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 5.9

Data Integrated from Existing Repositories V2

Start date of project:	01 December 2016	Duration:	60 months
Due date of deliverable:	31 May 2021	Actual submission date:	16 December 2021
Lead beneficiary for preparing the deliverable:		TDUE	
Other contributing partners		AWI, ARMINES, IMR, NUIM	
Person-months used to produce deliverable:		1.9 pm	

Authors: Ingo Schewe (AWI), Hervé Caumont (TDUE), Fabien Ors (ARMINES), Arnfinn Morvik (IMR), Peter Thorne (NUIM)

Version	DATE	CHANGE RECORDS	LEAD AUTHOR
0.1	06/01/2021	Initial draft - Table of content	Hervé Caumont
0.2	01/03/2021	Second draft	Ingo Schewe
0.3	15/04/2021	1st Review	Pedro Goncalves
0.4	21/04/2021	Third draft	Ingo Schewe Hervé Caumont
1.0	22/04/2021	Final for internal review	Ingo Schewe
2.0	25/05/2021	Review by coordination team	Hanne Sagen
2.1	02/07/2021	Updates	Ingo Schewe Hervé Caumont Fabien Ors
2.2	12/07/2021	Second review by coordination team	Hanne Sagen Torill Hamre
2.3	15/12/2021	Final for EC delivery	Hervé Caumont
2.4	16/12/2021	Final review and submission	Toril Hamre

Approval X	Date: 16/12/2021	Sign.  Stein Sandven

CONTRIBUTING PARTNERS					
No	***Beneficiary***	***PM***	***No***	***Beneficiary***	***PM***
1	NERSC		24	TDUE	0.8
2	UiB		25	GINR	
3	IMR	0.1	26	UNEXE	
4	MISU		27	NIVA	
5	AWI	0.8	28	CNRS	
6	IOPAN		29	U Helsinki	
7	DTU		30	GFZ	
8	AU		31	ARMINES	0.1
9	GEUS		32	IGPAN	
10	FMI		33	U SLASKI	
11	UNIS		34	BSC	
12	NORDECO		35	DNV GL	
13	SMHI		36	RIHMI-WDC	
14	USFD		37	NIERSC	
15	NUIM	0.1	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	

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	

EXECUTIVE SUMMARY

The INTAROS project develops an integrated Arctic Observation System (iAOS) by extending, improving and unifying existing systems in the different regions of the Arctic. Within INTAROS, WP5 (Data integration and management) is tasked with designing and implementing evolutions of the cloud platform with geo statistical tools for services as part of the iAOS. Partners in Task 5.3 are in charge of data integration from existing repositories, along with metadata, so that it is findable and exploitable by iAOS applications, and well referenced by the iAOS Portal and the INTAROS Data Catalogue. In particular, identification of the online data access points and the related protocols are offered to make the data accessible for use in the iAOS Showcase application services, as well as from other applications in WP6 which can utilize these data.

Deliverable D5.9 describes the work done in Task 5.3 with respect to identifying and clarifying the technical requirements for data integration of existing data repositories, and gives examples of such integrations. Section 2.1 provides information on the growth of registered datasets in the INTAROS Data Catalogue as of 22nd April 2021 compared to 1st December 2019. A summary of to what extent the original plans for integration of data repositories in Task 5.3 have been achieved is given in Section 2.2. AWI/Terradue conducted an email campaign to partners in order to identify which datasets are planned to be registered in the INTAROS Data Catalogue (Section 2.3). Section 3 describes relevant data repositories which in some cases are linked to the iAOS cloud platform and in other cases listed as potential source of data (when an actual integration into data processing applications was performed as part of other INTAROS tasks, the mention “iAOS Application” is added along with the reference to it). Section 4 describes the data sources for selected iAOS Showcase applications, the data needed by application developers, and how the foreseen output data products could be used.

In particular, this includes the integration of data repositories behind machine-to-machine interfaces, using an API (Application Programming Interface). Effort was made to identify iAOS Cloud Platform features related to data discovery and access useful for the showcase applications from WP6 partners. We finally describe the selected iAOS Showcase applications, how the data is delivered to these application developers, and how the foreseen output data products could be used.

Table of Contents

Acronyms and abbreviations

API	Application Programming Interface
CMEMS	Copernicus Marine Environment Monitoring Service
DIAS	Data and Information Access Service
EGI	European Grid Infrastructure
EO	Earth Observation
EPOS	European Plate Observing System
FRAM	FRontiers in Arctic marine Monitoring project
HTTP	HyperText Transfer Protocol
HYCOS	Hydrological Cycle Observing System
HYPE	Hydrological Predictions model
iAOS	integrated Arctic Observation System
NERSC	Nansen Environmental and Remote Sensing Center
NetCDF	Network Common Data Format
OGC	Open Geospatial Consortium
OpenDAP	Open Data Access Protocol
OSDD	OpenSearch Description Document
ORDP	Open Research Data Pilot
OWS	OGC Web Services
THREDDS	Thematic Real-time Environmental Distributed Data Services
CS-W	Catalog Service for the Web
WPS	Web Processing Service
XBT	Expendable Bathythermograph
XCTD	Expendable Conductivity/Temperature/Depth

1. Introduction

Existing observing systems, data repositories and infrastructure available from partners and collaborators are the building blocks of the iAOS. The observing systems and data repositories were originally assessed in WP2 (Exploitation of existing observing systems) at the beginning of the project. The aim of INTAROS was to exploit existing e-infrastructure for data storage and preservation, capitalizing on selected research data infrastructures in Europe, US, Canada and Asia, that hold environmental data for the Arctic.

This report follows the requirements of an Open Research Data Pilot (ORDP) as defined by The European Commission (EC)[1] and presents the status of the data registration into the iAOS INTAROS data catalogue and of the integration tools provided by the iAOS Cloud Platform.

1.1. Goals for the initial period (end of year 2)

The first phase of Task 5.3 focused on the creation of a framework for making data sources available to the iAOS Cloud Platform as online repositories. An assessment of technical metadata available from identified data led to general requirements for how data integration from existing repositories into the iAOS must be met. The initial process was described in deliverable D5.3 (Data integration from existing repositories V1) using xls sheets extracting information from the INTAROS survey (e.g. summarized in D2.10, D2.11). In Task 5.6 (see D5.4 iAOS portal with user manual-V1) a web-based data portal was established and populated with input from the individual partners. The partners collecting or generating data subsequently deposited the data into existing repositories and registered the data into the INTAROS Data Catalogue (WP2, WP3 and WP4). This data catalogue provides a joint access point to the data derived by the INTAROS project.

1.2. Focus of the second phase of data integration

In the second phase the following activities relevant for Task 5.3 were carried out:

1. Registration of new datasets into the INTAROS data catalogue: At the time of preparation of this report, 135 records have been registered by 37 different organizations within the INTAROS consortium. This activity was carried out by partners generating data in WP2,3,4. In the next period the registration is expected from WP3, WP6, including data generated by the iAOS Showcase applications developed from WP5/6 collaboration. The partners are supported by NERSC in Task 5.6.
2. Setup of a Task Force of 14 project partners, focused on the exchange of experience on using OpenDAP and NetCDF, was initiated in July 2018. The Task Force aimed at making the knowledge of experienced partners available to less experienced partners. A sequence of work sessions (email exchanges and calls) led by TERRADUE and AWI have been organised between September and November 2018, concluding on the 22nd November 2018 with a visio conference session gathering 12 participants (partners from Terradue, AWI, NERSC, FMI, IMR, ARMINES, University of. Bremen, University of Silesia in Katowice, and IGF PAN) to share views, present activities and contribute to the preparation of a NetCDF / OpenDAP session for the WP5 workshop, held as part of the 2nd INTAROS GA in Bremen, January 2019.

[1] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

3. "Data Delivery" Workshop session on competence building among the INTAROS partners, for the ingestion of data products into OPeNDAP servers. The session was conducted as part of the General Assembly in January 2019. The session's agenda was initially established to provide both a broad overview of the operational objectives with online data distribution, and some hands-on guidance:

    a. **From Data Collection to Long-Term Preservation and Use: Arctic data as part of a global system**
    (Peter L. Pulsifer; NERSC, Torill Hamre)
    b. **OpenDAP Use Cases - how-to essentials**
    (Terradue, Hervé Caumont)
    c. **Example for querying a remote OpenDAP server for Sea-Ice data-collections**
    (UB, Georg Heygster / Terradue, Hervé Caumont)
    d. **Data integration via IMR data model**
    (IMR, Arnfinn Morvik / ARMINES, Fabien Ors)
    e. Tools for data-mining in an online-repository
    (AWI/ PANGAEA, Ingo Schewe) - Not presented (cancelled)
    f. Using Geoserver to provide gridded datasets to iAOS
    (FMI, Mwaba Hiltunen) - Not presented (cancelled)

Through these activities listed above, the goal of the task has been to exemplify how the data flows within iAOS work from measurements to data products targeted at issues of societal importance for indigenous and local communities in the Arctic, for Europe and on global scale.

Besides the pure data repository integration, a key objective of Task 5.3 in the final phase of the INTAROS project was to focus more on those datasets of particular interest for use within the WP6 Applications of iAOS towards Stakeholders.

2. Status of existing repositories integrated into the iAOS

2.1. INTAROS Data Catalogue

A web based INTAROS Data catalogue has been established (Task 5.6), and it is mandatory for the partners to register all data produced with support from INTAROS. The Data produced covers in situ measurements, Citizen Science, Community Based Monitoring, remote sensing products, modeling products and results from showcases using the cloud platform or other applications.

At the time of document preparation (April 2021), a total of **135** records (in brackets: pre-announced additional data sources to be provided) were listed in the Data Catalogue, provided by 37 different partner-organizations.

Partner Institution	# datasets
Nansen Environmental and Remote Sensing Center (NERSC)	15
University of Bergen (UiB)	3
Institute of Marine Research (IMR)	13
Department of Meteorology Stockholm University (MISU)	7(+2)
Alfred-Wegener-Institute (AWI)	6
Institute of Oceanology of the Polish Academy of Sciences (IOPAN)	1(+1)
Technical University of Denmark (DTU)	6
Aarhus University (AU)	3

Partner Institution	# datasets
The University of Sheffield (USFD)	2
Institut Francais de Recherche pour L'exploitation de la Mer (IFREMER)	3
University Bremen (UB)	4
University Hamburg (UHAM)	1
Norwegian Research Institute AS (NORCE)	1
Norsk institutt for vannforskning (NIVA)	2
National Center for Scientific Research (CNRS)	4
University of Helsinki (U Helsinki)	2

Geological Survey of Denmark and Greenland (GEUS)	3	Deutsches Geoforschungszentrum (GFZ)	2
Finnish Meteorological Institute (FMI)	6(+1)	Institute of Geophysics Polish Academy of Sciences (IGPAN)	2
Nordic Agency for Development and Ecology (NORDECO)	6	University of Silesia (U SLASKI)	8(+1)
PISUNA Database	1	All Russian Research Institute for Hydro-meteorological Information (RIHMI-WDC)	23
Yukon River Inter-Tribal Watershed Council (YRTWC)	1	University of Alaska (UAF)	1
Centre for Support of Indigenous Peoples of the North (CSIPN)	1	The Institute of Remote Sensing and Digital Earth (RADI)	(+1)
Swedish Meteorological and Hydrological Institute (SMHI)	1		

After the establishment of the INTAROS Data Catalog from October 2018 to the end of 2019, 51 datasets were registered, providing a first package of data product candidates for integration as part of the iAOS Showcase applications. Thereafter, another **42** data sources were added as of end November 2020. From this point until the time of 22 April 2021, another group of **42** datasets was made available by partners. A detailed list of the datasets added to the INTAROS Data Catalogue in the last two development phases can be found in Annex 1 (Time Period: December 2019-November 2020) and Annex 2 (Time Period: December 2020-Spring 2021).

2.2. Data repositories accessible for data integration into applications

A review of the extent to which the initial Task 5.3 goals were achieved, in terms of integration of data repositories is summarized below.

These goals are defined in the INTAROS Description of Action (DoA) in order to support the demonstrations in WP6, by listing the datasets considered as relevant for the WP6 activities, and to be integrated from, among others:

- Sentinel-1 and Sentinel-2 data through the Thematic Exploitation Platform (TEP) tools (TERRADUE)
- Selected sea-ice parameters from CERSAT (IFREMER; from WP2)

- Selected satellite-based products from CMEMS, the Copernicus Marine Environment Monitoring Service (NERSC)
- Multidisciplinary data from PANGAEA, among others, oceanographic and biochemical data (AWI)
- Multidisciplinary data from SeaDataNet, among others, oceanographic and biological data (IMR)
- Multidisciplinary data from NMDC (Norwegian Marine Data Centre), e.g. oceanographic data (IMR)
- Ecosystem data from NORWECOM and ATLANTIS models for the Barents Sea. (IMR)
- Fish stock and other relevant data from ICES (IMR)
- Glaciological and seismological data from EPOS (GEUS)
- Carbon and other greenhouse gases data from ICOS (UIB; from WP2)
- Atmospheric data, e.g. aerosols and clouds, from the ACTRIS database (FMI)
- Atmospheric data, e.g. radiosonde measurements of pressure, temperature, humidity, and wind from the GRUAN network (NUIM)
- Community-based observations, e.g. observations of temperature, wind or seismic activities from the ELOKA network (NORDECO; from WP4)
- Glider data from the Coriolis GDAC (CNRS-LOCEAN; from WP3)

The task 5.3 therefore performed checks of the actual maturity level of data sources, and facilitated any technical action to reach satisfying maturity, so that the datasets are ready for integration by the iAOS tools and services (especially in support of the WP6 applications). The review in this section highlights which data repositories were actually made available for exploitation (with identified access protocols, client tools and platform services) within INTAROS.

Copernicus Sentinel-1 and Sentinel-2
Source: https://scihub.esa.int/
The INTAROS partner Terradue is curating a catalog of Copernicus Sentinel datasets supporting a data processing capability to move processing jobs where the data is.

The processing is done either physically on the same Cloud infrastructure, like a selected Copernicus DIAS, or connectivity-wise, considering high-bandwidth networking infrastructures like the EGI.eu Federated Cloud.

The capability was initially developed and used by Terradue to support the ESA-initiated Geohazards Thematic Exploitation Platform (GEP) and Hydrology Thematic Exploitation Platform (HEP). It is available to the INTAROS partners for enabling iAOS users to extract time series of Sentinel data products and process these as part of the iAOS cloud platform.

Data access is offered through the OGC OpenSearch protocol at:

- Backend data catalog providing data access links to Copernicus Sentinel-1 and -2: https://catalog.terradue.com

CERSAT
Source: http://cersat.ifremer.fr/
The INTAROS partner IFREMER provided three datasets for sea ice concentration and displacement from satellite data in various resolutions. These open access datasets are

registered in the INTAROS Data Catalogue with links to the CERSAT data center.
Data access is offered through FTP, at:

- ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/psi-drift/data/arctic/amsr2-merged/
- ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/psi-drift/data/arctic/merged-ascat-A-B-ssmi/
- ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/psi-concentration/data/arctic/daily/netcdf/

Copernicus Marine Service
Source: https://marine.copernicus.eu/
The Copernicus Marine Service is the marine component of the Copernicus Programme of the European Union. In INTAROS we have focused on integrating a subset of satellite-based products from the Copernicus Marine Environment Monitoring Service Copernicus - Marine environment monitoring service (CMEMS). As part of the iAOS Portal development in Task 5.6, NERSC harvested metadata for a limited set of ice charts for the product "Arctic Ocean - Sea Ice Concentration Charts - Svalbard and Greenland". This product was also used in the service TS-SIC developed in Task 5.5.

CMEMS offers access to a wide range of data (satellite based, model and in situ products) through their data catalogue that can be accessed at:

- https://resources.marine.copernicus.eu/?option=com_csw&task=results

PANGAEA
Source: https://www.pangaea.de/
Multidisciplinary data, among others, oceanographic and biochemical data
AWI provided a total of 6 data collections and registered these in the INTAROS catalogue:

- UDASH (Unified Database for Arctic and Subarctic Hydrography)
- DTM (Digital terrain model of the central Fram Strait)
- Biogeochemical parameters – Hausgarten
- Benthic Oxygen flux – FramStrait
- Biogenic particle flux – FRAM observatory
- Inorganic nutrients – Fram Strait

It is important to note that the links to the biogeochemical datasets are dynamic. This means that new datasets added to PANGAEA, e.g. through WP3 activities, are immediately available via the INTAROS data catalogue. A deeper integration into iAOS can also be realized with PANGAEA via direct API queries directly to the data warehouse. See further details in section 3.2.3.

Data access is offered through PANGAEA:

- UDASH:
  https://doi.pangaea.de/10.1594/PANGAEA.872931
- DTM:
  https://doi.pangaea.de/10.1594/PANGAEA.526589
- Hausgarten biogeochemical parameters:
  https://www.pangaea.de/?q=project%3AFRAM+OR+project%3AHAUSGARTEN
- Fram Strait benthic oxygen flux:
  https://www.pangaea.de/?q=project%3Ahausgarten+AND+parameter%3Aname%3A%22depth%2C+sediment%22+AND+parameter%3Aname%3Aoxygen
- FRAM observatory biogenic particle flux:
  https://www.pangaea.de/?q=parameter%3Aflux&f.device%5B%5D=Mooring+(long+time)&f.location%5B%5D=Arctic+Ocean&f.author%5B%5D=N%C3%B6thig%2C+Eva-Maria&maxlat=82&minlon=-15&maxlon=15&minlat=75

- Fram Strait inorganic nutrients: https://www.pangaea.de/?q=parameter%3Aname%3A%22depth%2C+water%22+parameter%3Aname%3Aphosphate+parameter%3Aname%3Asilicate+parameter%3Aname%3Anitrate+parameter%3Aname%3Anitrite&maxlat=82&minlon=-15&maxlon=15&minlat=75&mindate=1997-01-01

Norwegian Marine Data Centre

Source: https://nmdc.no/

Multidisciplinary data from SeaDataNet, among others, oceanographic and biological data. (collected and ingested by IMR)

CTD data collections over several years are provided from eight Research Vessels: Helmer Hanssen; Johan Hjort; G.O. Sars; Kristine Bonnevie; Sarsen; Håkon Mosby; Michael Sars and Jan Mayen. These data are used in iAOS Showcase 2 Barents Sea Multi-depth Temperature & Salinity Maps.

Multidisciplinary data from NMDC (Norwegian Marine Data Centre), e.g. oceanographic data. (collected and ingested by partners in NMDC)

From NMDC two oceanographic datasets from XCTD/XBT measurements have been provided.

For both collections, data access is offered through http://opendap1.nodc.no/opendap/physics/point/yearly/contents.html

Norwegian Institute of Marine Research

Source: https://www.hi.no/

Ecosystem data from NORWECOM and ATLANTIS models for the Barents Sea. (IMR)

A total of four major datasets have been published to date in connection with the NORWECOM simulations. One of these is a simulation of the overfishing of mackerel, herring or blue whiting for the period 2015-2025. Three other results come from three prognostic simulations of the distribution of fish species in the period 2010 - 2069.

Geological Survey of Denmark and Greenland

Source: https://www.geus.dk/

Glaciological and seismological data from Greenland (GEUS)

During the project, GEUS has prepared and published a total of five datasets of Greenland ice-related data. All five datasets are registered with data access links in the INTAROS Data Catalogue. These range from Greenland Freshwater Runoff and Ice Sheet solid ice discharge data and Ice velocity maps of the Greenland ice sheet margin based on Sentinel-1 data to data from a pilot community based seismic study for cryo- and tectonic seismological recordings.

Data access via GEUS DataVerse, ShakeNet and PROMICE:

- Greenland Freshwater Runoff: https://dataverse01.geus.dk/dataverse/freshwater
- Greenland Ice Sheet solid ice discharge: https://dataverse01.geus.dk/dataverse/ice_discharge
- Ice velocity maps of the Greenland ice sheet margin based on Sentinel-1 data: https://dataverse01.geus.dk/dataverse/Ice_velocity
- Citizen seismology program, Qeqertalik, Greenland: https://raspberryshake.net/stationview/
- PROMICE AWS data: http://promice.org/DataDownload.html

Finnish Meteorological Institute

Source: https://en.ilmatieteenlaitos.fi/

Atmospheric data, e.g. aerosols and clouds, from Sodankylä-Pallas observatory, following the data and metadata standards developed in ACTRIS

New atmospheric composition and cloud data sources in the framework of the collaboration with the ACTRIS community and infrastructures are provided through the INTAROS data catalogue.

Generally this work is going on, also with the contribution of the ACTRIS project, demonstrating also the good synergy between the two projects. In order to fill a geographically large gap in atmospheric data in Arctic Russia, also aerosol observations from Tiksi Hydrometeorological Observatory were added to the catalogue. In addition to this, we are creating Merged Observatory Data Files (MODFs) about Sodankylä-Pallas observatory, in the framework of the YOPPsiteMIP project (YOPP super-site Model Inter-comparison Project). In practice, data from Arctic supersites during YOPP Special Observing Periods are formatted and presented in a standardized way, common also to the NWP model outputs, with the highest possible resolution, to facilitate model validation and process studies.

Data access is offered through the following online data portals at:

- Arctic Russia, Tiksi site
  http://ebas.nilu.no/DataSets.aspx?stations=RU0100R&fromDate=1970-01-01&toDate=2021-12-31
- MODFs, Sodankylä-Pallas observatory
  https://litdb.fmi.fi/

GRUAN Network / Maynooth University (NUIM)
Source: https://www.gruan.org/
Atmospheric data, e.g. radiosonde measurements of pressure, temperature, humidity, and wind from the GRUAN network through the C3S Climate Data Store (CDS)

GRUAN data is now available via the CDS. Reference quality radiosonde measurements from the GRUAN network are available via the Copernicus Climate Change Service Data Store at https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-observations-gruan-reference-network?tab=overview.

This is potentially a little more user friendly than the NCEI ftp site. Datasets are available for sites at Barrow, Ny Alesund and Sodankyla in the Arctic region. All profiles that pass the GRUAN data processing are presented with up to 4 ascents per day per site (regular program at Barrow, campaign mode at other sites). All sites have used the Vaisala RS92 radiosonde and switched to the RS41. Presently data is only available for the RS92 but will become available once certification is completed for the RS41. Methods and documentation are given at www.gruan.org and in literature linked therefrom. GRUAN processing of these commercial sonde models is fully traceable to SI standards and each observation is presented alongside a quantified uncertainty estimate. Presented parameters are temperature, humidity, wind speed, wind direction and pressure. Data is served as CSV files via a web form and requests can also be submitted via API scripting.

Access to global land meteorological records is possible via the Copernicus Climate Change Service's data store. Datasets are available from several thousand stations within the Arctic domain. Data is available for monthly, daily and sub-daily (synoptic) aggregations from a broad range of sources. These data have been harmonised and are served via a web form interface that can be scripted. Data are available either as fully open or restricted to research and personal usage (consistent with WMO Resolution 40). Data are irregular and discontinuous in

both space and time. The record became progressively more complete until the 1970s with some recent period drop-off that may partially reflect reporting and collation latencies. Successive releases are incorporating new data sources including e.g. those secured under Copernicus collaboration agreements with countries such as Canada. Data rescue is also adding to early records with successive releases. A listing of all stations served and their periods of record is available via the documentation tab. Variables available vary by the timescale aggregation of the data and include: temperature, humidity, pressure, wind, precipitation, snowfall, snow water equivalent and snow depth. Data can be downloaded via a webform or user defined API scripts. The form allows simple rectangular domain definition. Requests are capped for performance presently.

Data access is offered through the Copernicus CDS portal at:

- https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-observations-surface-land?tab=overview

NORDECO
Source: various hosts (synthesis via https://catalog-intaros.nersc.no)
Community-based observations, e.g. observations of temperature, wind or seismic activities from the ELOKA network

NORDECO provided access to a total of sixteen different Citizen Science (CS) projects and Community-Based Monitoring (CBM) observations in conjunction with WP4 actions. These include subject areas such as: statistics of whale catches; Hunting harvest statistics; a web-based marine mammal photo platform; cloud observations; weather and sea ice conditions relevant to walrus and simple oceanographic observations made by Secchi-Disc. The data access links for the CS and CBM datasets are available at:

- https://catalog-intaros.nersc.no/dataset?q=CBM&sort=score+desc%2C+metadata_modified+desc

This is featuring, among others (excerpt):

- HappyWhale: https://happywhale.com/browse
- Sea Ice for Walrus Outlook: https://www.arcus.org/siwo/archive
- Pilot Whale Statistics: https://heimabeiti.fo/hagtol
- Piniarneq: https://www.sullissivik.gl/-/media/Sullissivik/Blanketter-og-pdf/Jagt_fangst_og_fiskeri/Piniarneq-2020-DA.ashx?la=da-DK
- GLOBE Observer: Clouds: https://observer.globe.gov/get-data/clouds-data
- Secchi Disk Study: https://www.playingwithdata.com/secchi-disk-project/

CNRS-LOCEAN
Source: http://www.coriolis.eu.org/About-Coriolis/Organisation
Glider and Argo float data from the Coriolis GDAC

In INTAROS, CNRS-LOCEAN has made available glider data from the Fram Strait, as well as Argo Float data monitoring the biogeochemical properties of the Baffin Bay (with the deployment of a fleet of bio-Argo floats dedicated to navigate in icy waters).

Data access is offered through the following online repositories:

- Glider, Taiga deployment
  https://www.seanoe.org/data/00403/51473/
- Argo floats
  http://www.coriolis.eu.org/Data-Products

2.3. Which data sources can still be expected

In April 2021, AWI and TERRADUE conducted an emailing campaign in preparation for this deliverable. The aim was to find out which additional datasets the partners are planning to make available (potentially for integration by the iAOS tools and services), and until the end of the project.

The list of contacted persons was defined out of the INTAROS interim reports, for the project partners having mentioned they made contributions to INTAROS through data provisioning activities in the course of WP3, 4 and 6.

The request was formulated in such a way that the provided responses to the survey were ready-made for the registration form available on the INTAROS Data Catalogue. In response to our request, we received feedback from a total of **5** partners and the prospect of **6** additional data sources. The individual data entry forms replied by partners can be found in Annex 3.

3. Data repositories linked to the iAOS as data access services

The INTAROS Data Catalogue is the entry point for obtaining in-depth metadata information on datasets to be used in showcasing applications running in the iAOS Cloud Platform. This provides information on how to access these datasets (e.g. URL endpoint, supported data access protocol).

The data repositories presented hereafter have been reviewed in support of the iAOS showcase applications identified as part of the collaboration between WP5 and WP6. They are accessible online by the iAOS-hosted applications either through OPeNDAP or as other API-based services.

When an actual integration into data processing applications was performed as part of other INTAROS tasks, the mention “iAOS Application” is added along with the reference to it.

When a data repository is listed as a candidate for data integration into iAOS-powered applications, the mention “iAOS relevance” is added along with the reference to the identified potential application.

3.1. OpenDAP servers contributed from INTAROS partners

In preparation of the INTAROS 2nd General Assembly (Bremen, January 2019) a dedicated INTAROS Partners Task Force was set up, for the exchange of experience on the use of OpenDAP servers and NetCDF files.

Over the 2019 and 2020 project activities, a number of OPeNDAP servers have been deployed by the project partners, to serve data repositories identified as relevant for the iAOS. These partner servers have been exploited in INTAROS along with external OPeNDAP servers, as listed below.

Some existing OPeNDAP servers, for example in the case of MET.NO (not an INTAROS partner), were used when identified as a needed data source for an INTAROS task.

SMHI OpenDAP server
Pan-arctic hydrological model Arctic-HYPE results
http://opendata-download.smhi.se/opendap
iAOS Application: this server is used in Showcase 1 for Task 6.1 with SMHI: Pan-Arctic Hydrological Modelling.

IMR/NMDC OpenDAP server
Content: Arctic CTD cruise campaigns
Access: http://opendap1.nodc.no/opendap
iAOS Application: this server is used in the iAOS Showcase 2 for Task 6.2 with IMR on "Barents Sea Multi-depth Temperature & Salinity Maps".

MET.NO OpenDAP server
Content: Arome Arctic archive, including model output for wind speed, surface air pressure and temperature, precipitations
Access: https://thredds.met.no/
iAOS Application: this server was tested as part of the iAOS Showcase 3 investigations for Task 6.4 with FMI "Maps for Svalbard Avalanche Forecast Modelling".
Note: WP5 efforts are reported in D5.13, which delivered only feasibility study results for this Showcase (no iAOS application could be finalized).

FMI OpenDAP server
Content: Snow depth data
Access: not implemented
iAOS relevance: this server was intended to be part of the iAOS Showcase 3 investigations for Task 6.4 with FMI "Maps for Svalbard Avalanche Forecast Modelling". The Showcase work has been conducted based on file sharing only.
Note: WP5 efforts are reported in D5.13, which delivered only feasibility study results for this Showcase (no iAOS application could be finalized).

Bremen University OpenDAP server
Sea Ice Concentration (AMSR-E/AMSR2) & Sea Ice Thickness (SMOS)
Access: OPeNDAP server access delegated to Terradue Cloud Platform (see further below)
Data sources:
https://seaice.uni-bremen.de/start/data-archive/
https://seaice.uni-bremen.de/data/
iAOS relevance: this server was tested only, as part of the iAOS Service.

Terradue Cloud Platform - Hyrax OpenDAP server
Content: all relevant Arctic dataset samples for Ellip workflows users (Cloud integration service), with on-demand creation of dataset collections (copy from reference data servers).
Access: https://opendap.terradue.com/hyrax (Ellip subscribers access only)
iAOS Application: this server is used in:

- the iAOS Showcase 2 for Task 6.2 with IMR on "Barents Sea Multi-depth Temperature & Salinity Maps" as a developer resource used for testing the application under development.
- the iAOS Showcase

Terradue Cloud Platform - TDS OpenDAP server
Content: all relevant Arctic dataset samples for Ellip workflows users (Cloud integration service), with on-demand creation of dataset collections (copy from reference data servers).
Access: https://opendap.terradue.com/thredds (Ellip subscribers access only)
iAOS Application: this server is used in the iAOS Showcase 2 for Task 6.2 with IMR on "Barents Sea Multi-depth Temperature & Salinity Maps" as a developer resource used for testing the application under development.

3.2. Other API-based service, exploitable due to their Cloud readiness

The iAOS cloud infrastructure builds on the principle of interoperable data access from trusted data repositories exposing machine-to-machine interfaces.

Data providers collecting datasets are deploying efforts to bridge the gap between a situation of spreaded datasets hosted on multiple machines or servers, in order to reach a first level of improvement through data curation procedures (typically under the FAIR principles) and the consolidation of datasets onto centrally managed data servers.

Having these data servers accessible online as online data repositories is the next step.

The ultimate step for Cloud readiness, defined as a more powerful data exploitation capacity within applications, is to expose these online data repositories through APIs, that can empower applications to programmatically search, filter and fetch datasets according to their specific applicative needs.

We present hereafter the list of online data repositories exposed to software applications through APIs, therefore qualifying under the category of "Cloud ready".

3.2.1 Terradue Data Agency service

The iAOS benefits from this EO data exploitation capabilities and several data processing campaigns have been conducted during the INTAROS project based on it (cf. descriptions in the INTAROS deliverable D5.11 - Integration of new processing services V2). We summarize hereafter the mechanisms in place that relate to the iAOS objectives for data integration from existing repositories (see also the INTAROS deliverables D5.2 - iAOS platform and tools V1 and D5.8 - iAOS platform and tools V2).

The Terradue Data Agency backbone is made of an OGC-compliant EO data catalogue service (OpenSearch profiles). Connected to this catalog service, a scheduled data staging service is coordinating the data ingestion for a processing service. It can submit bulk requests, following specific metadata parameters (e.g. geographic area, time span, product type) that can be operated for either time-driven or data-driven staging operations. To access multiple sources of EO products, the Catalogue includes Data Hoovers that create a data coordinator mechanism that harvests multiple data repositories to support the cost-effective selection of the processing infrastructure. The Data Hoovers are plugins, specially designed to harvest individual systems and abstract a data source to a common HTTP interface. A data caching system, or "data gateway" is also involved to organise dataset files from remote sources into an organised file system following a defined layout. For instance, most of the data hoovers accessing remote repositories of EO data will organise the data according to the mission, the product type and the date. This way the data gateway can query any remote data source via its data hoover and access the files in a common way. This capacity allows to remove the

complexity of the data source heterogeneity that is sometimes needed by data processing jobs, and only focuses on the fast data access path, files caching and access control. Finally, the data Catalogue component is able to store and query the EO metadata of these products in indexes and provides an interface for searching the dataset in a catalogue via an OpenSearch interface according to a data model.

This API was used by NERSC in the activity referenced as part of D5.13 under the sections 2.3.1.5. polarstern (UAK research school / INTAROS training session on processing Sentinel-1 products) and 2.3.4.1. collaboration with the EC H2020 project NextGEOSS (productions of Sea ice classification and Sea ice drift datasets from the processing of Sentinel-1 products).

3.2.2 Copernicus Marine environment monitoring service (CMEMS)

CMEMS provides a SUBSETTER download mechanism. Users can retrieve a subset of gridded datasets through http and https protocol. The main idea is: download exactly what the user needs. Users can retrieve the needed variable, the geospatial coverage, the temporal coverage. There are two ways to access this service: via GUI (through the web portal), or via a machine to machine interface (script). CMEMS provides several Application Programming Interfaces (APIs) to have access to its products and the related metadata. At the moment, the MOTU API is the most suitable for integrating CMS datasets into iAOS. There are two MOTU features, *Subsetter* and *Direct Get File*, each with its advantages and both available using any http(s) library/client consuming the MOTU API. The official MOTU clients available to users are a Python version and a Java version. In both cases, it is a basic script which integrates the MOTU API and allows the interaction with the Copernicus Marine Database. However for December 2021 the release of an HTTPs - OPeNDAP API is planned, which will then probably allow an even easier integration into iAOS. For further information:

https://help-cmems.mercator-ocean.fr/en/articles/4794731-which-apis-are-provided

3.2.3 PANGAEA data library

The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the hosting institutions. Each dataset can be identified, shared, published and cited by using a Digital Object Identifier (DOI Name). PANGAEA also allows data to be published as supplements to science articles (example) or as citable data collections in combination with data journals like ESSD, Geoscience Data Journal, Scientific Data, or others. The PANGAEA data editorial ensures the integrity and authenticity as well as a high usability of your data. Archived data are machine readable and mirrored into our data warehouse which allows efficient compilations of data.

For direct data access PANGAEA provides a service which queries the PANGAEA data warehouse and returns values in a tab delimited text file, based on search criteria that (1) specify a bounding box in time and space (latitude, longitude and water depth), and (2) specify a individual list of parameters.

The returned text file contains following columns: Date/Time, Latitude (north), Longitude (east), Depth water [m], Parameter 1, Parameter 2, ..., Data source (DOI)

Base URL for all queries is: http://ws.pangaea.de/dds-fgp

3.2.4 EMODnet web service

Generally the EMODnet web services allows search, visualisation and download of EMODnet data and data products. Different metadata catalogs exist for the different disciplines offering the ability to search collections of metadata for data, services and related information objects related to the EMODnet Marine Data. The data catalogues also offer an OGC CS-W endpoint to other client applications to connect to the service and query the metadata held in the catalogue. The EMODnet web service documentation provides detailed information on how to handle the provided web-services (https://emodnet.eu/en/data). It also provides a tutorial on how to include the OGC services in R programming for example.

3.2.5 FMI OGC web services

The objective here is to disseminate all datasets generated at the National Satellite data centre using one interface with OGC web services capabilities. The metadata about FMI datasets can be harvested to iAOS using web catalogue service.

The types of datasets that are served cover Raster and gridded dataset with varying spatial and temporal resolution (for Temporal, there are daily, weekly, monthly aggregates). Satellite data products include Sentinel-1 mosaics and cryospheric products. Vector datasets comprise In-situ measurements (point data), collected in databases or Shapefiles. The main component is the Geoserver software solution, used to create mini datacubes of various layers that can be accessed based on a location bounding box and time. Geoserver allows command line operations, users can use scripts, or other GIS software (Erdas Imagine, Matlab, QGIS, GDAL, etc. to access the data).

3.2.6 Zenodo

Zenodo 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. Zenodo provides a unique identifier, a digital object identifier (DOI), for each uploaded resource. This offers a means to credit the providers of data and other resources (e.g. webinars).

INTAROS has established a Community on Zenodo where all resources generated with support from INTAROS are linked:
https://zenodo.org/communities/intaros-h2020

3.3. Recommendation for other API based, institutional repositories

In addition to the API-based data repositories described above, there are a large number of institutional repositories. These repositories have variable support for standard access protocols, but many of those are connected to large national repositories e.g. NMDC, SEANOE.

Our recommendation is that long term storage of datasets should be done in open access repositories providing OPENDAP or other open standard API services, to facilitate standardized access to data and promote interoperability between data repositories (D1.8 INTAROS Data Management Plan, section 8).

4. Data repositories used in the iAOS Showcase applications

WP6 integrates remote sensing data and in situ observations delivered from a variety of platforms at different geographical scales and from different locations. Incorporation of these data into analysis and modelling systems, including physical and ecological process models, climate models and forecast methods, is providing support for better products to key societal areas.

A collaboration between WP5 and WP6 was established to develop Showcase applications using the iAOS Cloud Platform. This section presents these showcase applications, the required input data and the data products generated from these showcases.

For each showcase, the contribution from WP5 is described and in particular how this can benefit selected showcases within WP6 following the structure below:

- **Objectives:** definition of the showcase objective.
- **Data sources:** description of the data access and data integration to be based on cloud technology (online data storage, data service APIs).
- **Data exploitation:** description of the data processing tasks and the related tools.
- **Intermediary data products:** the by-products of the WP5 activities to support the definition of a Showcase application.
- **Foreseen scientific data productions:** description of the foreseen benefits delivered by a showcase application running on the cloud platform tools and services.

The showcases were defined in a series of events as follows:

- **Helsinki, 2018** - Joint WP5-6 Workshop– Identification of the application types, data sources and data outputs that are relevant as potential iAOS showcase applications.
- **Bremen, 2019** - RGeostats Workshop & Data Distribution workshop – Introduction to the INTAROS support material (including tutorials) for building iAOS Showcase applications.
- **Sopot, 2020** - Interviews with WP6 task leaders for their work plan analysis, and identification of best ‘showcase’ opportunities to be supported by the iAOS (WP5)
    - As return of experience (INTAROS internal) on how the WPs interact in order to illustrate the iAOS added value
    - As a set of results-oriented data collections and services, which can support the INTAROS outreach activities in 2020-2021

      – Identification of the WP6 work plan for their Showcase applications, and identification of best ‘showcase’ opportunities to be supported by the iAOS (the selected Showcase applications are presented further below).
- **Remote, 2020** - Intermediate results reviews – Definition of the final objectives for each Showcase, especially in terms of dissemination / promotion of the results as INTAROS outcomes.
- **Remote, 2021** - Peer-to-Peer workshops presenting the iAOS support to actual and potential users of each selected Application – (on-going)

From the perspective of Data integration from existing repositories, the WP5 support to the showcases has been to:

- Guarantee a good data availability (check the data sources descriptions, verify the online accessibility of data services, perform interoperability testing)

- Leverage the iAOS Portal and INTAROS Data Catalogue to guide on the selection of most relevant datasets for the set up of the selected show cases.
- Show the data flow from a source, to the application developer, and to its end users (a summary is provided in the iAOS Showcase applications infosheets)
- Help to implement a smooth data flow between systems

4.1. Showcase 1 (Task 6.1 SMHI): Pan-Arctic Hydrological Modelling

4.1.1. Objectives

- Have the "observational" data available for search and download from iAOS
- Have the Arctic-HYPE run at SMHI and provide the data as open data from SMHI repositories:
  - Daily analyses of last 60 days
  - Medium range forecast of coming 10 days
- Improve predictions of spring floods, river ice breakup and freshwater flow to Arctic Ocean, cf. INTAROS D6.1 Climate model initialization v1.4

4.1.2. Data sources

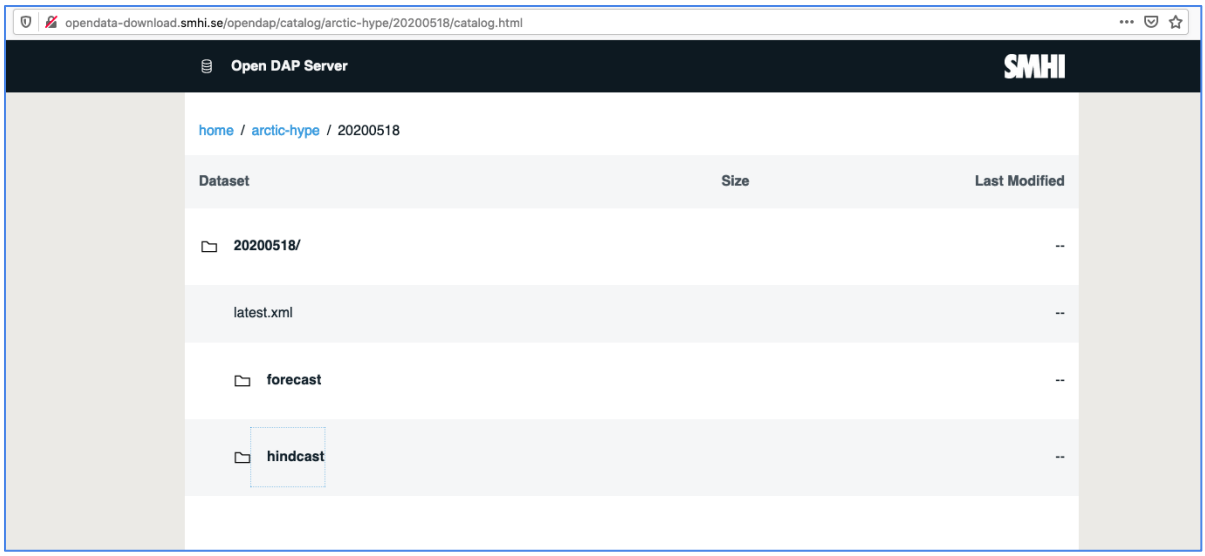
- River discharge data from the Arctic Hydrological Cycle Observing system (Arctic-HYCOS) - assessed and enhanced in INTAROS WP2
  https://catalog-intaros.nersc.no/dataset/arctic-hycos-hydrological-data/
- HydroGFD v3 temperature and precipitation data (Global Forcing Data for Hydrology) SMHI's FTP server (ftp.smhi.se)
  Read more: http://hypeweb.smhi.se/wp-content/uploads/2019/04/What-is-HydroGFD2.pdf
- ECMWF deterministic medium range weather forecasts
  https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/

4.1.3. Data exploitation

- Implement HYCOS pre-processing (both archive of quality controlled data with 4months/2years lag, or provisional datasets)
  - In-house server at SMHI
  - Cloud-based, using Ellip, to compare
- Schedule HYCOS pre-processing operations to be made daily at a certain time
- Setup OpenDAP server for publishing the Arctic-HYPE model results

4.1.4. Intermediary data products

- OpenDAP server publishing Pan-arctic hydrological model Arctic-HYPE results provided by SMHI
  http://opendata-download.smhi.se/opendap/catalog/catalog.html



4.1.5. Foreseen scientific data productions

iAOS Showcase (demo)	**Use of Arctic-HYCOS sources (multi-provider)**	**Use of Arctic-HYPE (SMHI) outputs (per station ID, with 1 dimension time)**
Flood forecasting use cases Select a point, or some AOI (in Russia MPI in Yakutsk - L. Lebedeva)	For checking the initial conditions	For checking the model analysis of current condition and the model forecast
Entire freshwater inflow to the ocean (in collaboration with IMR - R.Hordoir)	To integrate discharge into the ocean from observations (only represent 60% draining land, see WP2 deliverables).	To integrate model analysis/predictions on river discharge into the ocean on user defined resolution

4.2. Showcase 2 for T6.2 with IMR: Barents Sea Multi-depth Temperature & Salinity Maps

4.2.1. Objectives

- Use the Geostatistical Library (RIntaros / RGeostats) and build the R software for interpolating maps from CTD datasets
- Generate temperature and salinity fields for:
  - modelling of Arctic Ocean biogeochemistry
  - validation of climate model projections (NorCMP)
- Build a Web Processing Service for iAOS

4.2.2. Data sources

- Institute of Marine Research (IMR) center of Norway, 7 research vessels
- collecting conductivity, temperature, depth (CTD) data in the North Sea.
    - Acquisitions between 7th January 1995 and 29th November 2016.
    - 5.5 billion of samples measured over 63 500 positions (vertical profiles).
- All files are freely available on an OpenDAP server: http://opendap1.nodc.no/opendap/physics/point/yearly/contents.html
- NetCDF files (one file by year and per vessel). Coordinates are in degrees (Long/Lat) and the timestamp is the number of minutes since the 1st January 1950. The whole dataset volume is 880 GB.

4.2.3. Data exploitation

- Use of the iAOS OpenDAP server at NODC
- Explanatory data analysis and variography
- Modelisation of spatial behavior for Temperature and Salinity

4.2.4. Intermediary data products

- Map productions per run have been performed on the Standalone solution (RGeostats R software):
    - Base maps
    - Average per cell
    - Cross-validation (blind test) maps
    - Estimation (Temperature / Salinity) and corresponding uncertainty maps

4.2.5. Foreseen scientific data productions

It is foreseen to improve the initial geostatistical spatial analysis performed in 2018 by:

- building a unique spatio-temporal variogram model for Temperature
- taking into account co-variables like (salinity, bathymetry...)
- handling currents nonstationarities and distance to coastline

4.3. Showcase 3 for T6.4 with FMI: Maps for Svalbard Avalanche Forecast Modelling

4.3.1. Objectives

- Use the Geostatistical Library (RIntaros / RGeostats) and build the R software for interpolating maps from snow stations, arome model output and terrain model
- Generate snow depth maps at regular time intervals as input for avalanche forecast model

4.3.2. Data sources

- NMI Frost API historical weather and climate data stations

(selected files on shared Drive)
- NMI arome model: https://thredds.met.no/thredds/catalog/aromearcticarchive/catalog.html
- Norwegian Polar Institute Svalbard Terrain Model: https://doi.org/10.21334/npolar.2014.dce53a47

4.3.3. Data exploitation

- Explanatory data analysis and variography
- Handle the different spatial distributions and resolutions of the data ("support")
- Modelisation of spatial and temporal behavior for snow depth through co-variables (temperature, wind speed by class of wind direction)

4.3.4. Intermediary data products

Pre-analysis of the data :
- Few stations: temporal series of snow thickness measured at short time steps
- Arome models: various maps covering the whole area, every 6 hours, on a large scale grid (incl. snow thickness derived from model)
- Several co-variables

Processing :
- Regularization of the station time series by averaging over 6 hours
- Correlation (space-time) of snow depth variable with arome model output
- Estimation using both informations (with relevant co-variables) over a small scale grid, at regular 6 hours intervals.

4.3.5. Foreseen scientific data productions

- It is foreseen to finalize the showcase with a proof-of-concept:
    - Create snow depth variogram models (with relevant co-variables) for each class of wind orientation
    - Generate snow depth map for:
        - a given date and time interval (6 hours)
        - accounting for a global wind orientation

4.4. Showcase 4 for T6.8 with Aarhus: Baffin Bay Bottom Temperature Maps

4.4.1. Objectives

- Use the Geostatistical Library (RIntaros / RGeostats) and build the R software for interpolating ocean floor temperature maps from CTD and Trawl datasets
- Generate temperature fields at bottom of the ocean in support of:
    - Analysis of long term global warming influence
    - Analysis of the fish stock correlation to bottom temperature

4.4.2. Data sources

CTDs (ICES)

- 1977 to 2017
- 3700 vertical profiles
- 1.34M samples
- 1 CSV file (95Mo)

Bottles (ICES)

- 1960 to 2017
- 7800 vertical profiles
- 167K samples
- 1 CSV file (11Mo)

Trawls (GINR)

- 1988 to 2016
- Catches near sea floor only
- 51K samples
- 1 CSV file (11Mo)

Bathymetry (Gebco)

- Grid lag = 1/250 degree
- Grid size = 10320x4560 nodes
- 47M samples
- 1 NetCDF file (94Mo)

Additional data sources provided by M. Sejr (Nov. 2020):

- Better quality of arctic bathymetry: https://www.nature.com/articles/s41597-020-0520-9
- World ocean database (Temp fields): https://www.nodc.noaa.gov/OC5/WOD/pr_wod.html
- Global temperature and salinity profile program (data?): https://www.nodc.noaa.gov/GTSPP
- NASA project in Greenland (new CTD): https://omg.jpl.nasa.gov/portal/browse/OMGEV-AXCTD/

4.4.3. Data exploitation

- Explanatory data analysis and variography
- Local and global temperature evolution analysis
- Modelisation of spatial and temporal behavior for ocean floor temperature through Bathymetry co-variable

4.4.4. Intermediary data products

- Temporal evolution (global and local):
  - Between 1960 and 2015, T°C has gained 1°C (around 1995)
  - Lower T°C values around 2008 have been recorded
- Map productions:

- Basemap of data
- Bottom temperature estimation by year and its standard deviation

- Time series of average temperature by region (Lat/Depth)

4.4.5. Foreseen scientific data productions

- It is foreseen to improve first the kriging estimation:
  - Reduce the estimation error (currently around 0.7°C)
  - Reduce estimation smoothing
- followed by improvements on:
  - Multi-directional variography and zonal anisotropies
  - Local cross-validation with additional abyssal (or new) data
  - Bathymetry as external drift using non linear regression
  - Salinity as co-variable to be studied

5. Conclusions

This deliverable describes the status of data integration capabilities empowered by the iAOS in year four of the INTAROS project. The ease of integration of Arctic-related data into applications by using the iAOS Cloud Platform tools and services was the main goal of task 5.3. This work builds on a process that started in Task 2.3 in the beginning of the project using the results from the assessment carried out in Task 2.1, and synthesis of results in Task 2.4 (D2.10, D2.11).

One essential achievement of task 5.3 was to guide the INTAROS end users on the data access mechanisms that are usable in support of a range of data exploitation tasks: data screening, data analysis and related experiments to investigate phenomena, and systematic data processing campaigns. In this context, it was essential to support multi-dimensional data sub-setting and data processing, which has been performed in the context of the tasks T5.4 “Geostatistical methods”, T5.5 “Integration of new processing services” and WP6 tasks related to implementing the iAOS Showcase applications. Such guidance is formulated for the iAOS Cloud Platform users (cf. D5.8 Cloud Platform tools & services, section 3) and can be tried out in connection with the iAOS Showcase applications.

The heterogeneity of the methods used by data providers to store their data in repositories proved to be a particular challenge already at a very early stage of the project. A very large proportion of the data systems in use were not providing suitable machine-to-machine interfaces to access the available datasets within software applications via the iAOS cloud platform.

An overview of the maturity level of individual data repositories was provided in version 1 of this deliverable (D5.3) in year two of the project. The individual repositories were of different levels of maturity for deeper integration into iAOS cloud applications.

Therefore, it was decided that it would be necessary to work towards two important types of improvements:

1) make it possible to obtain a good overview of which data products are available for exploitation (section 2 of this report) and to provide support and guidance for data products and services endpoints registration in the INTAROS Data Catalogue.

2) Identify and promote the use of data sources having already software interfaces for data integration as part of the iAOS. Through a deeper integration layer into the iAOS, system-specific discovery (i.e. search) and retrieval (i.e. access) frameworks are directly supported (Pangaea, OpenDAP Servers, Terradue Data Agency for Copernicus Sentinel datasets). In the end the aim is to promote mechanisms which support data access to the relevant remote repositories, to establish a workable environment that federates data infrastructures.

In addition to the integration of large repositories that come with the corresponding APIs, a partner task force was set up quite early on to promote the use of OpenDAP servers and NetCDF files among the partners and to provide corresponding support. As a result, a total of seven OpenDAP platforms have been referenced as part of iAOS.

Moreover, in order to promote the transfer of know-how between the project partners, a number of coordination activities have been coordinated by the Task T5.3 for data integration from existing repositories: workshops, training material, and collaborations for the definition of iAOS Showcase applications.

In summary, support for integration into the data flow of iAOS-powered applications was delivered for almost 20 data sources. In order to demonstrate this data flow adequately, a close cooperation was sought with partners from WP6, aiming at the incorporation of these data products into analysis and modelling systems, exploiting data sources that are particularly relevant for the iAOS.

Annex

Annex 1: INTAROS Data Catalogue integrations reported as part of the 3rd reporting period

Time Period: December 2019-November 2020

Organization	Title of dataset	Last update
NERSC	Fram Strait/ACOBAR: XBT measurements and derived values - Fram Strait - September 2010	January 2020
NERSC	Fram Strait/ACOBAR: XBT measurements and derived values - Fram Strait - September 2012	January 2020
NERSC	Fram Strait/UNDER-ICE: XBT measurements and derived values - Fram Strait - September 2013	April 2020
RIHMI-WDC	Climatic data on weather conditions of the coastal zone of the Chukchi Sea: wind speed. (RU_RIHMI-WDC_2361)	June 2020
RIHMI-WDC	Climate data of the high seas: Wind speed. Arctic Ocean (Western Sector). (RU_RIHMI-WDC_2168)	June 2020
RIHMI-WDC	Climatic data on the weather conditions of the coastal zone of the White Sea: water temperature. (RU_RIHMI-WDC_966)	June 2020
RIHMI-WDC	Climatic data on the hydrometeorological conditions of the coastal zone of the Kara Sea: wave height. (RU_RIHMI-WDC_2368)	June 2020
RIHMI-WDC	Climatic data on weather conditions of the coastal zone of the Laptev Sea: water temperature. (RU_RIHMI-WDC_965)	June 2020
RIHMI-WDC	Climate data of the high seas: Wind speed. Arctic Ocean (Eastern sector). (RU_RIHMI-WDC_2169)	June 2020
RIHMI-WDC	Climatic data on the weather conditions of the coastal zone of the Barents Sea: water temperature. (RU_RIHMI-WDC_918)	June 2020
RIHMI-WDC	Climatic data on weather conditions of the coastal zone of the Laptev Sea: wave height. (RU_RIHMI-WDC_2369)	June 2020
RIHMI-WDC	Climatic data on weather conditions of the coastal zone of the Chukchi Sea: wave height. (RU_RIHMI-WDC_2371)	June 2020
RIHMI-WDC	Climatic data on the hydrometeorological conditions of the Kara Sea coastal zone: water temperature. (RU_RIHMI-WDC_969)	June 2020
RIHMI-WDC	Climatic data on the hydrometeorological conditions of the coastal zone of the Kara Sea: wind speed. (RU_RIHMI-WDC_2356)	June 2020
RIHMI-WDC	Climatic data on the hydrometeorological conditions of the coastal zone of the East Siberian Sea: wind speed. (RU_RIHMI-WDC_2355)	June 2020
RIHMI-WDC	Climatic data on weather conditions in the coastal zone of the East Siberian Sea: wave height. (RU_RIHMI-WDC_2373)	June 2020
RIHMI-WDC	Hazardous events and analyzes. Hazardous Storm Alerts (O) for the last 30 days. (RU_RIHMI-WDC_1752)	June 2020
RIHMI-WDC	Danger over the past 24 hours (WAREP code) .NEW. (RU_RIHMI-WDC_2892)	June 2020
RIHMI-WDC	Climatic data on the hydrometeorological conditions of the coastal zone of the East Siberian Sea: water temperature. (RU_RIHMI-WDC_970)	June 2020
RIHMI-WDC	Climatic data on weather conditions in the coastal zone of the Chukchi Sea: water temperature. (RU_RIHMI-WDC_968)	June 2020
RIHMI-WDC	Integrated AARI map of the state of ice cover of the Laptev Sea in GIF format. (RU_AARI_1135)	June 2020

RIHMI-WDC	Integrated AARI map of the state of the ice cover of the East Siberian Sea in GIF format. (RU_AARI_1136)	June 2020
RIHMI-WDC	Predictive map of surface wind in the Arctic Ocean with a lead time of 1-6 days. (RU_AARI_3251)	June 2020
RIHMI-WDC	Integrated AARI map of the state of the Kara Sea ice cover in GIF format. (RU_AARI_1134)	June 2020
RIHMI-WDC	Integrated AARI map of the state of the Barents Sea ice cover in GIF format. (RU_AARI_1132)	June 2020
RIHMI-WDC	Integrated AARI map of the state of the ice cover of the Chukchi Sea in GIF format. (RU_AARI_1137)	June 2020
NORCE Norwegian Research Institute	High resolution sea-ice mosaic	July 2020
Institute of Marine Research (IMR)	CTD data collected with R/V Michael Sars	July 2020
Institute of Marine Research (IMR)	CTD data collected with R/V Jan Mayen	July 2020
Institute of Marine Research (IMR)	CTD data collected with R/V Håkon Mosby	July 2020
Institute of Marine Research (IMR)	CTD data collected with R/V Sarsen	August 2020
Uni Stockholm, Meteo (MISU)	Radiosonde data from the Arctic Ocean 2018 expedition	August 2020
Uni Stockholm, Meteo (MISU)	Ceilometer backscatter, cloud base height and cloud fraction data from the Arctic Ocean 2018 expedition	August 2020
Uni Stockholm , Meteo (MISU)	Present weather sensor visibility and precipitation data from the Arctic Ocean 2018 expedition	August 2020
Uni Stockholm , Meteo (MISU)	Navigation, meteorological and surface seawater data from the Arctic Ocean 2018 expedition	August 2020
Uni Stockholm , Meteo (MISU)	Weather data from MISU weather station during the Arctic Ocean 2018 expedition	August 2020
Uni Stockholm , Meteo (MISU)	Micrometeorological data from an ice-floe adjacent to an open lead during the Arctic Ocean 2018 expedition	August 2020
Uni Stockholm , Meteo (MISU)	Micrometeorological data from icebreaker Oden's foremast during the Arctic Ocean 2018 expedition	August 2020
IOPAN	CTD measurements during the INTAROS 2018 cruise	October 2020
NERSC	Fram Strait/ACOBAR: Ambient noise in water - From Strait - Aug 2010 - Jul 2011 and Sep 2011 - Jul 2012	October 2020
NERSC	Fram Strait/ACOBAR: Range and depth average sea water temperature from acoustic tomography measurements - Fram Strait - Sep 2010 - Jul 2011 and Sep 2011 - Jul 2012	October 2020
NERSC	CTD data collected in Storfjorden, Svalbard, during the UAK 2020 Cruise	October 2020

Annex 2: Most actual integrations to the INTAROS Data Catalogue

Time Period: December 2020-Spring 2021 (query date: 19.04.2020)

Organization	Title of dataset	Last update
NERSC	UAK - Active and Passive Acoustic data - 23 Jun 2020	December 2020
NERSC	UAK - Active and Passive Acoustic data - 24 Jun 2020	December 2020
NERSC	UAK - Active and Passive Acoustic data - 25 Jun 2020	December 2020
NERSC	UAK - Active and Passive Acoustic data - 26 Jun 2020	December 2020
PISUNA	PISUNA database	December 2020
UiT The Arctic University of Norway	Phenology of spring bird migration to North Norway	December 2020
University of the Faroe Islands	Hunters' Self-Monitoring in the Faroe Islands	December 2020
Cornell Lab of Ornithology	eBird	December 2020
NORDECO	Secchi Disk Study	December 2020
NORDECO	GLOBE Observer: Clouds	December 2020
CSIPN	Yakutia CBM	December 2020
NORDECO	Pilot Whale Statistics	December 2020
University of Alaska Fairbanks	Alaska Arctic Observatory & Knowledge Hub (AAOKH)	December 2020
NORDECO	Piniarneq	December 2020
Yukon River Inter-Tribal Watershed Council	Yukon River Inter-Tribal Watershed Council	December 2020
NORDECO	Sea Ice for Walrus Outlook	December 2020
GEUS	Citizen seismology program, Qeqertalik, Greenland	December 2020
Finnish Meteorological Institute (FMI)	FMI Snow Depth Measurement Citizen Science Program	December 2020
Húsavík Research Center, University of Iceland	Spotter Pro Marine Mammal Records	January 2021
Institute of Marine Research (IMR)	Mooring data from the Barents Sea Opening – Atlantic Water inflow	January 2021
Institute of Marine Research (IMR)	NORWECOM.E2E simulation, herring distribution 2010-2069	January 2021
Institute of Marine Research (IMR)	NORWECOM.E2E simulation, mackerel distribution 2010-2069	January 2021
Institute of Marine Research (IMR)	NORWECOM.E2E simulation, blue whiting distribution 2010-2069	January 2021
Institute of Marine Research (IMR)	NORWECOM.E2E overfishing simulation, mackerel, herring or blue whiting distribution 2015-2025	January 2021
Uniwersytet Śląski	Seasonal fluctuations of Hansbreen terminus position	January 2021
Uniwersytet Śląski	Hornsund Fiord area 1936-2010	January 2021
Uniwersytet Śląski	Height of the Hansbreen terminus	January 2021
Uniwersytet Śląski	Changes in position of Hansbreen front	January 2021
Uniwersytet Śląski	Inventory of tidewater glaciers of Svalbard	January 2021

Institute of Marine Research (IMR)	CTD data collected with R/V G.O.Sars	January 2021
Institute of Marine Research (IMR)	CTD data collected with R/V Helmer Hanssen	January 2021
Institute of Marine Research (IMR)	CTD data collected with R/V Johan Hjort	January 2021
Institute of Marine Research (IMR)	CTD data collected with R/V Kristine Bonnevie	January 2021
NORDECO	Happywhale	January 2021
UiB	Optical property measurements collected in Storfjorden, Svalbard, during the UAK 2020 Cruise	January 2021
Technical University of Denmark (DTU)	Arctic Steric Heights	January 2021
Technical University of Denmark (DTU)	Arctic Vertical Land Motion	January 2021
CNRS	Passive acoustic data from Kongsfjorden, Svalbard, September 2018	February 2021
Universität Hamburg	Ocean-Sea Ice Synthesis from 2007-2016	March 2021
GEUS	Greenland Ice Sheet solid ice discharge	March 2021
GEUS	Greenland Freshwater Runoff	March 2021
Uniwersytet Śląski	Glacier mass balance: Werenskioldbreen	April 2021

Annex 3: Partner feedback on the request for expected data sources

Engagement email (April 2021):

Dear INTAROS project partner,

you have previously interacted with WP5 for registering datasets in the INTAROS Data Catalogue.
As of now, you might have some, or not any, dataset registered.

Please provide below details for the description of any newly available datasets from your INTAROS year 3 activities. We'll coordinate with you afterwards for the INTAROS Data Catalogue registration steps.

We present hereafter the responses collected from the partner organisations contacted about any additional datasets made available as part of the last INTAROS project reporting period (from December 2020 onwards).

Partner: Uni Stockholm , Meteo (MISU); provided by Michael Tjernström

Description	Value
Title Short text naming the dataset and outlining its content.	Atmospheric data from weather station on icebreaker Oden during the Ryder 2019 expedition
Parameter name(s) List of parameters contained in the dataset. E.g. "Ocean Temperature (profiles)".	Air temperature and relative humidity, atmospheric pressure, wind speed and direction (absolute and relative to ship), surface temperature, broadband incoming solar and terrestrial radiation, cloud cover cloud bases and lidar backscatter from ceilometer and visibility
Project/Program name(s) Name of project(s)/program(s) that supported data acquisition.	Arctic Climate Across Scales (ACAS)
Observing system Name of observing system that collected the data.	N/A
Description Short description of dataset contents, how data was collected or generated, what processing steps & quality control procedures were applied, estimated uncertainty.	Data was collected during the 2019 Oden expedition to the north-west passage and the Ryder glacier and fjord, from three instrument systems; a weather station, a laser ceilometer and a so-called present weather sensor.
Tags Keywords associated with the dataset, for fast search.	Arctic meteorology, Arctic surface fluxes, Arctic clouds
License License under which the dataset is made available.	Open Data Commons Attribution License (ODC-By) v1.0
Organisation Name of the organisation owning the dataset.	Department of Meteorology, Stockholm university
Source URL Online access point for dataset, e.g. for a Thredds server.	www.bolin.su.se/data
Version Version of the dataset.	Not ready yet
Principal Investigator Name of principal investigator(s).	Michael Tjernström and John Prytherch
PI e-mail Email address of principal investigator(s).	michaelt@misu.su.se and john.prytherch@misu.su.se
Data Curator Name of data curator for the dataset.	Bolin Centre for Climate Research
Data Curator e-mail E-mail address of the data curator.	bolindata@su.se

Description	Value
Title Short text naming the dataset and outlining its content.	Atmospheric data from eddy-covariance measurements of turbulent fluxes from icebreaker Oden during the Ryder 2019 expedition
Parameter name(s) List of parameters contained in the dataset. E.g. "Ocean Temperature (profiles)".	Air temperature and relative humidity, wind speed and direction (absolute and relative to ship), turbulent fluxes of momentum, sensible heart, water vapor, carbon dioxide and methane.
Project/Program name(s) Name of project(s)/program(s) that supported data acquisition.	Arctic Climate Across Scales (ACAS)
Observing system Name of observing system that collected the data.	N/A
Description Short description of dataset contents, how data was collected or generated, what processing steps & quality control procedures were applied, estimated uncertainty.	Data was collected during the 2019 Oden expedition to the north-west passage and the Ryder glacier and fjord, from three instrument systems; a weather station, a laser ceilometer and a so-called present weather sensor.
Tags Keywords associated with the dataset, for fast search.	Arctic meteorology, Arctic surface fluxes, Arctic clouds
License License under which the dataset is made available.	Open Data Commons Attribution License (ODC-By) v1.0
Organisation Name of the organisation owning the dataset.	Department of Meteorology, Stockholm university
Source URL Online access point for dataset, e.g. for a Thredds server.	www.bolin.su.se/data
Version Version of the dataset.	Not ready yet
Principal Investigator Name of principal investigator(s).	Michael Tjernström and John Prytherch
PI e-mail Email address of principal investigator(s).	michaelt@misu.su.se and john.prytherch@misu.su.se
Data Curator Name of data curator for the dataset.	Bolin Centre for Climate Research
Data Curator e-mail E-mail address of the data curator.	bolindata@su.se

Partner: RADI (CKAN); provided by Yubao Qiu (qiuyb@aircas.ac.cn)

Description	Value
Title Short text naming the dataset and outlining its content.	HiMAC is a data category portal of High Mountain and cold regions, supported by MARIS-INTAROS project of MOST and other projects of CAS.
Parameter name(s) List of parameters contained in the dataset. E.g. "Ocean Temperature (profiles)".	Several datasets, and met dataset about the snow, ice, buoy observations, in the domain of Arctic and High Mountain Asia.
Project/Program name(s) Name of project(s)/program(s) that supported data acquisition.	MARIS-INTAROS and several projects support by CAS
Observing system Name of observing system that collected the data.	Remote sensing products, in-situ measurement, and buoy.
Description Short description of dataset contents, how data was collected or generated, what processing steps & quality control procedures were applied, estimated uncertainty.	Remote sensing datasets are all valued-added dataset, and buoys datasets were from original dataset.
Tags Keywords associated with the dataset, for fast search.	Sea ice, SWE, HMA, Arctic, glacier
License License under which the dataset is made available.	
Organisation Name of the organisation owning the dataset.	Aerospace Information Research Institute, Chinese Academy of Science (AIR-CAS)
Source URL Online access point for dataset, e.g. for a Thredds server.	http://47.104.68.151/
Version Version of the dataset.	Testing version (living)
Principal Investigator Name of principal investigator(s).	Yubao Qiu
PI e-mail Email address of principal investigator(s).	qiuyb@aircas.ac.cn, yubaoqiu@gmail.com
Data Curator Name of data curator for the dataset.	
Data Curator e-mail E-mail address of the data curator.	

Partner: Institute of Oceanology PAS (IOPAN); provided by Agnieszka Beszczynska-Möller

Description	Value
Title Short text naming the dataset and outlining its content.	Physical oceanographic data from measurements by IOPAN moorings north of Svalbard in 2017-2020
Parameter name(s) List of parameters contained in the dataset. E.g. "Ocean Temperature (profiles)".	Ocean temperature (point and profiles), ocean salinity (point and profiles), ocean currents (point and profiles), sea ice drift and draft, dissolved oxygen
Project/Program name(s) Name of project(s)/program(s) that supported data acquisition.	INTAROS, A-TWAIN, IOPAN statutory research
Observing system Name of observing system that collected the data.	INTAROS moored array, A-TWAIN moored array
Description Short description of dataset contents, how data was collected or generated, what processing steps & quality control procedures were applied, estimated uncertainty.	Data were collected during one-year long deployments of ocean moorings in 2017-2020 north of Svalbard at two mooring lines (INTAROS line along 22°E and A-TWAIN line along 31°E). Data were collected using temperature sensors (RBR Solo3 and Duet3), CTD sensors (SeaBird SBE37 and RBR Concerto3), TRDI QM and LR ADCPs, Nortek Signature 55 and Signature 250 AD2CPs, and AADI SeaGuard. Data processing and quality control is specific for each type of instrument and described in a separate document.
Tags Keywords associated with the dataset, for fast search.	Arctic physical oceanography, ocean temperature, ocean salinity, ocean currents
License License under which the dataset is made available.	TBD
Organisation Name of the organisation owning the dataset.	Institute of Oceanology PAS, Sopot, Poland
Source URL Online access point for dataset, e.g. for a Thredds server.	To be provided later
Version Version of the dataset.	To be provided later
Principal Investigator Name of principal investigator(s).	Agnieszka Beszczynska-Möller
PI e-mail Email address of principal investigator(s).	abesz@iopan.pl
Data Curator Name of data curator for the dataset.	Institute of Oceanology PAS, Sopot, Poland

Partner: University Bergen (UiB); provided by Mathilde B. Sørensen

Description	Value
Title Short text naming the dataset and outlining its content.	Ocean Bottom Seismometer data collected west of Svalbard and in Storfjorden (may be one or two datasets)
Parameter name(s) List of parameters contained in the dataset. E.g. "Ocean Temperature (profiles)".	3-component ground motion waveforms and hydrophone waveforms
Project/Program name(s) Name of project(s)/program(s) that supported data acquisition.	INTAROS, EPOS-Norge
Observing system Name of observing system that collected the data.	OBS network
Description Short description of dataset contents, how data was collected or generated, what processing steps & quality control procedures were applied, estimated uncertainty.	Data have been collected during two OBS deployments. 3 instruments collected data near the spreading ridge west of Svalbard for one year summer 2018-summer 2019. 3 instruments collected data for one year in Storfjorden summer 2019-summer 2020.
Tags Keywords associated with the dataset, for fast search.	OBS, earthquake, seismological data
License License under which the dataset is made available.	
Organisation Name of the organisation owning the dataset.	UiB
Source URL Online access point for dataset, e.g. for a Thredds server.	Data will be made available through the UiB-NORSAR EIDA node: http://eida.geo.uib.no
Version Version of the dataset.	
Principal Investigator Name of principal investigator(s).	Mathilde B. Sørensen
PI e-mail Email address of principal investigator(s).	mathilde.sorensen@uib.no
Data Curator Name of data curator for the dataset.	Jan Michalek
Data Curator e-mail E-mail address of the data curator.	eida@geo.uib.no

Partner: Finnish Meteorological Institute (FMI); provided by Roberta Pirazzini

Description	Value
Title Short text naming the dataset and outlining its content.	Snow spectral albedo in Sodankylä wetland site
Parameter name(s) List of parameters contained in the dataset. E.g. "Ocean Temperature (profiles)".	Spectral downwelling irradiance Spectral reflected irradiance Spectral albedo
Project/Program name(s) Name of project(s)/program(s) that supported data acquisition.	INTAROS (EU H2020), SnowAPP (Academy of Finland)
Observing system Name of observing system that collected the data.	Sodankylä supersite
Description Short description of dataset contents, how data was collected or generated, what processing steps & quality control procedures were applied, estimated uncertainty.	The dataset was collected during three consecutive springs (2019-2021) from mid March until the snow had melted. The spectra of incoming and reflected irradiance were automatically collected every 2 minutes with a custom made SVC dual-sphere spectro-albedometer. The data processing included: the calculation of irradiance from raw photoelectric counts applying the calibration, correction from temperature drift of the sensor's sensitivity, correction from the deviation from the perfect cosine response of the receiving integrating spheres, check of data quality through the comparison with radiative transfer calculations and broadband albedo measurements. As we are still completing the last steps of the processing, we cannot yet provide the final estimation of the uncertainty.
Tags Keywords associated with the dataset, for fast search.	Snow spectral albedo, spectral irradiance
License License under which the dataset is made available.	Open access
Organisation Name of the organisation owning the dataset.	Finnish Meteorological Institute
Source URL Online access point for dataset, e.g. for a Thredds server.	The data will be published (latest in early 2022) in association with the publication (INTAROS special issue) in the recently developed FMI data repository http://fmi.b2share.csc.fi which fulfills the FAIR principles.
Version Version of the dataset.	-Raw data; -1st processed version (calibration and corrections applied, quality check done) is under preparation.
Principal Investigator Name of principal investigator(s).	Roberta Pirazzini
PI e-mail Email address of principal investigator(s).	roberta.pirazzini@fmi.fi
Data Curator Name of data curator for the dataset.	Henna-Reetta Hannula
Data Curator e-mail E-mail address of the data curator.	henna-reetta.hannula@fmi.fi

---------- 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: