

Status of data catalogue and data repositories in INTAROS

INTAROS Steering Committee Meeting

GEUS
Øster Voldgade 10, Copenhagen

May 6-7, 2019

Torill Hamre, NERSC
Contributions from data providing partners



INTAROS

Agenda

- **Status of Data Catalogue**
- **WP2 Data Sets in preparation**
- **Data repositories used by WP2**
- **Data Providers Working Group**



NERSC

Status of Data Catalog



Status of Data Catalog

- Status
 - **42** datasets registered (Task 2.3++) – **~30** of these with data access
 - All datasets are searchable (free text, tag, provider, parameter, format, ...)
 - All dataset have key metadata elements
 - Contact person
 - Parameters
 - Data License
 - Data Access links (which may be a plot for now)
 - Display of individual datasets
- Next steps
 - Get WP3 and WP4 for register their new datasets (ref DP-WG)
 - Harvest metadata from external providers
 - Extend metadata schema(tbd)



Status of Data Catalog

- **AWI datasets: (6)**
 - **UDASH - Unified Database for Arctic and Subarctic Hydrography (in PANGAEA, DOI: <https://doi.pangaea.de/10.1594/PANGAEA.872931>)**
 - Benthic oxygen flux data from the Arctic Fram Strait, composite of several PANGAEA datasets, links to 56 datasets in PANGAEA
 - **Digital terrain model (DTM) of the central Fram Strait (in PANGAEA, DOI: <https://doi.pangaea.de/10.1594/PANGAEA.526589>)**
 - Biogenic particle flux through the water column at the FRAM observatory from mooring sediment traps (link to 55 datasets in PANGAEA)
 - Biogeochemical parameters from deep-sea sediments obtained at HAUSGARTEN permanent sampling stations (link to 800 datasets PANGAEA)
 - Inorganic nutrients measured on Fram-Strait water samples since 1997 (links to 29 datasets in PANGAEA)



Status of Data Catalog

- FMI datasets: (4)
 - Hydrometeor classification from ceilometer profiles in Arctic stations (in preparations, will be stored in FMI repo for ACTRIS (not ready yet))
 - ERA-CLIM2 in situ SWE (Arctic Space Centre at FMI, data from ERA-CLIM project)
 - **In situ snow observations from the Sodankylä supersite (stored in database at FMI)**
 - SMOS Soil Freeze/Thaw (will be stored in Arctic Space Centre at FMI)



Status of Data Catalog

- **Technical University of Denmark datasets: (4)**
 - Accumulated ice loss from GRACE for the Greenland Ice Sheet basin 1-8 (text file uploaded to Data Catalog, no DOI)
 - Altimetric Sea Level (data at FTP site at DTU, data available, no DOI)
 - Arctic and Greenland tide gauge data (data at FTP site at DTU, no DOI)
 - Ice mass change of the Greenland ice sheet (data at FTP site at DTU, no DOI)
- **University of Bremen datasets: (4)**
 - Atmospheric Total Water Vapor over ice and open ocean (in prep OPeNDAP server at UB, no DOI)
 - Arctic Multiyear ice concentration (available on HTTPS, in prep OPeNDAP server at UB, no DOI)
 - Thickness of thin sea ice combined data product from SMOS and SMAP (same as previous)
 - ASI sea ice concentrations (same as previous)



Partner	#datasets	# in repository
AWI	6 (4 are aggregates of 50-800 individual datasets)	6 (PANGAEA)
FMI	4	1 (repos are planned for the others)
DTU	4	3 (4th dataset uploaded to Data Catalog)
UB (Bremen)	4	4 (OPeNDAP server is being set up)
AU (Århus)	3	0 (DC includes plots)
IFREMER	3	3 (CERSAT ftp server)
NERSC	3 (4 more in preparation)	2 (NMDC)
USlaski	3	0 (institutional repo under development)
GEUS	2	2 (PROMICE)
GFZ	2	0
IGPAN	2	2 (institutional repo, requires login)
UiB (Geofag)	2	2 (1 uploaded in Data Catalog, NNSN)
NIVA	1	1 (institutional repo)
SMHI	1	1 (project database Arctic HYCOS)
UHelsinki	1	1 (project database, PEEX)
USFD	1	1 (Arctic Data Centre, US)
Total	42	29



NERSC

WP2 Data sets in preparation



WP2 Data sets in preparation

- Many WP2 partners work with
 - Setting up own data repositories
 - Finding a repo that will assign DOI to their data
 - Small adjustments of file formats
 - Preparing remaining datasets from Task 2.3



NERSC

Data repositories used by WP2



Data repositories used by WP2

- Numerous, among others:
 - PANGAEA
 - NMDC
 - Institutional repos
 - ICOS
 - EPOS
 - CERSAT
 - INTAROS Data Catalog (which is not a repo)
 - ...



Data Providers Working Group

- **Members:**
 - Torill (lead), Roberta, Agnieszka, Finn and Arnfinn
 - Open for any data provider in INTAROS
- **Mandate:**
 - Produce a roadmap for delivery of INTAROS datasets from WP2, WP3, WP4 to data repositories
 - Map requirements and issues of data providers (partners) for preparing and making available their INTAROS datasets to the iAOS
 - Compile guidelines for the preparation of different categories of datasets
 - Clarify the purpose of the INTAROS Data Catalog and the iAOS Portal, and how the data providers in the project can contribute to these



Data Providers Working Group

- **Work done:**
 - Prepare workflow diagram for data delivery
 - Contact data providing partners in WP2, 3 and 4
 - Follow up on responses
 - Summarise in Google sheet
(https://docs.google.com/spreadsheets/d/1hxpSmUmlj7anc83xG1w2nMjEUels62BvVTEjRw3_IFg/)
 - Start work of joint format for ocean mooring data
 - Collect documents on relevant standard formats
 - Plan for meetings
- **Group work area on Google Docs**
(<https://drive.google.com/drive/folders/1pvdL-URWy1u3CUCMi3gkLtQ1PT1mexSB>)



Dataset

Briefly state the content/nature of your INTAROS dataset(s):
...

Data and metadata format

(need to be standard, or such that users can easily understand)

You know which metadata&data standard to apply.
Please describe it:
...

Guidance from Data Providers WG

You don't know which metadata&data standard to apply.
Please describe what information you need to decide on a metadata&data standard:
...

Data and metadata tools

You know the software needed to format your data/metadata.
Please provide name and/or link to it:
...

Guidance from Data Providers WG

You don't know the software needed to format your data/metadata.
Please provide name and contact info of specialists who may know it:
...



Data delivery time

When will your dataset be ready for uploading to a data repository/server?
(realistic estimate of delivery month):
...

Data repository

(needs to have services that allow the data to satisfy the FAIR principles)

You know where to store your data.

Please provide name and URL to data repository/server:
...

You don't know where to store your data.

Please explain what information you would need to be able to select a data repository/server:
...

Data repository search

List of repositories used by INTAROS WP2-6 partners:
([Link](#))

Search engines for data repositories that enable the FAIR principle:
(re3data.org)

List of recommended repositories from the Data Delivery Task Force
(in progress)



End of presentation

