INTAROS

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 2.6

Catalogue of products and services based on atmospheric data

Start date of project:	01 December 2016	Duration:	60 months
Due date of deliverable:	30 November 2018	Actual submission date:	30 November 2018
Lead beneficiary for preparing the deliverable:	SMHI		
Person-months used to produce deliverable:	5.52 pm		

Authors: Ewan O'Connor (FMI), Eija Asmi (FMI), Roberta Pirazzini (FMI), Andreas P. Ahlstrøm (GEUS), Michele Citterio (GEUS), Robert S. Fausto (GEUS), Katrin Kohnert (GFZ), Andrei Serafimovich (GFZ), Torsten Sachs (GFZ), Alexander Mahura (U Helsinki), Hanna Lappalainen (U Helsinki), Tomasz Wawrzyniak (IGPAN), Piotr Glowacki (IGPAN), Abhay Devasthale (SMHI), Ingo Schewe (AWI), Torill Hamre (NERSC), Morten J. Stette (NERSC)

Version	DATE	CHANGE RECORDS	LEAD AUTHOR
1.0	10/09/2018	Template	Roberta Pirazzini
1.1	10/10/2018	Draft	Roberta Pirazzini
1.2	23/11/2018	Minor modifications after project meeting review	Torill Hamre
2.0	28/11/2018	Final submission version	Abhay Devasthale
	30/11/2018	Review	Kjetil Lygre

Approval X	Date: 30 November 2018	Sign.  Coordinator

USED PERSON-MONTHS FOR THIS DELIVERABLE					
No	Beneficiary	PM	No	Beneficiary	PM
1	NERSC		24	TDUE	
2	UiB		25	GINR	
3	IMR		26	UNEXE	
4	MISU		27	NIVA	
5	AWI		28	CNRS	
6	IOPAN		29	U Helsinki	1.5
7	DTU		30	GFZ	1.5
8	AU		31	ARMINE	
9	GEUS	1.29	32	IGPAN	0.23
10	FMI	0.5	33	U SLASKI	
11	UNIS		34	BSC	
12	NORDECO		35	DNV GL	
13	SMHI	0.5	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	

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

This document contains the description of the INTAROS online data catalogue, which includes data collections that were not previously available, and of the effort done in INTAROS to make selected in situ atmospheric data collections accessible through existing repositories.

Table of Contents

1. Introduction

This document contains the description of the INTAROS online data catalogue (Chapter 2) that includes data collections that were not previously available, and of the effort done in INTAROS to make selected in situ atmospheric data collections accessible through existing repositories (Chapter 3). Purpose of the dynamic, web-based catalogue is to give an overview of the present observing capability after exploitation in WP2, and to provide the necessary metadata to allow for the exploration of datasets by users, and for the data usage in the demonstrative applications (WP6).

2. The INTAROS data catalogue

2.1. Purpose and content

The INTAROS data catalogue will contain descriptions of and links to all datasets collected or generated through exploiting existing datasets and/or estimating new parameters within the project. In the first version of the INTAROS data catalogue, released at end November 2018, partners have registered the datasets that are resulting from their work in WP2 (Exploitation of existing observing systems) during the two first years of the INTAROS project. The data catalogue is the tool for displaying the data collected and/or exploited within INTAROS. It is therefore a major component of the IAOS portal described in INTAROS D5.4. The main technical development of the data catalogue is therefore carried out in WP 5 - Task 5.6. Details about the design and implementation is found in D5.4.

Each dataset is described by a set of metadata elements that capture key characteristics of the dataset. The following metadata elements have been defined in the first version of the INTAROS data catalogue:

- Title (mandatory): A descriptive title of the dataset.
- URL (mandatory): A link (URL) to the dataset within the data catalogue.
- Parameter name(s): A list of the parameters contained in the dataset.
- Project/program name(s): The project(s)/program(s) that supported data collection and/or preparation. The funding agency (e.g. European Commission) and the contract number should be included here.
- Observing system name: The name of the observing systems (if any) that collected the measurements on which the dataset is based.
- Description: A short text describing the content of the dataset. The following elements could be included
  - What kind of parameters are included in the dataset.
  - Geographic area (by name) and time period covered by the dataset.
  - Summary of processing to generate the dataset (e.g. with reference to paper or report).
  - Indication of possible use of the dataset (e.g. model validation).
- Tags: Keywords associates with the dataset, e.g. “ocean acoustics”, “sea ice”. These keywords will be used a way to quickly search for data.

- License: What license the data are/will be provided under. A list of commonly used licenses is provided in the data catalogue.
- Organisation: Which organisation the dataset belongs to.
- Visibility: This is a flag controlling the publication of a dataset registration. If the datasets description is not completed yet selecting “private” will ensure the entry is only visible to the metadata editor. When the dataset description has been completed this flag can be changed to “public” to make the description visible to all users.
- Source: a link (URL) to the dataset file(s), e.g. on a Thredds Data Server or a FTP server.
- Version: A number denoting the version of the dataset.
- Principal Investigator: The name of the PI for dataset. Multiple persons can be named.
- PI Email: The e-mail address of the PI
- Data Curator: The name of the person in charge of maintaining the dataset and its metadata.
- Data Curator Email: The e-mail address of the data curator.

One or more links to datasets and/or illustrations of the content can be added to the description. These data resources are described by:

- Name: A title of the graphics or link to the dataset.
- Description: A short description of what the graphics is illustrating, or link is referring to.
- Format: What format the graphic resource is stored in or protocol of the data access link.

2.2. Catalogue

The INTAROS data catalogue is online at https://catalog-intaros.nersc.no/. The following shows some illustration of its contents at the time of writing.

Figure 1 shows the home page of the INTAROS data catalogue. This page is comprised of a search component (upper left), short statistics of how many datasets and organisations are registered (middle left) and a general information component identifying the areas and spheres addressed by INTAROS (right). From the home page, users can also get access to Dataset pages, Organisation pages, Group pages (currently not used) and the About page. It also provides a link to the login in page (top right).

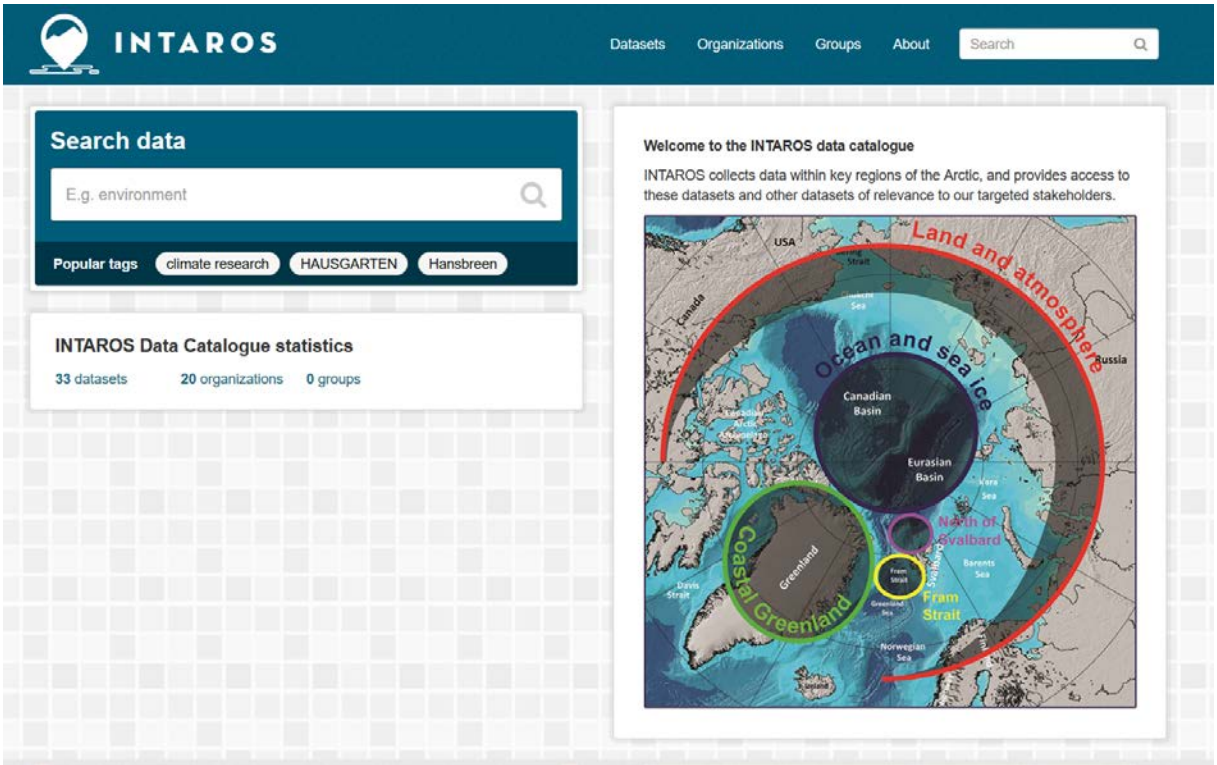

Figure 1 INTAROS Data Catalogue home page (https://catalog-intaros.nersc.no/).

When entering the Dataset page (Figure 2), users can easily search using free text search or by selecting on of the tags associated with the dataset. It is also easy to select all datasets from a specific organization (Figure 3). After identifying a dataset of interest, the user can view its metadata and proceed to view or download the dataset (Figure 4).

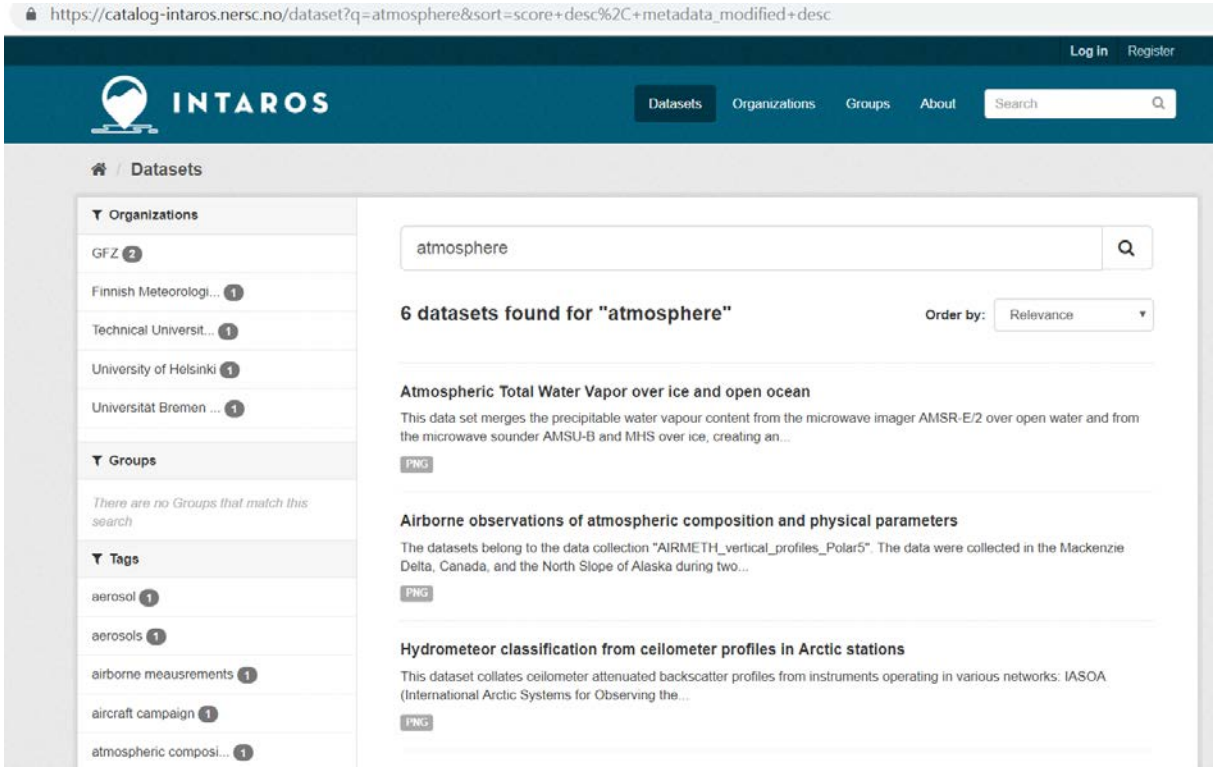

Figure 2 INTAROS Data Catalogue - Datasets page (https://catalog-intaros.nersc.no/dataset?q=atmosphere) .

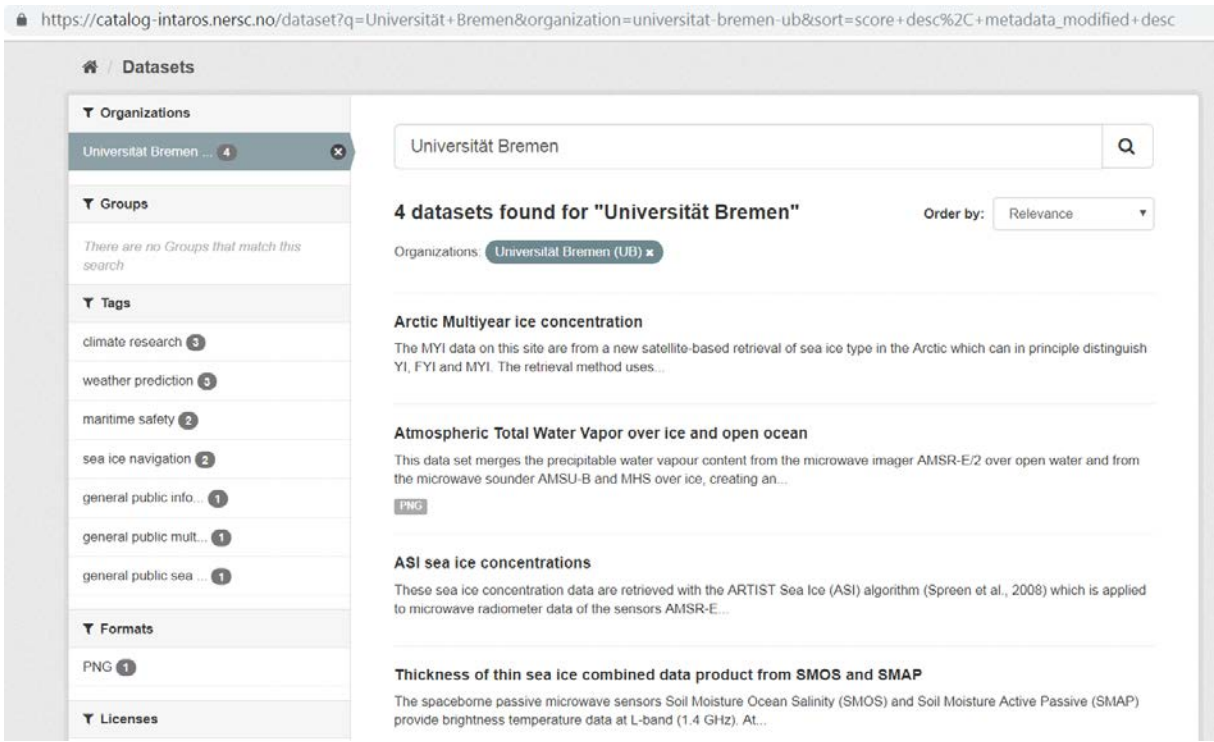

Figure 3 INTAROS Data Catalogue – Searching all dataset from a given organisation (https://catalog-intaros.nersc.no/dataset?q=Universit%C3%A4t+Bremen&sort=score+desc%2C+metadata_modified+desc).

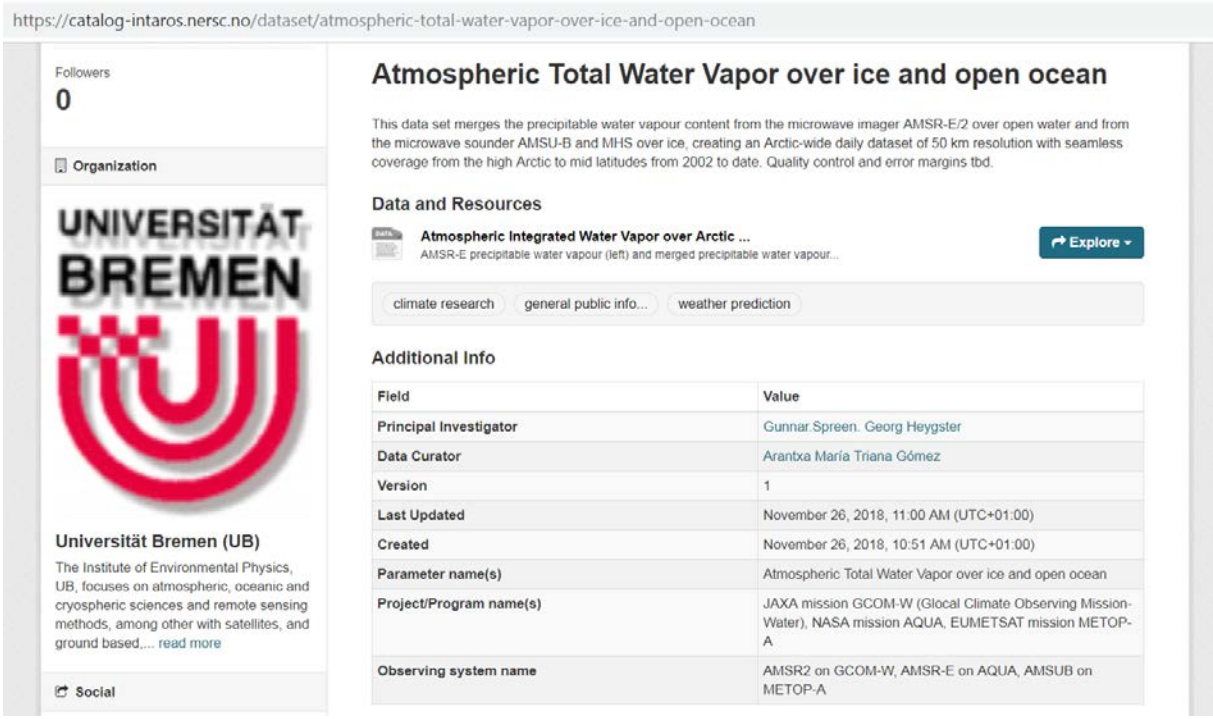

Figure 4 INTAROS Data Catalogue – Viewing (some of the) metadata for a selected dataset (https://catalog-intaros.nersc.no/dataset/atmospheric-total-water-vapor-over-ice-and-open-ocean).

3. Integration of in situ data collections into existing data repositories

This chapter provides a short description of the steps taken by the partners to make their data openly accessible and integratable into the iAOS. Different institutions have very different levels of data management and data infrastructure maturity, therefore the undertaken work is very diversified. In those cases when the data infrastructure is under building, an outlook of data repositories that will be utilized and a timeline of the work is provided.

3.1. FMI: aerosol, cloud, and greenhouse gas data from Pallas-Sodankylä (Finland), and Tiksi and Baranova stations (Russia)

Aerosol data: Aerosol data from Tiksi (Russia) and Pallas-Sodankylä (Finland) are archived in the EBAS atmospheric database (http://ebas.nilu.no/). EBAS was originally designed for the European Monitoring and Evaluation Programme (EMEP), and archives today data on atmospheric composition from ground stations around the globe as well as aircraft platforms. It is hosted in nilu, and it is also serves as data repository for the ACTRIS Data Centre (http://actris.nilu.no/) and GAW.

The repository is not yet interoperable, but this is a demand of ACTRIS, that requires the adhesion to the FAIR data principle. This is expected for European Research Infrastructures, it is stated in the ACTRIS DMP (Data Management Plan), and work is underway at the ACTRIS

Data Centre to move towards compliance. ACTRIS is member of the RI consortium participating to the ENVRI-FAIR project, which will ensure that ACTRIS will complies with the FAIR principles by the end of the project (in 4 years).
The present plan is that FMI will submit aerosol data from Tiksi by the end of 2018 (first 4 years of data). Pallas data is submitted to EBAS according to ACTRIS deadlines, i.e. always ~1 year after collection. BUT! Submission and appearance of the data are not simultaneous. Baranova Aerosol data are not yet in a state to be put in any repository this year, later will be made accessible in EBAS.

Cloud data: FMI provides ceilometer data and full cloud-profiling data from ACTRIS-Cloudnet. Ceilometer sites: Pallas, Sodankylä, Kittilä, Ivalo, Tiksi, Ny Ålesund, maybe Baranova ACTRIS-Cloudnet sites: Ny Ålesund, some data from Pallas and Sodankylä, possibly also from ARM stations Barrow, Oliktok and Summit.

All these cloud data are stored in the Cloudnet database (located in FMI), which is accessed through the ACTRIS portal. Ceilometer data is not yet accessible outside of FMI. Cloudnet ACTRIS database has an API for data access. This does not yet provide full operability as requested by FAIR (not yet discoverable for example), but it will be achieved by the end of the ENVRI-FAIR project.

ACTRIS-Cloudnet data is usually available within 3 days. ARM data processed by Cloudnet has no timeline (it is only produced on request/capacity basis). Ceilometer data is similar (usually within 1-2 days).

Officially, ACTRIS data is available through the ACTRIS portal at actris.nilu.no but may suffer from the delay in accessibility mentioned above. Direct access to ACTRIS and other Cloudnet data is via devcloudnet.fmi.fi. Ceilometer data is not yet accessible outside FMI but will have a similar location: ceilometer.fmi.fi. ARM data can also be accessed via http://www.arm.gov/data

3.2. GEUS: near-surface meteorological data from the PROMICE network

Near-surface meteorological data from the PROMICE network are already stored in a public repository, accessible via http://promice.org/DataDownload.html. To download the data, only a simple registration is required. This registration is about to be removed (by the end of 2018), as a machine-to-machine interface is going to be established for INTAROS as well as for others portals with this need. GEUS will provide to the INTAROS iAOS only a subset of the available data, as not all datasets are mature/useful enough to enter into iAOS.

3.3. IGPAN: Daily and sub-daily in situ meteorological and aerosol data from Svalbard

Atmospheric and snow data from Hornsund Station used to be available via the GLACIOTOPOCLIM Database. As this database no longer exists, IGPAN is working on opening it again. Possibly it will be re-open and accessible at the end of November/ beginning of December 2018. IGPAN is considering to make the data machine-to-machine interoperable by

constructing the spatial data infrastructure (SDI) using GeoServer. GeoServer is an open source software for sharing geospatial data. The SDI will provide services based on Open Geospatial Consortium OGC compliant standards. As for now, data remains available only via personal contact with Tomasz Wawrzyniak (tomasz@igf.edu.pl) or, in the case of diurnal data, from eklima.no.

3.4. GFZ: airborne observations of atmospheric composition and physical parameters

Airborne observations of atmospheric composition and physical parameters have been harmonized following the format requirements of the open access data repository Pangaea (https://www.pangaea.de/), and, together with their metadata, will be made openly available through Pangaea end of 2018/beginning of 2019. The database will will be populated with further related data sets in the future.

3.5. U Helsinki: Pan-Eurasian Experiment (PEEX) network

U Helsinki has built an e-catalogue "*In-Situ Atmosphere-Ecosystem Collaborating Stations - Russian Federation*" (in form of PDF file) of the PEEX stations and data, available at https://www.atm.helsinki.fi/peex/images/PEEX-catalogue-october3_Optimized-locked.pdf. In this catalogue, for each station, information in addition to measurements for the Atmosphere (meteorology) and Atmosphere (composition) is also included for the Hydrosphere, Cryosphere, Pedosphere, Biosphere, and Fluxes (where it is applicable). The access to the data is not automatic, it requires contacting the data manager. The name and e-mail address of the contact person is provided for each station. No machine-to-machine interoperability is allowed for the Russian stations that belong to the network.

---------- 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: