: Torill Hamre, Hanne Sagen and Stein Sandven (NERSC), Finn Danielsen (NORDECO), Geir Ottersen (IMR), Agnieszka Beszczynska-Moller (IOPAN), Arnfinn Morvik and Asuka Yamakawa (IMR), Ingo Schewe (AWI), Martin Enghoff (NORDECO)

Reviewed by: Tor I. Olaussen (NERSC)

Version	DATE	CHANGE RECORDS	LEAD AUTHOR
0.1	18/02/2021	Initial version, based on D1.6 (data governance framework, including data management plan).	T. Hamre
0.2	05/04/2021	First input from partners.	A.Morvik, G.Ottersen, I.Schewe, F.Danielsen
0.3	30/04/2021	Update on in situ data datasets and developed templates for metadata for ocean mooring data.	A.Beszczynska-Moller, A.Yamakawa, T.Hamre, A.Morvik
0.4	06/05/2021	Incorporated comments from review.	T.Hamre, Tor I. Olaussen
0.5	28/05/2021	Update of metadata template for ocean mooring data, template for data management plan, and recommendations for data formats and protocols.	T.Hamre, A.Morvik, A.Yamakawa, H.Sagen, A.Beszczynska-Moller
0.6	31/05/2021	Technical review and submission	K. Lygre

Approval X	Date: 31 May 2021	Sign. Coordinator

USED PERSON-MONTHS FOR THIS DELIVERABLE					
No	***Beneficiary***	***PM***	***No***	***Beneficiary***	***PM***
1	NERSC	***0.30***	**24**	TDUE	
2	UiB		**25**	GINR	
3	IMR	***0.05***	**26**	UNEXE	
4	MISU		**27**	NIVA	
5	AWI	***0.05***	**28**	CNRS	
6	IOPAN	***0.05***	**29**	U Helsinki	
7	DTU		**30**	GFZ	
8	AU		**31**	ARMINE	
9	GEUS		**32**	IGPAN	
10	FMI		**33**	U SLASKI	
11	UNIS		**34**	BSC	
12	NORDECO	***0.05***	**35**	DNV GL	
13	SMHI		**36**	RIHMI-WDC	
14	USFD		**37**	NIERSC	
15	NUIM		**38**	WHOI	
16	IFREMER		**39**	SIO	
17	MPG		**40**	UAF	
18	EUROGOOS		**41**	U Laval	
19	EUROCEAN		**42**	ONC	
20	UPM		**43**	NMEFC	
21	UB		**44**	RADI	
22	UHAM		**45**	KOPRI	
23	NORUT		**46**	NIPR	
			47	PRIC	

Listed person-months (PM) include the work with the writing of the report

DISSEMINATION LEVEL		
PU	Public, fully open	**X**
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

NERSC Technical Report Series. Report number 409.
DOI: https://doi.org/10.5281/zenodo.7015039

EXECUTIVE SUMMARY

This document contains the Data Management Plan (DMP) for the INTAROS project. The DMP describes how new datasets collected or generated by partners in the project, are managed according to guidelines for FAIR data management in Horizon 2020. This includes the procedures for planning and conducting data management within the project, i.e., the data governance framework. Data governance in INTAROS is pragmatic and geared towards supporting partners in preparing and publishing their data collections. The planning and monitoring activities are carried out by the Data Management Theme Leader (NERSC) and the leaders of the data generating work-packages in the project (FMI, IOPAN, NORDECO, TDUE, IMR). Partners generating data are responsible for making their collections available in line with the recommendations of the DMP. The Data Management Theme Leader (NERSC) and data centre partners (AWI, CNRS, FMI, IMR, IFREMER, ONC, RADI, RIHMI-WDC) provide support with data publication in open data repositories. Close collaboration between data managers and data providers has been key to implementing sound data management in the project.

INTAROS is pan-Arctic in scope and collects *in situ* observations, extract parameters from satellite data and model projections in several regions and across multiple spheres (themes). The focus areas of INTAROS include Coastal Greenland, North of Svalbard, Fram Strait, the Eurasian Basin, and (5) selected sites in Siberia, Finland, Canada and Alaska. Within these areas, INTAROS partners are collecting new observations and generating high-level data products from different spheres: (1) Atmosphere, (2) Ocean, (3) Sea ice, (4) Marine ecosystems, (5) Terrestrial, (6) Glaciology, (7) Natural hazards, (8) Community-based monitoring.

Datasets collected or generated within these spheres by the time of writing are summarised in this document, based on the deliverables from WP 2 ("Exploitation of existing observing systems"), datasets collected in WP 3 ("Enhancement of multidisciplinary in situ observing systems") and WP 4 ("Enhance community-based observing programs for participatory research and capacity-building"), as well as model products and derived datasets from WP6 ("Applications of iAOS towards Stakeholders"). The published datasets have been registered in the INTAROS Data Catalogue, available at https://catalog-intaros.nersc.no/. This data catalogue is updated as new datasets are prepared during the remainder of the INTAROS project.

The DMP recommends standards for metadata and data standards that INTAROS partners should prepare their datasets in, to make it easier for other scientists and stakeholders to reuse the data. Some widely used open-source tools and libraries that can help scientists generate metadata and data in standard formats are described. Use of servers that support the OPeNDAP standard protocol is recommended to facilitate data extraction from distributed sources. INTAROS, together with the Useful Arctic Knowledge (UAK) project has organised several user meetings and one research school to build competence in data management within the INTAROS consortium. Training material from these and other events with INTAROS contribution is made publicly available on the INTAROS web site http://intaros.eu/.

Major changes since the previous release of the DMP (Hamre et al., 2019) include:

- Revised the list of recommended metadata and data formats (section 2)
- Updated the list of datasets collected or generated (sections 3-6)
- Updated list of recommended data repositories for long-term storage of data (section 8)
- Added a template for planning data management for a field experiment or a monitoring programme (Appendix D)
- Added a template for metadata for timeseries data collected by ocean moorings (Appendix E)

Table of Contents

1. Introduction

The Data Management Plan (DMP) for INTAROS describes how new datasets collected or generated by partners in the course of the project, are managed according to guidelines for FAIR data management in Horizon 2020 (European Commission, 2016). This includes a plan for preparation of datasets collected or generated in the project in standard formats documented according to best practices, ingestion of the new data in an established data repository, long-term storage and curation of data, an overview of capacity building material for data management that is prepared together with the Useful Arctic Knowledge (UAK) project, as well as the the procedures for organising data management in the project. The UAK project (led by NERSC) is funded by the Research Council of Norway under the INTPART program.

The Data Governance Framework (DGF) for INTAROS defines the procedures for how data management is carried out in the project, including the planning, conducting and monitoring the preparation and distribution of data collections. The planning and monitoring activities are carried out by the Data Management Theme Leader and the leaders of the data generating work-packages in the project (WP2, WP3, WP4, WP5, WP6). Partners generating data are responsible for making their collections available in line with the recommendations of the DMP. The Data Management Theme Leader (NERSC) and data centre partners (AWI, CNRS, FMI, IMR, IFREMER, ONC, RADI, RIHMI-WDC) provide support with data publication in open data repositories. Close collaboration between data managers and data providers has been key to implementing sound data management in the project.

Scientific datasets collected or generated during the project are registered in the INTAROS Data Catalogue (https://catalog-intaros.nersc.no), which holds key metadata elements for each dataset (Appendix A). For community-based data special conditions may apply; Appendix B contains the information and consent form for all participants in the community-based activities within INTAROS. Within the stakeholder application work-package, WP6, a range of models will be run, and this will generate model-based data collections. Appendix C contains a template with key metadata elements describing the models used. Appendix D contains a template for planning data management for a field experiment or monitoring programme that is carried out with support from INTAROS. Template E holds a metadata template for time series of point and profile measurements from ocean moorings. This template is developed in 2021 by a designated working group comprised of experts in ocean data collection and in data management, ensuring that both domain knowledge and technical expertise are incorporated in the metadata. Partners with INTAROS funded instruments on ocean moorings are now using the template to publish their data in a standard compliant format and with a metadata structure following international standards e.g. CF, CMEMS, ACDD, OceanSITES.

2. General data management principles

FAIR principles for data management

Data management in INTAROS is carried out in accordance with guidelines for FAIR data management in Horizon 2020 (European Commission, 2016). This means data collected or generated in the project must be:

- F (Findable) – “making data findable, including provisions for metadata”.

- A (Accessible) – “making data openly accessible”.
- I (Interoperable) – “making data interoperable”.
- R (Reusable) – “increase data re-use (through clarifying licenses)”.

A key element of making data *findable* is ensuring that all datasets are accompanied with rich metadata describing the contents and how data has been processed and quality controlled, as well providing a persistent identifier that uniquely identifies every dataset and supports versioning of datasets.

All data generated by the project are made *available* according to an Open Data policy in line with the recommendations from Horizon 2020. This will enable external parties the right to access and use the digital datasets created by INTAROS, while respecting general terms and conditions as defined in the Grant Agreement and acknowledging that the originating partner retains the ownership of their datasets. The data governance framework defines a set of guidelines for governance of the datasets generated by INTAROS, considering, among others, the needs for documenting data quality, procedures for secure long-term storage and curation, as well as a mechanism for data search, retrieval, and use.

INTAROS partners make the datasets collected or generated in the project available in standard data formats, with discovery metadata encoded using standard vocabularies. This enables INTAROS datasets to be easily integrated in the iAOS portal through the data catalogue and used by the iAOS cloud platform. In this way the data sets from the INTAROS project will be findable, accessible and *interoperable* with in other data systems holding Arctic data. As a result the INTAROS data are made available for reuse.

Data collected or generated by INTAROS during the planned field campaigns and extensions to ongoing observations are published in accordance with the FAIR data principles. Data from community-based monitoring programmes and collaboration with local communities in the project, are made available as agreed with the respective communities. Access restrictions may apply to community data. This is done to protect the rights and interests of the local communities and the individuals living there. Both scientific and community-based datasets are accompanied by a data license clearly stating how the data can be *reused* by the scientific and wider user community.

Data governance

INTAROS uses a simplified version of the DAMA-DMBOK2 data governance framework (Cupola et al., 2014). Focus has been on establishing formal procedures for planning, conducting and monitoring data management activities in the project, given its scope and international character. The planning and monitoring activities are carried out by the Data Management Theme Leader and the leaders of the four data generating work-packages in the project. Partners generating data are responsible for making their collections available in line with the recommendations of the DMP. The Data Management Theme Leader and data centre partners provide support with data publication in open and sustained data repositories ensuring data remain available long-term.

INTAROS relies on existing data infrastructures holding Arctic data. Thus, the organisation and long-term storage of data will be handled by the respective infrastructures. INTAROS role is to

recommend a set of such data infrastructures that are mature and sustained through national or international long-term commitments. This ensures INTAROS data are securely stored and remain accessible for future Arctic projects and programmes.

Data security is key aspect of INTAROS, especially for community-based monitoring (CBM) data. These data are typically managed by the communities themselves, or information and communications technology (ICT) specialists of collaborating research organisations or community networks like ELOKA[1]. INTAROS has taken measures for protecting CBM data from the beginning. For instance, the original (sensitive) CBM data are password protected, and non-sensitive data intended for wider use can only be browsed by guests if the guests first agree to adhere to a set of ethical and appropriate use guidelines, and to cite the data if it is used in publications. Within the project, the data collected as part of CBM activities have been managed by the corresponding local community or community network. The CBM data ownership and data use rights follow principles of "Free, Prior and Informed Consent" (Appendix B). Some of the field experiments in WP3 have been conducted by PhD students or young scientists. In these cases, INTAROS has taken measures to protect their data until they have published a peer reviewed paper or thesis based on these data. In addition to traditional scientific papers, all data providers are encouraged to publish data papers describing their data and to reuse data in other scientific studies or (if permitted) in commercial settings.

Selected aspects of the DAMA-DMBOK2 data management knowledge area "Data integration & interoperability" is addressed by the development of the iAOS cloud platform in WP5. This cloud platform will provide a set of services to illustrate integration of multi-source data, various geo-statistical operations, and other services relevant to stakeholder applications in WP6.

The INTAROS Data Catalogue holds metadata describing the data collections generated by the project. All partners generating data collections in the project, are obliged to register their data in this catalogue. Additional metadata for data collections that can be used in the stakeholder applications in WP6 can be registered as well or harvested from other catalogues offering a standard interface like OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) and OGC CSW (Open Geospatial Consortium Catalogue Service for the Web).

Data collected or otherwise generated within INTAROS originate from many different spheres. Consequently, the processing chains and standards used for calibration and quality checking differ accordingly. Thus, each partner in WP2, 3, 4, 5 and 6, must follow the best practices and use the community standards for data collection, processing and analysis within their scientific domain or community-based monitoring. Resulting data collections will be stored in one of the recommended data infrastructures in this document (Section 8), with metadata describing the data quality.

Data value chain

The INTAROS project addresses the full data value chain from observation or simulation to delivery to users through the iAOS Portal or custom applications. All data are stored in open data repositories with long-term funding to ensure data remain available for future use in science, public and private sector. Figure 1 shows schematic workflows for the major

[1] Exchange for Local Observations and Knowledge of the Arctic

categories of data collected and generated for integration in iAOS portal, cloud platform and applications: (1) in situ observations, (2) remote sensing data and (3) model simulations. Each workflow uses established practices to process and quality control the data before preparing datasets for publication using standard formats. Data are stored in open data repositories enabling the iAOS portal, cloud platform and applications to integrate data using standard protocols. The iAOS does not include a data repository so the applications and cloud based services must establish the link directly to the data center holding the required data sets (dotted arrows in Figure 1).

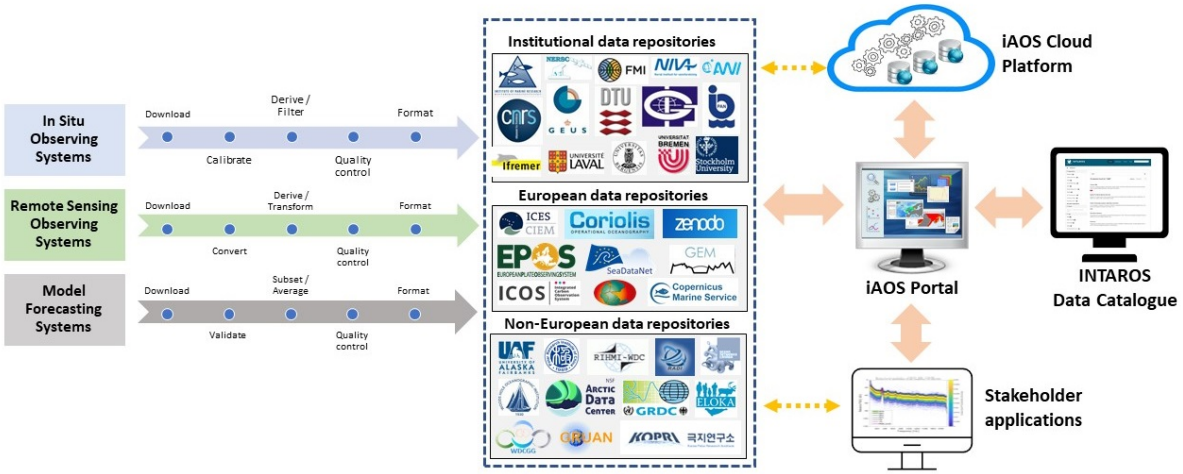

Figure 1. Data value chains for integrating INTAROS data into the various iAOS subsystems.

Metadata and data standards

For the INTAROS datasets, partners will use standards such as GCMD DIF and ISO 19115 to hold general descriptive metadata, i.e. discovery metadata, to support flexible search in iAOS. Standard vocabularies such as GCMD Science Keywords and SeaDataNet-2 vocabularies (e.g. for parameter and sensor names, units, keywords) will be used to mark up the metadata in a manner that facilitates machine readable search and retrieval. Metadata following GCMD DIF must be encoded using DIF-9 or DIF-10 XML schemas; metadata following ISO 19115 must be encoded in XML using the ISO 19139 standard. To facilitate reuse of a dataset, a more detailed description of its content, including among others, parameters and units used, is needed. Such usage metadata are well defined for some types of data, e.g. physical oceanography data from CTDs and EO data from satellites. However, for other types of data, there is a lack of standards. For instance, while GCMD and ISO19115 can represent general descriptive metadata for acoustic data from scientific experiments, there is currently no standard that can fully represent the usage metadata. The lack of standardized representation of metadata for passive acoustic data from distributed ocean acoustics observation networks, has been addressed by NERSC in other ongoing research projects.

In the NorDataNet project (https://www.nordatanet.no/) funded by the Research Council of Norway, NERSC has developed a new data format for the acoustically sensed ocean temperatures from the Fram Strait Multipurpose Acoustic system (Yamakawa et al., 2019). This format is based on the NetCDF Climate and Forecast (CF) Metadata Conventions and the metadata structure developed by the OceanSITES program (OCEANSites, 2010). NERSC has

developed a second format, also based on NetCDF/CF and OceanSITES, for the ambient noise data from the Fram Strait Multipurpose Acoustic system. The datasets from EC projects DAMOCLES and ACOBAR have been converted to this format and stored at NERSC. During INTAROS, these datasets are made available in iAOS as well as NMDC.

To describe usage metadata in a consistent and machine-actionable manner, it is imperative to use standard vocabularies. These offer unambiguous definitions of key concepts and terminology within a scientific domain, with each term having its unique identifier (URIs). This enables computer programs to distinguish between syntactical similar terms and to deduce the semantics (i.e. meaning) of the metadata. As an example, the GCMD keyword vocabulary[2] describes Science and Services Keywords (Olsen et al., 2013), Data Centers, Projects, Instruments, Platforms, Locations, Horizontal Data Resolution, Vertical Data Resolution, Temporal Data Resolution and URL Content Types. The NERC Vocabulary Server (NVS) Common Vocabulary[3] holds an extensive set of vocabularies for parameters, sensors, platforms, ships, organisations, projects, to name a few. These vocabularies have been developed by the British Oceanographic Data Centre in collaboration with the scientific community in a series of projects for the past decades. Currently, the development of the NVS Common Vocabulary is carried out within the frame of SeaDataNet[4], a distributed Marine Data Infrastructure with more than 50 partners and sub-contractors from Europe, Russia, Africa and Australia. A third highly relevant vocabulary is the list of standard parameters names and units in the NetCDF Climate Forecast (CF) conventions (Eaton et al., 2011).

NetCDF/CF defines a set of required elements, including descriptive and usage metadata, that should be stored in a NetCDF file to allow users to both discover datasets and decide whether they are fit for a particular purpose. With NetCDF/CF each variable can be described in detail using standard names and units, allowing computer applications to extract, process (e.g. re-grid) and display user selected values. In addition to parameter names and units, the CF also enables definition of, amongst others, time (point or interval), coordinate axis, depth axis and map projection. To describe the data set as a whole, CF includes metadata elements for data set title, name of institution producing the data originally, data source, history (of processing), references (e.g. scientific or technical literature) and comments. The Attribute Convention for Data Discovery (ACDD) defines a set of attributes recommended to make a dataset stored in NetCDF easier to discover by data systems (ESIP, 2020). These metadata attributes can also be used by data servers such THREDDS to extract metadata from the datasets that are published and exporting those metadata to other metadata standards (e.g. Dublin Core (ISO, 2017), DIF, ISO 19115). This will also help both data systems (e.g. data catalogues, portals and digital libraries) and users in finding and using the published data more efficiently.

For the datasets themselves, the following standard formats are recommended (in alphabetical order):

- CSV (comma-separated values)
- Darwin Core (Taxonomic Databases Working Group Standard)
- GeoJSON
- GeoTIFF

[2] http://gcmd.nasa.gov/learn/keywords.html
[3] https://www.seadatanet.org/Standards/Common-Vocabularies
[4] https://www.seadatanet.org/

- HDF-EOS (Hierarchical Data Format - Earth Observing System)
- JSON (JavaScript Object Notation)
- NetCDF/CF1.8+ACDD1.3 (NetCDF/Climate and Forecast metadata convention + Attribute Convention for Data Discovery)
- OGC GeoPackage (Open Geospatial Consortium GeoPackage)
- Shapefile

Several of these formats allow for metadata to be encoded as part of the data file(s). This is highly recommended since it ensures the data consumer has access to alle relevant metadata when downloading the data file(s). For the other formats, metadata should be placed in an accompanying file, according to the agreed structure using standards such as the Directory Interchange Format (DIF) Standard from NASA or the ISO 19115:2013 "Geographic Information — Metadata" standard from ISO/TC 211.

Data repositories

INTAROS does not establish new data repositories, but instead utilizes existing repositories and data infrastructures holding Arctic data in Europe, US, Canada and Asia. With data being collected in or estimate for many different regions of the Arctic and across multiple spheres, there is no single data repository that is an optimal choice for storing INTAROS data. The partners have therefore formulated general requirements for the data repositories to be recommended for long-term storage and curation of INTAROS datasets. These data repositories must, among others

- Support metadata standards DIF-9, DIF-10 or ISO 19115.
- Prescribe use of standard vocabularies for metadata (such as those listed above).
- Dates shall always be represented in ISO 8601 format.
- Prescribe use of standard data formats (such as those listed above).
- Offer a unique identifier for each dataset, such as a Digital Object Identifier (DOI),
- Provide an open data policy and a data license for all datasets.
- Have an established and well documented data governance framework.
- Offer secure storage with appropriate access control for scientific and sensitive data.
- Have sustained funding for regular operation and maintenance the next 10+ years.

In addition, the data repositories must provide machine-to-machine interfaces for metadata and data exchange. Specifically, they must support:

- Metadata search and access through OAI-PMH and OGC CSW.
- Data access through OPeNDAP, OGC WFS or OGC WCS for extraction of user defined subsets of datasets.
- Data access through HTTPS or HTTP for download of full datasets (as files).
- Generation of simple maps (raster images) through OGC WMS of selected parameters in a dataset.

Data access through FTP for download of full datasets (as files) is optional, but no longer recommended for long-term storage of INTAROS datasets. Leading browsers such as Google Chrome and Mozilla Firefox have stopped support of FTP, meaning external applications must be used to access FTP sites from these browsers.

Long-term plans for data management in a Sustainable Arctic Observing System

In 2014, the Sustained Arctic Observing Networks (SAON) established the Arctic Data Committee (ADC) to map Arctic data management projects and services with their locations and relationships, to identify and promote common metadata elements, to provide a guide on data publication and citation, and to conduct interoperability experiments for selected regions and SBAs (Social Benefit Areas). They also established the Committee on Observations and Networks (CON), to advise the SAON Board on funding, coordinating, and extending existing observation systems, and planning for their sustainability. Several Spatial Data Infrastructures (SDIs) and data repositories hold data for either the whole, or part of, the Arctic. These SDIs are operated by different organizations and communities world-wide, making it a challenge to reach agreement on common metadata and data standards, data policies and governance frameworks.

INTAROS has established contact with both SAON committees, ADC and CON, as well as with other initiatives, such as YOPP and GEOCRI, that address data management and interoperability between spatial data infrastructures for Arctic regions. INTAROS is a member of the SAON Roadmap Task Force, comprising the pan-Arctic forum for developing a Roadmap for a Sustainable Arctic Observing System (SAOS). Expertise and experience from the INTAROS project will provide supplement the SAON Roadmap, through the major deliverable from INTAROS: The INTAROS Roadmap(Sandven et al., forthcoming) towards sustainable data delivery chains from instruments to stakeholder defined applications.

3. In situ datasets collected by INTAROS

INTAROS collects data using platforms such as aircrafts, research vessels and ships of opportunity, stations on land, ice-tethered instruments (SIMBAs and IAOOS platform), gliders, bottom-anchored systems and fixed ocean moorings in WP3 (Enhancement of multidisciplinary in situ observing systems). The data collection in INTAROS is designed to fill selected gaps in the following regions: (1) Coastal Greenland, (2) North of Svalbard, (3) Fram Strait, (4) Eurasian Basin, and (5) sites in Siberia, Finland, Canada and Alaska for terrestrial and atmospheric measurements. Additional activities are focused on integration of seismometers and biogeochemical sensors into the existing monitoring programs by adding an Arctic component to EPOS and ICOS infrastructures.

In situ observing systems in the Arctic are limited due to logistical constraints and the high cost of deploying and maintaining equipment in this region. The sparseness of in situ data is therefore the largest gap in the overall observing system. Therefore, the in situ data collection activities in INTAROS aim at filling some of these gaps, especially in the ocean which is severely under sampled compared to the other spheres (terrestrial, atmosphere, cryosphere).

During the 2017-2018 field season, the pilot phase of INTAROS fieldwork, INTAROS partners have organised or participated in campaigns in Alaska, Eastern Canadian Arctic, Greenland, Fram Strait, North of Svalbard, central Arctic, northern Finland and Siberia. The in situ datasets collected during the above field campaigns will be registered in the INTAROS Data Catalogue once they have undergone quality control and are made available for use by other scientists and stakeholders.

During the first and second INTAROS field season, 2018-2019 and 2019-2020, the observing infrastructure have been implemented by INTAROS partners or already in operation on Greenland, including coastal regions of the Greenland Ice Sheet (GIS) and Greenland coastal waters (in particular the Young Sound), in the Nansen Basin north of Svalbard, in Fram Strait and Svalbard fjords (Kongsfjorden), in the central Arctic Ocean, in Alaska, Eastern Canadian Arctic, northern Finland and Siberia. Data collected by the INTAROS partners cover many different spheres and include tens of different variables, specific to each sphere. Consequently, the processing chains and standards used for calibration and quality checking differ accordingly as well as data formats and data repositories selected for long-term preserving of collected observations. In general, data from instruments and platforms, operating during the 2018-2019 and 2019-2020 field season, will be available after the relevant data processing and quality control is completed for each different type of data and the final data products are submitted to open data bases or made available through dedicated ftp servers or websites. Detailed information on data types, data processing and planned storage for each activity in WP3 are provided in WP 3 deliverables (D3.6, D3.7, D3.8, D3.9, D3.14), focused on the results from the first implementation of INTAROS observing platforms and sensors and will be updated in the following WP3 deliverables (D3.10, D3.11, D3.12, D3.13, and D3.15) describing final implementation of and preliminary results from INTAROS in situ observing system.

New data collected in the coastal Greenland during the first and second field season include time series of physical ocean variables collected on four moorings in the Young Sound (e.g. temperature, salinity, particles) and measurements on stations during the research cruises, including CTD, carbonate chemistry (pCO2, DIC, and TA) and nutrients profiles. New data on snow water equivalent were collected with four SnowFox instruments on the GIS together with standard AWS measurements and new measurements of rainfall rate were also gathered. Improved radiation measurements were collected by AWS equipped with new tilt and azimuth sensors provided by INTAROS. Acoustic recordings were collected under the sea ice in the Young Sound for bio- and geo-sound studies. Sea-ice inherent optical properties were measured in the coastal Baffin Bay in 2019 and in the central Arctic in 2020.

The array of seven INTAROS moorings north of Svalbard deployed in 2018 for the first field season was recovered in 2019 during different cruises. Due to very heavy ice conditions in the mooring area in summer and autumn 2019 and limited ship time available, one mooring had to remain in water until 2020. Additional two moorings were deployed in 2019 for the second field season and one was recovered in 2020 (the second one could not be recovered due to the COVID-related lack of ship time and will be retrieved in June 2021). During the first field season 2018-2019, instruments deployed on the INTAROS moorings provided time series of point and profile measurements of physical (ocean pressure, temperature, salinity, turbidity, ocean currents, sea ice drift and draft), biogeochemical (dissolved oxygen, pCO2, nitrates, chlorophyl a, FDOM) and biological (visual particle sampling with underwater vision profiler) variables. Passive samplers collected samples of inorganic contaminants on 3 of 7 moorings. Moorings redeployed in 2019 for the second field season measured mostly physical variables (point and profile measurements of ocean pressure, temperature, salinity, currents, sea ice, dissolved oxygen). Two INTAROS instruments (Signature 100 and Signature 250) were deployed on the ATWAIN moorings. INTAROS Signature 100 and another Signature 100 (part of the RCN-funded SIOS InfraNor project) in November 2019 were recovered during the IMR

ecosystem survey in September 2020. The INTAROS S100 that covered the lower part of the water column unfortunately suffered a memory failure but the SIOS S100, covering the upper 400 m recorded successfully for the entire deployment period, contributing data also to INTAROS. Three Ocean Bottom Seismometers were deployed for 2018-2019 in Fram Strait and for 2019-2020 in Storfjorden, providing seismic data for solid Earth processes and geohazards.

In Fram Strait the Autonomous arcFOCE (arctic Free Ocean Carbon Enrichment) system was deployed for the first operational period in 2018-2019 and provided observations of deep-sea bacteria and meiofauna organisms (focus on nematodes) under ocean acidification, i.e. changes in community structure and ecosystem functioning. In 2019-2020 the arcFOCE system was reconfigured and rebuilt, based on results from the first deployment and will be launched again in 2021. Time series of temperature, salinity and carbonate chemistry (pCO2, pH, TA) were continuously collected with real-time data availability in Kongsfjorden at the AWIPEV observatory. The long-term acoustic recordings to study the soundscape diversity including benthic fauna sounds, marine mammals' vocalisations, ice sounds, boat noise, wind/wave noise were also collected in Kongsfjorden during both field seasons (2018-2020).

In the central Arctic Ocean, the upper ocean CTD profiles were collected in autumn 2018 with the ice-tethered IAOOS platform together with data from sea ice mass balance instrument (SIMBA), meteorological sensors and microlidar. Unfortunately, the IAOOS platform stopped working after 4 weeks, the most likely due to damage by polar bears. Data collected include ocean temperature, salinity and dissolved oxygen profiles, temperature profiles through air, snow, ice and surface ocean, air temperature and atmospheric pressure, and microlidar profiles. For the second field season 2019-2020, instead of deploying the second ice-tethered platform, the deep ocean mooring was deployed for one year in the deep Nansen Basin. Instruments deployed on this mooring provided point measurements of temperature, salinity, dissolved oxygen, profiles of ocean currents, carbonate chemistry (pCO2 and pH), sea ice drift and draft and passive acoustic recordings. Measurements of snow-ice-ocean temperature profiles, snow depth and ice thickness were collected with nearly 40 ice-tethered SIMBA buoys (nearly 20 provided by the INTAROS partner FMI) deployed in 2018-2020 during different campaigns in the central Arctic Ocean. Physical ocean measurements (temperature, salinity, dissolved oxygen and fluorescence) were collected by gliders in the upper 1000 m in the northern Fram Strait (in open water) along repeated sections in 2017, 2019 and 2020. Surface CTD, pCO2 (and pH/pCO3), CDOM, chlorophyll a, phytoplankton accessory pigments and turbidity measurements as well as microplastic samples were collected by the FerryBox system on MS Norbjørn along the repeated transect between Tromsø and Svalbard. BGC-Argo floats deployed in the Baffin Bay in 2016, 2018 and 2019 provided standard CTD data and additional biogeochemical measurements (O2, PAR, CDOM, backscatter , chl backscatter a, nitrates). Surface albedo measurements were collected in the Central Arctic during the MOSAiC campaign in 2020 as well as measurements of optical and structural properties of the Snow and Sea-ice Scattering Layer (SSL).

On Greenland (Station North) GHG were sampled with automated flask sampler to measure trace gases in the high Arctic atmosphere starting in summer 2019. Continuous concentrations of CO2and CH4 in various soil depths of the active layer to the permafrost as well as high resolution soil temperature profiles and a suite of meteorological measurements were collected at five sites in Alaska (BES, BEO, IVO, ATQ and CMDL) in 2018-2020. In the Canadian

Arctic, measurements of air temperature, wind, radiation, snow temperature, thermal conductivity, snow height, soil temperature, thermal conductivity and water content were collected at 4 monitoring sites. Measurements with the unmanned atmospheric observatory on the icebreaker Oden during the Arctic Ocean 2018 cruise included wind, humidity and air temperature observations, ocean surface temperature, incoming BB LW and SW radiation, precipitation, visibility, cloud bases and fractions, and backscatter, and eddy covariance fluxes of momentum, sensible and latent heat, CO2 and CH4, and aerosols. Similar measurements were collected during the IB Oden expedition to the Ryder glacier and fjord in 2019. New SVC-spectro-albedometer and polarized radar data were collected in 2019 and 2020 at the Sodankylä supersite in the northern Finland.

Data collected by different systems and platforms in 2018-2019 and 2019-2020 will be stored in relevant data repositories, listed in Section 8. Detailed information on planned data repository for each type of newly collected observations can be found in INTAROS deliverables D3.10, D3.11, D3.12, D3.13, D3.15 that will be finalized by autumn 2021 and in the INTAROS Data Catalogue for those data that have already been published.

4. Community-based datasets collected by INTAROS

INTAROS partners mainly work with community-based (including citizen science) observing systems in two Arctic local communities, Longyearbyen, Svalbard, and Disko Bay, Greenland. The two selected communities are high-risk regions in terms of climate change impacts as well as loss of biological diversity and can potentially benefit significantly from community-based observing programs. Furthermore, the two communities are characterized by economies which will benefit from efficient and low-cost observing programs at local levels.

During workshops and meetings in the two communities INTAROS partners in WP 4 (Enhance community-based observing programs for participatory research and capacity building) have discussed with the local authorities (Governor of Svalbard; Qeqertalik and Avannaata municipalities) to define a set of topics to be addressed in the project. Based on these topics, relevant existing datasets have been identified, and agreements made to collect new data that can augment the existing records, within the framework of INTAROS and/or existing community-based observation networks. The community-based datasets have been registered in the INTAROS Data Catalogue.

First, in Qeqertalik and Avanaatta municipalities, contributing to Greenland Government's observing network *Piniakkanik Sumiiffinni Nalunaarsuineq* (PISUNA), local natural resource experts, typically fishermen and hunters, have formed Natural Resource Councils. The members compile data on species and resource uses. The data are stored at pisuna.org and PISUNA-net (see https://eloka-arctic.org/pisuna-net/). At pisuna.org, the data are stored in their original words, language and format, exactly as they were reported by the fishermen and hunters. In PISUNA-net, English versions of the data are stored in a searchable, web-based database. The data collection is also registered in the INTAROS Data Catalogue. Second, in Longyearbyen and Qeqertalik municipality, INTAROS has established pilot community based seismic stations for cryoseismological recordings, led by GEUS and UiB. The community-based seismometers collect data from the local seismic activity (Jeddi et al. 2020). These data are accessible in real-time on https://raspberryshake.net/stationview/. Observations from the

community-based seismic stations are included in the GEUS' earthquake bulletin (see https://www.geus.dk/natur-og-klima/jordskaelv-og-seismologi/registrerede-jordskaelv-i-groenland/). They will also become part of the iAOS developed by INTAROS. Third, in Svalbard and Greenland, INTAROS with many partners has begun testing environmental monitoring by tourist cruise expedition operators. This initiative builds on four citizen science programs: Secchi Disk Study, Cloud Observations, Happywhale, and eBird. The data are stored within databases held by each citizen science programme, and the data collections are registered in the INTAROS Data Catalogue.

Finally, as a spin-off of the INTAROS capacity-development for local communities, community-based monitoring tools are used by Indigenous fishermen, hunters and reindeer herders in Zhigansk and Olenek Districts in Yakutia (Enghoff et al. 2019). The data are stored within databases held by the Republic Indigenous People Organisation of Sakha Republic, and additionally the data collection is registered in the INTAROS Data Catalogue.

INTAROS has been taking the gender dimension into account in its work with community-based observing systems (Lygre and Sagen, 2017). An important element in involving local community members in observing efforts in INTAROS, regardless of gender, is informing of the project objectives and benefits of participation for their community. The template for information and consent form for involving local communities in data gathering for INTAROS, focused on hunting and fishing, is provided in Appendix B.

5. Higher level products based on satellite, in situ and model data

In WP2 ("Exploitation of existing observing systems"), new products derived from existing satellite, in situ and model data are exploited and registered to improve the overall availability through existing repositories. Categories of such products include:

- Regionally and seasonally downscaled products based on existing datasets of carbonate system chemistry; nutrients and phytoplankton
- Unified collection of temperature and salinity data for the Barents Sea, including the area North of Svalbard, compiled from a wide range of data sources and thoroughly quality-checked. Fields interpolated from these data.
- Improved satellite-based ice concentration, ice thickness and ice drift products using new Sentinel data, passive microwave and optical data.
- ESA Sentinel/Copernicus data integrated with *in situ* observations and climate modelling to deduce sub-annual mass loss estimates from the Greenland ice sheet at sub-basin scale.
- The Alaskan transect of 5 eddy covariance towers measuring CO2, H2O, CH4 and energy fluxes will be enhanced to provide continuous, year-round data on GHG concentrations and fluxes, as well as active layer depth, water table depth and snow depth.

Higher level data products based on in situ or satellite data prepared as part of Task 2.3 are registered in the INTAROS Data Catalogue. In addition, new integrated products are developed for selected stakeholders in WP 6 (Applications of iAOS towards Stakeholders) using the iAOS cloud platform (WP 5). WP 6 itself generates a series of model products, from e.g., ecosystem models (fish stock), hydrological models (river discharge) and ice-ocean models (temperature,

sea ice thickness, and more). These higher level integrated and model-based products are registered in the INTAROS Data Catalogue.

6. Model results generated for stakeholder applications

Within INTAROS WP6 a range of different models are applied to provide model output and products targeting diverse stakeholder groups. Because of the diversity in model output and user groups (scientists, fisheries and environmental managers, local communities and more) the model output will be diverse both in size and format. NetCDF is widely used within the ocean physical modelling communities and is recommended for physical model output as well as ecosystem model output. Likewise, climate models often store their output in NetCDF files. Other models may store their output in different formats, but in general NetCDF is recommended for model output.

A few examples of the many INTAROS model applications are given below:

- Two well-established end-to-end ecosystem models, NORWECOM E2E and NoBa Atlantis have been used to evaluate to what degree indicators from the Norwegian integrated environmental management plan for the Barents Sea provide useful information on ecosystem state and trends. Results point to certain geographic areas and times in the year that are especially important to sample. Output: Integrated physical and biological marine data and indicators based on this data. Storage: Data is published through NMDC and available by ftp.In situ and remote sensing observations has been combined with ice-ocean models though data assimilation methods in an Arctic ocean–sea ice reanalysis covering the period 2007–2016. The reanalysis is based on the adjoint approach of the Estimating the Circulation and Climate of the Ocean (ECCO) consortium. Output: parameters such as ocean potential temperature, salinity, sea ice thickness, sea ice concentration, sea ice drift, and heat fluxes at the sea surface. Storage: Data is published in NetCDF format through NERSC thredds server and Zenodo. Data are used in acoustic models to understand the Arctic acoustic environment.
- Integration of satellite data from the Copernicus programme (Sentinel-1), enhanced in situ observations of surface mass balance and mass loss, and innovative modelling to deliver the contribution of mass losses from the Greenland ice sheet and Svalbard glaciers and ice caps to sea-level rise. Output: Calculated solid and fluid ice discharge and total mass loss. Storage: Greenland Ice Sheet solid ice discharge data is published in OGC GeoPackage and KML format and made available through the GEUS Dataverse repository.

7. Resources for data preparation and distribution

The costs of making datasets and products obtained within the INTAROS project available in standard formats with ample metadata, and deposit these in an approved data repository, are eligible costs under the Horizon 2020 Grant Agreement. The respective data owners in the INTAROS consortium are responsible for ensuring that their datasets and products are uploaded to one of the existing data repositories recommended by WP1 (see section 8). In general, all data and products obtained within INTAROS should go into a secured and long-term storage. In cases where the cost and potential value for long-term storage is questioned

(e.g. erratic data), the data provider, in dialogue with the INTAROS Executive Board, can decide not to save the data.

Open source tools can help scientists generate metadata and data in standard formats, such as Rosetta, GDAL (Geospatial Data Abstraction Library) and NetCDF utilities. Widely used programming languages, such as Python, MATLAB and R, offer libraries that can be used to write customised format converter tools. Some web sites with training material for metadata and data preparation include:

- DataONE Data Management Planning: https://www.dataone.org/data-management-planning
- DMPTool: https://dmptool.org/
- NMDC and NorDataNet use of Rosetta: http://tomcat.nersc.no/rosetta/
- GDAL documentation and tutorials: https://www.gdal.org/
- NetCDF utilities: ncgen and ncdump
- NetCDF/CF validation: http://puma.nerc.ac.uk/cgi-bin/cf-checker.pl
- Global Change Master Directory (GCMD) Keywords
- Links to Training and reference materials compiled by DCC: http://www.dcc.ac.uk/training/training-and-reference-materials

A dataset prepared in NetCDF format can be made publicly available using data publishing tools like the THREDDS Data Server (TDS). In November 2018, a webinar was held to exchange experience and build competence in publishing scientific data using TDS. This activity will be followed up by a Data Management Training Workshop at the upcoming INTAROS General Assembly in January 2019 and complemented with training material prepared as part of the Useful Arctic Knowledge (UAK) Winter School held in Longyearbyen, Svalbard, 3-7 December 2018. Training material prepared for these events will be used to build competence in data management within the INTAROS consortium, to facilitate preparation of INTAROS datasets in standard data formats with ample metadata compliant with established standards.

A working group aiming to standardize metadata and data formats from ocean mooring datasets in INTAROS started after the INTAROS General Assembly in 2021. The outcome is a recommended NetCDF format for INTAROS (see Appendix E). This format follows the CF 1.8 and ACDD 1.3 conventions. Additional standards for NetCDF formats for ocean datasets, like defined by CMEMS, SeaDataNet, and OceanSITES have been investigated and selected parts adopted to generate a standard format that fulfils the needs from INTAROS. The template is tool agnostic. Each partner can develop their own tool for generating NetCDF files or use a tailored Rosetta setup if their processing chain outputs a CSV file that can be converted to NetCDF by this tool.

An essential part of a Sustainable Arctic Observation System (SAOS) is long-term and secure preservation of observations and derived products. Monitoring of resources used for preservation of data and products within INTAROS will provide input to the estimation of resources needed for SAOS. This is an important part of the roadmap for SAOS to be developed in WP 1 (Requirements and strategy for Pan-Arctic Observing Systems).

8. Long-term data preservation and curation

Datasets collected or generated during the INTAROS project are stored in established data repositories with secured funding for long term preservation and curation. Partner institutions follow their institutional data policies or policies from research infrastructures including the recommended data repository for their data. The aggregated list of recommended data repositories includes (in alphabetical order):

- **Bolin Centre of Climate Research database**, operated by Stockholm University, is used for atmospheric data from the icebreaker Oden.
- **CERSAT** – a data repository of satellite-based products operated by IFREMER, e.g. Arctic sea ice drift.
- **Coriolis GDAC** (Global Data Assembly Centre) – for glider data (e.g. ocean temperature, biogeochemistry, and acoustic data). Operated by IFREMER.
- **European Plate Observing System (EPOS) RI** – a distributed network of seismic stations across Europe, and a data infrastructure for long term storage of these data.
- **Exchange of Local Observations and Knowledge for the Arctic (ELOKA)** – Global repository founded for the purpose of making community-based observing datasets publicly available while respecting sensitivities specific to the management of data related to Indigenous knowledge.
- **Global Runoff Data Centre (GRDC)** – provides access to data from the Arctic-HYCOS observing system (daily and monthly gauged river discharge data from a selection of stations operated by the national hydrological services (NHS) in the Arctic Council member states).
- **Greenland Ecosystem Monitoring Programme (GEM) database** (http://data.g-e-m.dk/) for data from the coastal Greenland observations.
- **Integrated Carbon Observation System (ICOS) RI** - a distributed network of collection of carbon data across Europe, and a data infrastructure for long term storage of these data.
- **IOPAN Oceanographic Data and Information System (eCUDO)** – a national data infrastructure (under development) operated by IOPAN for oceanographic data collections with a network of data providers from several Polish scientific institutions will be used for physical ocean observations from IOPAN moorings.
- **NOAA CMDL database and WDCGG (World Data Centre for Greenhouse Gases)** at https://gaw.kishou.go.jp/ for GHG measurements on NE Siberia
- **Nordicana D** collection at www.cen.ulaval.ca/nordicanad/en_index.aspx for thermal and carbon data of permafrost in the Canadian Arctic.
- **Norwegian Marine Data Centre (NMDC)** – a national data infrastructure with a distributed network of providers holding data for ocean areas of Norwegian interest, with a long-term mandate to preserve marine data for these ocean areas. NMDC holds a wide range of in situ observations and some model output. NMDC is a national infrastructure operated by IMR.
- **Norwegian node of EIDA (European Integrated Data Archive)** at http://www.orfeus-eu.org/data/eida for seismic data from OBS.
- **NSF Arctic Data Centre** for soil temperatures in the Eastern Canadian Arctic.
- **PANGAEA (Data Publisher for Earth & Environmental Science)** is a data centre holding a wide range of environmental datasets; global coverage. Operated by AWI.
- **Sea-Ice Portal of University Bremen (UB)** holds long time series of sea ice parameters derived from satellite data (e.g. ice concentration and ice thickness) for polar regions.

- **SEANOE (Sea scientific open data publication),** https://www.seanoe.org/, operated by Ifremer Department of Information Systems and Marine Data for all French data (acoustic recordings, glider data, French mooring data, etc.);
- **SIMBA data repository at FMI** at https://simba.srsl.com/fmi/ for data from SIMBA buoys deployed as part of the INTAROS project.
- **Zenodo** at https://zenodo.org/ is a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN. It holds not only datasets, but also other forms of digital research assets such as technical reports, scientific papers, software (source code), and videos.

Additional data repositories can be added to this list of recommendation when the datasets from the last field season in WP3 are finalised and published (ref. Section 3).

Most of the data and products from INTAROS are open, but INTAROS will respect the need to restrict access to sensitive data, such as data collected through community-based monitoring programmes. The ethical aspects of data management in INTAROS are further described in the next section.

9. Ethical aspects

All activities in the proposal meet the national legal and ethical requirements of the 8 Arctic countries. Specifically, the ethics requirements will address: (1) research related to humans, (2) protection of personal data, and (3) third countries.

1. Humans. The procedures and criteria that will be used to identify and recruit research participants for the community observing systems of INTAROS are described below.

a) Procedures: The recruitment will be undertaken by the project in cooperation with representatives of the local communities. Participation will be entirely voluntary. The community members that the project will invite to participate in project workshops, discussions, sharing of knowledge, observations and experience are seen as 'co-creators' (and not as 'objects' of research).

b) Criteria: The project together with representatives of the local communities will identify who among the local communities may be interested in participating. The criteria will be the knowledge and experiences of the community members and their interest in participating. The aim is to obtain the participation of both men and women and different age classes, as they tend to use different natural resources and have knowledge about different topics.

2. Protection of personal data. Below we describe the justification for collection and processing of personal data, the procedures involved, and the information sheet.

a) Justification for collection and processing of personal data: Information on gender, age classes, and the experiences and interest of community members in community-based observing will be used to contact and recruit participants in relation to the community-based observing systems. Aside from this, personal information is not used in the project. No further personal data are collected.

b) Procedures: The project together with representatives of the local communities will identify who among the local communities may be interested in participating. The criteria will be the knowledge and experiences of the community members and their interest in participating. The aim is to obtain the participation of both men and women and different age classes (as they tend to use different natural resources and have knowledge about different topics). Information on gender, age, and their experiences and interest in community-based observing is not stored. **The project will comply with the EU directive on data protection** and with any updates it might receive during the lifetime of the project.

c) Templates of the information sheet/informed consent: The information sheet/ informed consent form that will be communicated to the participants is provided in Annex 2 in the Description of Action.

d) The participants will be compensated for the time they use on community-based observing. In each area, *every participant will be treated equally favourable and they will obtain the same daily compensation, irrespective of personal characteristics*, e.g. religion, gender or age. The project has set aside funding for compensating community members for lost work-time for the time they spend on community monitoring activities in the two focal communities of community-based observing in INTAROS: Longyearbyen in Svalbard and Disko Bay in Greenland.

3. Third countries. The ethical standards and guidelines of Horizon2020 will be rigorously applied, regardless of the country in which the research is carried out. For the activities in Svalbard, the project will need a research permission from the local authorities. The project will obtain the research permission from the local authorities in Svalbard before the start of the research. The project will keep these permissions available at any time if needed for the European Commission representatives or for ethics reviewers.

10. References

Cupola, P., Earley, S., & Henderson, D. (2014). DAMA-DMBOK2 Framework. Online at https://www.dama.org/sites/default/files/download/DAMA-DMBOK2-Framework-V2-20140317-FINAL.pdf (retrieved 17 Oct 2019).

Danielsen, F., M.K. Poulsen, M. Enghoff, and L. Iversen. 2020. Connecting Arctic Community-Based and Citizen Science Observations with Existing Recognized Databases. Deliverable 4.4. INTAROS.

Eaton, B., J. Gregory, B. Drach, K. Taylor and S. Hankin, NetCDF Climate and Forecast (CF) Metadata Conventions, Version 1.6, 5 December, 2011.

Eaton, B., J. Gregory, B. Drach, K. Taylor, S. Hankin, J. Blower, J. Caron, R. Signell, P. Bentley, G. Rappa, H. Höck, A. Pamment, M. Juckes, M. Raspaud, R. Horne, T. Whiteaker, D. Blodgett, C. Zender, and D. Lee. 2020. NetCDF Climate and Forecast (CF) Metadata Conventions Version 1.8.

Enghoff, M., N. Vronski, V. Shadrin, R. Sulyandziga, and F. Danielsen. 2019. INTAROS Community-Based Monitoring Capacity Development Process in Yakutia and Komi Republic, Arctic Russia. CSIPN, RIPOSR, NORDECO and INTAROS. Available at: http://www.intaros.eu/media/1650/process-report-yakutia-cbm-dec-2019-final.pdf

ESIP. 2020. Attribute Convention for Data Discovery 1-3. https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3

European Commission, 2016. H2020 Programme Guidelines on FAIR Data Management in Horizon 2010. Version 3.0. 26 July 2016.

Hamre, T., H. Sagen, S. Sandven, F. Danielsen, G. Ottersen, A. Beszczynska-Moller, A. Morvik, I. Schewe, and M. Enghoff, 2019. INTAROS Deliverable 1.6 Data Governance Framework (including updated Data Management Plan).

ISO, 2017. ISO 15836-1:2017 - Information and documentation - The Dublin Core metadata element set - Part 1: Core elements".

ISO, 2014. ISO 19115 Topic Category vocabulary. https://apps.usgs.gov/thesaurus/thesaurus-full.php?thcode=15

Jeddi, Z., P.H. Voss, M. B. Sørensen, F. Danielsen, T. Dahl-Jensen, T. B. Larsen, G. Nielsen, A. Hansen, P. Jakobsen, and P. O. Frederiksen. 2020. Citizen Seismology in the Arctic. Frontiers in Earth Science 8: 139.

Lygre, K. and H. Sagen, 2017. INTAROS Deliverable 8.2 Gender and diversity action plan.

OceanSITES, 2010. OceanSITES User's Manual NetCDF Conventions and Reference Tables. Version 1.2. June 29, 2010.

NOAA, 2015. NOAA Role vocabulary CI_RoleCode. https://data.noaa.gov/resources/iso19139/schema/resources/Codelist/gmxCodelists.xml#CI_RoleCode

NVS, 2018. Role vocabulary CI_RoleCode. http://vocab.nerc.ac.uk/collection/G04/current/

OceanSITES, 2020. OceanSITES Data Format Reference Manual, NetCDF Conventions and Reference Tables Version 1.4.

Olsen, L.M., G. Major, K. Shein, J. Scialdone, S. Ritz, T. Stevens, M. Morahan, A. Aleman, R. Vogel, S. Leicester, H. Weir, M. Meaux, S. Grebas, C.Solomon, M. Holland, T. Northcutt, R. A. Restrepo, R. Bilodeau, 2013. NASA/Global Change Master Directory (GCMD) Earth Science Keywords. Version 8.0.0.0.0.

Sandven, S. et al., forthcoming. INTAROS Deliverable D1.10 Roadmap for sustainable Arctic Observing System. Yamakawa, A., Dushaw, B., Sagen, H. and T. Hamre (2018). Data standardization for long-term underwater acoustic observation - Ocean temperature and ambient noise data in Fram Strait, NERSC Technical Report No. 391, Version 1.0, Revised March 2019.

Appendix A. Template for dataset descriptions.

Table 1. Template for describing datasets in the INTAROS Data Catalogue.

Description	Value
Title:	Short text naming the dataset and outlining its content.
Abstract:	Short description of contents of dataset, how data was collected or generated, what processing steps data have undergone, quality control procedures applied, estimated uncertainty.
Parameter name(s):	List of parameters contained in the dataset.
Project/Program name(s):	Name of project(s)/program(s) that supported data acquisition.
Observing system:	Name of observing system that collected the data.
Tags:	Keywords associated with the dataset. Supports fast search.
License:	License under which the dataset is made available.
Organisation:	Name of the organisation owning the dataset.
Source:	Online access point for dataset, e.g. URL to a Thredds server.
Version:	Version of the dataset.
Principal Investigator:	Name of principal investigator(s).
PI e-mail:	Email address of principal investigator(s).
Data Curator:	Name of data curator for the dataset.
Data Curator E-mail:	E-mail address of the data curator.

Appendix B. Information and consent form for participants in community-based observing

Monitoring in (..name of area..) of natural resources by local people for improved management

Background. The climate is changing. Many people in the Arctic face huge challenges. They rely on natural resources for food, income or both. Maintaining life requires observation of the environment.

Scientific knowledge of the environment is incomplete. Scientific monitoring in the Arctic is difficult. Local herders, fishers, hunters and other environmentally interested people observe the environment all year-round. Their observations and knowledge are, however, not consistently quantified, analyzed, or used for resource management.

INTAROS is a new international project. The project aims at developing an Arctic observation system to improve resource management. The observation system will be based on both community members' and scientists' monitoring. With support from the European Union, the project runs for 5 years from 2016 to 2021. This form is about the community-based monitoring activities of INTAROS. If you are interested in participating in the community-based monitoring activities of INTAROS, it is necessary for us that you fill out the form.

What. A simple system for community-based monitoring by herders, fishers and hunters of:

- Animals that you hunt (such as geese, ducks, foxes),
- Attacks by predators,
- Fishing activities and fishing methods,
- Quality of pasture and reindeer conditions in your area,
- Use of resources in your area by people from within and outside community,
- Changes in climate and the environment around you (snow, ice, pollution)

Why. Your observations, when regularly collected and shared, can be used to influence the way resources are being used in your area. To improve your livelihoods. To strengthen your rights to the use of the land. Your knowledge is important.
Examples of results may be:

- Better hunting regulations for animals that you hunt (such as geese and other)
- Better management of predators,
- Improved and more sustained access to fish,
- Improved addressing of pollution,
- Better addressing of challenges to management of pasture,
- Better acknowledgement of the rights of your own community to use of your land

How. Five steps:

1. The most experienced and interested herders, fishers and hunters establish a community monitoring group.
2. Notebooks. You record observations of natural resources during field trips

3. You summarize your observations in a summary format at meetings in the group every 3-month, you analyse trends, discuss challenges and management initiatives
4. You provide your summarized information to (..organisation..) and authorities
5. You present key observations at community meetings one time each year

When. Monitoring should only be done as part of your routine herding, hunting or fishing activities. After every field trip you note your observations in a calendar. Every three month you meet with other members of the community monitoring group. You discuss and agree on trends in natural resources. If you want, you propose management actions to (..institution..) and the authorities.

Who. Any local person interested in natural resources of their areas in (..municipality..) can participate. People and communities will decide on their own if they see a benefit in this and if they want to participate in the project. They will decide what they want to monitor. Participation is on a voluntary basis and people are not paid to do the monitoring; they should do it because they think it may help them sustain their resource use.

Questions:

1. Do you agree to participate in this activity? _____
2. Do you agree to be contacted later on, in the context of INTAROS? _____

Personal data will be treated with confidentiality. Participation will be entirely voluntary. Participants in community-based monitoring are free to leave the activity at any time. Just inform the project's contact person.

More information: (..contact person in the local area, institution, telephone numbers..).

Appendix C. INTAROS model questionnaire

Template for model descriptions with examples.

Table 2. Template for describing of models used in the INTAROS project and two examples.

Description	Value
Model name:	Full name and abbreviation
Brief description of model:	Short description of the model.
Use/application in INTAROS:	Short description of its use/application in the project.
Used in these INTAROS WP(s) and task(s):	Identify which task(s) the model is used in.
Theme(s) (ocean, atmosphere, sea ice, glaciology, terrestrial, communities):	List the theme(s) in which the model is used within the project.
Geographical area:	Name the regions the model cover in the project.
Grid size and time step:	List the spatial and temporal resolution
Driving forces (input data and forcing):	Describe the input data and forcing fields used.
Model period (years, season…):	Time period(s) and type of simulation (e.g. hindcast).
Model used in other models within INTAROS:	Yes/No. If yes, name the other model(s).
References (maximum 3):	List papers or reports describing the model or analysis of model results.
Additional comments:	Other relevant information.
Contact person:	Name and e-mail of contact person for the model description.
Contact organisation	Name of the organisation of the contact person.

Description	Value
Model name:	NORWECOM.E2E
Brief description of model:	Module-based End-to-end model where an NPZD model is coupled to IBMs for some selected main species, an ocean acidification module and a contaminant module.
Use/application in INTAROS:	Indices and monitoring program evaluation in the Barents Sea
Used in these INTAROS WP(s) and task(s):	Tasks 6.2 and 6.8
Theme(s) (ocean, atmosphere, sea ice, glaciology, terrestrial, communities):	Ocean
Geographical area:	Barents and Nordic seas
Grid size and time step:	10-20 km, 1 hour

Driving forces (input data and forcing):	Atmosphere (wind and short wave radiation), ocean (currents, water level, salinity, temperature, ice)
Model period (years, season…):	1995-2017 (hindcast), 2006-2070 (downscaled climate simulations)
Model used in other models within INTAROS:	No, but model output has been registered in the INTAROS data portal, published at nmdc.no and made available on a ftp server
References (maximum 3):	Hjøllo, S. et al. 2012. Modelling secondary production in the Norwegian Sea with a fully coupled physical/primary production/ individual-based Calanus finmarchicus model system. Marine Biology Research, 8: 508–526 Utne, K. et al. 2012. Estimating consumption of Calanus finmarchicus by planktivorous fish in the Norwegian sea using a fully coupled 3d model system. Marine Biology Research, 8: 527–547
Additional comments:	
Contact person:	Morten D Skogen
Contact organisation	Institute of Marine Research, Norway

Description	**Value**
Model name:	ElmerIce
Brief description of model:	Parallel finite-element model based on the open-source multi-physics code Elmer.
Use/application in INTAROS:	Model-based demonstration of calculations of ice discharge from selected glaciers to the ocean, aimed to predict the contribution of glaciers to sea level rise.
Used in these INTAROS WP(s) and task(s):	Tasks 6.4.3 and 6.4.4
Theme(s) (ocean, atmosphere, sea ice, glaciology, terrestrial, communities):	Glaciology
Geographical area:	Northwest of Greenland and South of Svalbard
Grid size and time step:	50-100 m, 1 day-1 week
Driving forces (input data and forcing):	Glacier geometry, Surface mass balance, Surface velocity, submarine melt rate
Model period (years, season…):	Melt season
Model used in other models within INTAROS:	No
References (maximum 3):	Gagliardini, O. et al. 201). Capabilities and performance of Elmer/Ice, a new-generation ice sheet model. Geosci. Model Dev. 6, 1299–1318. Otero J. et al. 2017. Modeling the Controls on the Front Position of a Tidewater Glacier in Svalbard. Front. Earth Sci. 5:29.
Additional comments:	
Contact person:	Jaime Otero
Contact organisation	Universidad Politécnica de Madrid

Appendix D. Template for planning data management

The following template for planning data management and the data delivery chain was developed for the UAK 2021 Research School north of Svalbard.

Data management and Data delivery chain

Cruise name:
Ship:
Cruise leader:
Responsible Scientist:

Photo of instrument(s) or map of e.g. ice stations or CTD section.

[page break]

Collection *and download (1 page)*

What type of raw data and meta data is collected?

Instruments	Data parameters	Data format	Meta data format	Comments

Where is the downloaded raw data stored?

Where is the meta data stored?

Securing data during the cruise:
Describe your backup plan to ensure data is not lost during and immediately after the cruise is completed.

Securing raw data for longer term preservation:

Responsible for data collection and download:

[page break]

Processing and quality control (Level 1,2) (max 1 page)

What parameters are produced?

Pre-processing steps of do you use including conversion algorithms and averaging?

Quality control procedure:

[page break]

Formatting processed data and meta data (max 1 page)

Which meta data template do you use?

Prepare metadata record and where is the metadata form stored?

What is the level 1/2 data format used? (e.g. NetCDF, wav, xls)

Responsible for formatting data:
Responsible for collecting and formatting metadata:

[page break]

Store and publish (max 1 page)

What are the intellectual property rights for the data products?
Open data, Creative Commons license, copyright retained:

Choose a repository for long term preservation of data products (long-term plan)

What are the steps in publication of data and meta data?

Responsible for storing and publishing of data:
Responsible institution for long-term storage:

Appendix E. Metadata template for timeseries of ocean mooring data

The following templates for global and variable metadata are developed for ***point measurements at fixed depths*** and ***profile measurements at fixed depths*** on ocean moorings in the INTAROS project. This includes, among others, data from the following sensors used in INTAROS:

- microCAT (SBE37; temperature, salinity, pressure) [point data]
- RBR Concerto 3 (CTD; temperature, salinity, pressure) [point data]
- RBR Solo/Duet (temperature(Solo), temperature+pressure(Duet)) [point data]
- RBR T.ODO (temperature, dissolved oxygen) [point data]
- AADI SeaGuard SW (current meter; temperature, salinity, pressure, dissolved oxygen, turbidity) [point data]
- SAMI-CO2 (Ocean CO2 Sensor; pCO2) [point data]
- microAURAL (acoustic hydrophone) [point data]
- ADCP (different instruments, e,g, TRDI LR-ADCP, Signature 250) [profile data]
- McLane Moored Profiler (with different oceanographic sensors) [profile data]

The templates are based on the following metadata standards and vocabularies:

- Climate and Forecast (CF) convention, version 1.8 (Eaton et al., 2020)
- Attribute Convention for Dataset Discovery (ACDD), version 1.3 (ESIP, 2020)
- OceanSITES Data Format Reference Manual, NetCDF Conventions and Reference Tables, Version 1.4 (OceanSITES, 2020).
- NERC Vocabulary Server (NVS) Role vocabulary CI_RoleCode (NVS, 2018)
- NOAA Role vocabulary (ISO19139) (NOOA, 2015)
- ISO 19115 Topic Category vocabulary (ISO, 2014)

The metadata templates are defined by the INTAROS Ocean Mooring Working Group. Members of this working group that have contributed to this joint specification of metadata for ocean mooring data include (in alphabetical order):

- Agnieszka Beszczynska-Moller (IOPAN) [domain expert]
- Angelica Renner (IMR) [domain expert]
- Arnfinn Morvik (IMR) [data manager]
- Asuka Yamakawa (IMR) [data manager]
- Christophe Herbaut (CNRS) [domain expert]
- Delphine Mathias (CNRS) [domain expert]
- Espen Storheim (NERSC) [domain expert]
- Gaetan Richard (CNRS) [domain expert]
- Hanne Sagen (NERSC) [domain expert]
- Marie-Noelle Houssais (CNRS) [domain expert]
- Nicholas Roden (University of Bergen) [domain expert]
- Torill Hamre (NERSC) [data manager]
- Truls Johannessen (University of Bergen) [domain expert]

The following global attributes are ***mandatory*** for both point data and profile data:
(Example values in bold must be included exactly as shown.)

Name	Convention	Description	Example
title	CF	A succinct description of what is in the dataset.	Real time CIS-1 Mooring Temperatures
id	ACDD	An identifier for the data set, provided by and unique within its naming authority. The combination of the "naming authority" and the "id" should be globally unique, but the id can be globally unique by itself also. IDs can be URLs, URNs, DOIs, meaningful text strings, a local key, or any other unique string of characters. The id should not include white space characters.Each institute decide IDs	OS_CIS1_200502_TS
source	CF	Method of production of the original Data. If it is observational, source should characterize it	"surface observation" or "radiosonde" or "(10) hydrophones"
summary	ACDD	A paragraph describing the dataset, analogous to an abstract for a paper.	
institution	CF	Specifies where the original data was produced. Add the country in paranthesis.	Nansen Environmental and Remote Sensing Center (Norway) Institute of Oceanology Polish Academy of Sciences (Poland)
contact	CMEMS	E-mail address of support desk (PI and contributor emails are in other metadata fields)	cmems-service@imr.no
keywords	ACDD	A comma-separated list of key words and/or phrases. Keywords may be common words or phrases, terms from a controlled vocabulary (GCMD is often used), or URIs for terms from a controlled vocabulary	Earth Science > Oceans > Ocean Temperature > Water Temperature; Earth Science > Oceans > Salinity/Density > Conductivity
keywords_vocabulary	ACDD	If you are using a controlled vocabulary for the words/phrases in your "keywords" attribute, this is the unique name or identifier of the vocabulary from which keywords are taken. If more than one keyword vocabulary is used, each may be presented with a prefix and a following comma, so that keywords may optionally be prefixed with the controlled vocabulary key.	**GCMD:GCMD Keywords**
featureType	CF	Description of a single feature with this discrete sampling geometry	**timeSeries**
geospatial_lat_min	CF	Describes a simple lower/upper latitude limit; may be part of a 2- or 3-dimensional bounding region. Geospatial_lat_min specifies the	79.995
geospatial_lat_max	CF		79.995

Name	Convention	Description	Example
		southernmost latitude covered by the dataset. (in decimal)	
geospatial_lon_min	CF	Describes a simple longitude limit; may be part of a 2- or 3-dimensional bounding region.	4.723
geospatial_lon_max	CF		4.723
geospatial_vertical_min	CF	Describes the numerically larger vertical limit; may be part of a 2- or 3-dimensional bounding region. See geospatial_vertical_positive and geospatial_vertical_units.	229
geospatial_vertical_max	CF		229
geospatial_vertical_positive	CF	One of 'up' or 'down'.	**down**
time_coverage_start	CF	Describes the time of the first data point in the data set. Use the ISO 8601:2004 date format, preferably the extended format as recommended in the Attribute Content Guidance section. (a string according to the ISO8601 standard: "YYYY-MM-DDThh:mm:ssZ")	2014-01-01T00:00:00Z
time_coverage_end	CF		2014-12-31T00:00:00Z
Conventions	CF	Name of the conventions followed by the dataset.	**CF-1.8; ACDD-1.3**
license	ACDD	Provide the URL to a standard or specific license, enter "Freely Distributed" or "None", or describe any restrictions to data access and distribution in free text.	**https://creativecommons.org/licenses/by/4.0/**
project	ACDD	The name of the project(s) principally responsible for originating this data. Multiple projects can be separated by commas, as described under Attribute Content Guidelines. Examples: 'PATMOS-X', 'Extended Continental Shelf Project'.	**INTAROS**, XXX, XXXX
principal_investigator	OceanSITES	Name of the person responsible for the scientific project that produced the data contained in the file. If needed, use a comma-separated list. Participating investigators e.g. adding interdisciplinary studies to an existing site/cruise should specify the PI of the site/cruise, and use the 'creator' attributes to identify themselves.	
principal_investigator_email	OceanSITES	Email address of the project lead for the project that produced the data contained in the file. If needed, use a comma-separated list.	
cruise_id	INTAROS	Unique identifer of the cruise(s) where the data have been collected, or moorings deployed/recovered. Separate cruise names with semi-colon.	INTAROS-2018-KVS
cruise_responsible	INTAROS	Principle Investigator(s) of the cruise(s), and their organisation.	Hanne Sagen

Name	Convention	Description	Example
		Names, separated by semi-colon if multiple cruises.	
cruise_responsible_email	INTAROS		hanne.sagen@nersc.no
acknowledgement	ACDD	A place to acknowledge various types of support for the project that produced this data.	These data were made freely available by NERSC under projects Arctic Ocean under Melting Ice - UNDER-ICE (project no. 226373), and **INTAROS (GA No. 727890)**
citation	OceanSITES	The citation to be used in publications using the dataset;	Melissa Chierici, Elizabeth Jones, Helene Hodal Lødemel (2019) Interannual variability of the marine CO2 system and nutrients in the Barents Sea from 2011 to 2017 https://doi.org/10.21335/NMDC-1738969988
date_update	SDN	Timestamp specifying when the contents (i.e. its attributes and/or values) of the file were last changed Time is specified as a string according to the ISO8601 standard: "YYYY-MM-DDThh:mm:ssZ"	2021-02-16T13:00:15Z
standard_name_vocabulary	ACDD	The name and version of the controlled vocabulary from which variable standard names are taken. (Values for any standard_name attribute must come from the CF Standard Names vocabulary for the data file or product to comply with CF.)	'CF Standard Name Table v77'

The following global attributes are ***recommended*** for both point data and profile data: (Example values in bold must be included exactly as shown.)

Name	Convention	Description	Example
platform_code	OceanSITES	A unique platform code. This code is either assigned by the site PI (see principle_investigator below) or by the data provider. Required. (GDAC)	CIS-1 NERSC-4
data_mode	OceanSITES	Indicates if the file contains real-time, provisional or delayedmode quality controlled data. The list of valid data modes is in reference table 4. Required. (GDAC)	R / D
naming_authority	ACDD	The organization that provides the initial id for the dataset. The naming authority should be uniquely specified by this attribute. We	edu.ucar.unidata no.hi

Name	Convention	Description	Example
		recommend using reverse-DNS naming for the naming authority; URIs are also acceptable.	
netcdf_version	OceanSITES	NetCDF version used for the data set	netCDF-4 classic model
comment	CF	Miscellaneous information about the data or methods used to produce it	
source_platform_category_code	SDN	SeaDataNet vocabulary L06 (SeaVoX) reports platform categories, as a code and a label. E.g. '48' for a mooring.	**43** (subsurface mooring)
data_assembly_center	OceanSITES	Data Assembly Center (DAC) in charge of this data file. A partial list of the data assembly centers is in reference Table 5	IMR
contributor_name	ACDD	The name of any individuals, projects, or institutions that contributed to the creation of this data. May be presented as free text, or in a structured format compatible with conversion to ncML (e.g., insensitive to changes in whitespace, including end-of-line characters).	Lorenzo Corgnati; Carlo Mantovani
contributor_role	ACDD	The role of any individuals, projects, or institutions that contributed to the creation of this data. Multiple roles should be presented in the same order and number as the names in contributor_names.	metadata expert; HFR expert
contributor_email	ACDD	The email addresses of any individuals or institutions that contributed to the creation of this data, separated by semicolons.	lorenzo.corgnati@sp.ismar.cnr.it; carlo.mantovani@cnr.it
iso_topic_category	INSPIRE	A high-level classification scheme to help with grouping and topic-based search of spatial data resources.	**OCEANS**
creator_name	ACDD	The name of the person (or other creator type specified by the creator_type attribute) principally responsible for creating this data. Up to the partner to decide to include.	
creator_email	ACDD	Email of the person principally responsible for creating this data.	
creator_url	ACDD	The URL of the of the person (or other creator type specified by the creator_type attribute) principally responsible for creating this data.	
geospatial_vertical_units	ACDD	Units for the vertical axis described in "geospatial_vertical_min" and "geospatial_vertical_max" attributes. The default is EPSG:4979 (height above the ellipsoid, in meters); other vertical coordinate reference systems may be specified.	meters

Name	Convention	Description	Example
publisher_name	ACDD	The name of the person (or other entity specified by the publisher_type attribute) responsible for publishing the data file or product to users, with its current metadata and format.	European HFR Node
publisher_email	ACDD	The email address of the person (or other entity specified by the publisher_type attribute) responsible for publishing the data file or product to users, with its current metadata and format.	euhfrnode@azti.es
publisher_url	ACDD	The URL of the person (or other entity specified by the publisher_type attribute) responsible for publishing the data file or product to users, with its current metadata and format.	http://eurogoos.eu/

The following global attributes are ***suggested*** for both point data and profile data:
(Example values in bold must be included exactly as shown.)

Name	Convention	Description	Example
creator_type	ACDD	Specifies type of creator with one of the following: 'person', 'group', 'institution', or 'position'. If this attribute is not specified, the creator is assumed to be a person. Up to the partner to decide to include.	
format_version	OceanSITES	OceanSITES format version; may be 1.1, 1.3, 1.5.	1.0
update_interval	OceanSITES	Update interval for the file, in ISO 8601 Interval format: PnYnMnDTnHnM where elements that are 0 may be omitted. Use “void” for data that are not updated on a schedule. Used by inventory - software. (GDAC)	PT12H void
cdm_data_type	ACDD	Suggested by Unidata	**timeSeries**
publisher_type	ACDD	Specifies type of publisher with one of the following: 'person', 'group', 'institution', or 'position'. If this attribute is not specified, the publisher is assumed to be a person.	
publisher_institution	ACDD	The institution that presented the data file or equivalent product to users; should uniquely identify the institution. If publisher_type is institution, this should have the same value as publisher_name.	
cruise_info	INTAROS	Link to a cruise report or CSR that provide addtional information about the cruise. Separated by semi-colon if multiple documents.	

The variables should be defined as follows:

Name	data type	dimension	attribute	Example
instrument_id	char	(num_instruments, shortstr)	_FillValue = ' '	e.g. serial number, or 'microCAT 1'

instrument_owner	char	(num_instruments, shortstr)	_FillValue = ' '	e.g. 'IOPAN', 'IOPAN', 'NERSC', 'NERSC'
longitude	float32	(num_instruments)	units = "degrees_east" standard_name = "longitude" axis = "X"	NB: NO _FillValue attribute
latitude	float32	(num_instruments)	units= "degrees_north" standard_name = "latitude" axis = "Y"	NB: NO _FillValue attribute
depth	float32	(num_instruments)	units = "m" standard_name = "depth" axis = "Z" _FillValue= 9.96921E+36	nominal depth
instrument_description	char	(num_instruments, maxstr)	_FillValue = ' '	Instrument type, sensors and additional information can be detailed in this variable
data_processing	char	(num_instruments, maxstr)	_FillValue = ' '	e.g. "measurements averaged to hourly values after despiking and removing erronous values"
timeseries_id	char	(num_instruments, shortstr)	cf_role=”timeseries_id” _FillValue = ' '	e.g. NERSC4_11661 (<mooring name>_<serial number>)
time	double	(num_instruments, obs)	units = "days since 1950-01-01T00:00:00Z" standard_name = "time" axis = "T"	NB: NO _FillValue attribute
var (placeholder)	float32 or Int	(num_instruments, obs)	coordinates = "time latitude longitude depth instrument_name" standard_name = "sea_water_temperature" units = "degrees_C" _FillValue = 9.96921e+36f	e.g. measured depth, temperature
profile_var (placeholder)	float32 or Int	(num_instruments, obs, profiledim)	coordinates = "time latitude longitude depth instrument_name" standard_name = "sea_water_speed" units = "m/s" _FillValue = 9.96921e+36f	Variable holding a profile of e.g. currents for each timestep.
datafile_ref	char	(maxstr)	_FillValue = ' ' long_name = 'Unique identifier of the file(s) holding the timeseries of data.'	URL of a directory or catalogue of the data files, allowing download of

				individual files or all files as a package. Example: URL to a THREDDS data catalogue.

The following variables are used by all sensors collecting point measurements at fixed depths:

- instrument_id– providing a unique identifier for the sensor.
- instrument_owner – identifying the organisation owning the sensor.
- longitude, latitude, depth – defining the location and depth of the sensor.
- instrument_description, data_processing – providing technical information about the instrument and the processing that the data have gone through after acquisition.
- timeseries_id – providing a unique identifier for the dataset.

Time series of ***point measurements and profiles at fixed depths*** hold the time of each measurement in the variable named "time". Note that the dimensions of the time variable are num_instruments by obs, i.e. the number of instruments (sensors) mounted on the ocean mooring and the maximum number of measurements collected by any of these instruments. This means that each sensor on an ocean mooring can collect data at different times and have differing sampling frequencies. The "obs" dimension is the largest number of observations collected; sensors collecting fewer observations will pad their parameter variable(s) with fill values to mark that no data were collected for the last time steps.

Time series of ***point measurements at fixed depths*** holds the values of ***parameters measured*** in the variable named "var". E.g. data from a microCAT instrument will have variables named "temperature", "salinity", "pressure", all with the same dimensions. The dimensions of the "var" variable are the same as for the "time" variable.

Time series of ***profile measurements at fixed depths*** holds the values of ***parameters measured*** in the variable named "profile_var". E.g. data from an Signature 250 instrument could have variables named "sea_water_speed", "sea_ice_drift", "sea_ice_draft", all with the same dimensions. The dimensions of the "profile_var" are num_instruments, obs, profiledim, representing the number of profiling instruments on the ocean mooring, the maximum number of observations for these instruments, and the number of observations in each of the profiles measured at each time step.

The "datafile_ref" variable is used for datasets where the values for the observations and their time steps are stored in a separate file. The reason for not storing the data in the NetCDF file together with the metadata is e.g. that another file format is substantially more space efficient or that another format is dominant in a scientific community. An example of such a format is WAV, which is the most commonly used format in ocean acoustics and can store large amounts of acoustic recordings very efficiently.

For fill values in NetCDF, we follow the CMEMS conventions, i.e.:

Attribute	Data type	Value
_FillValue	double	9.96920996838686E+36

	float	9.96921E+36
	int	-2147483647
	byte	-127
	char	" "

---------- END of DOCUMENT----------

INTAROS

This report is made under the project
Integrated Arctic Observation System (INTAROS)
funded by the European Commission Horizon 2020 program
Grant Agreement no. 727890.

Project partners: