WP5: Data integration and management

Synthesis of work on data catalogue and iAOS portal

Torill Hamre (WP5 co-lead)
Nansen Environmental and Remote Sensing Center, Norway

INTAROS Final Meeting, Online meeting, 20-21 January 2022

Synthesis of Task 5.6 iAOS portal development

1. Objective and Tasks
2. Main achievements
3. Expected impact
4. Challenges
5. Recommendations

WP5 Task 5.6:
iAOS portal and the data catalogue

Lead: NERSC, T. Hamre;

Contributors: TERRADUE

Objective: Provide: (1) access to all data generated with support from INTAROS in the other workpackages (WP2-6); (2) a common entry point for selected datasets collected or generated as part of other project and relevant datasets from other Arctic data systems; (3) links to services from INTAROS and spin-off projects.

Task 5.6 iAOS portal development

Specific tasks:

- Development of iAOS portal comprised of
  - INTAROS data catalogue describing all datasets generated with support from the project
  - Portal Data catalogue harvesting dataset descriptions (metadata) from other Arctic Data Systems
  - ARCMAP inventory of Arctic in situ observing systems with tools for presenting statistics and maps
- Supporting partners of WP2, WP3, WP4, WP5, WP6 in registration of dataset in the INTAROS Data Catalogue
- Promoting services, tools and showcases developed in the project or in collaborating projects

INTAROS

European Commission

Main achievements

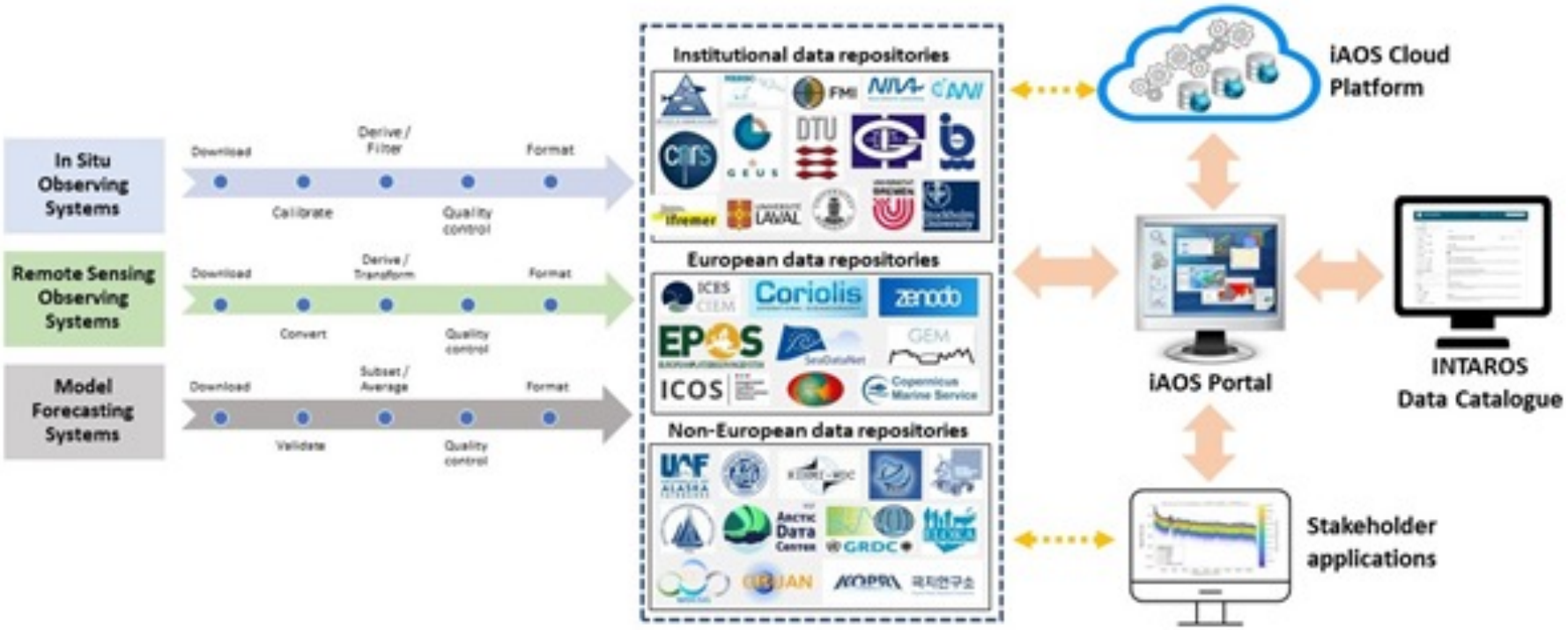


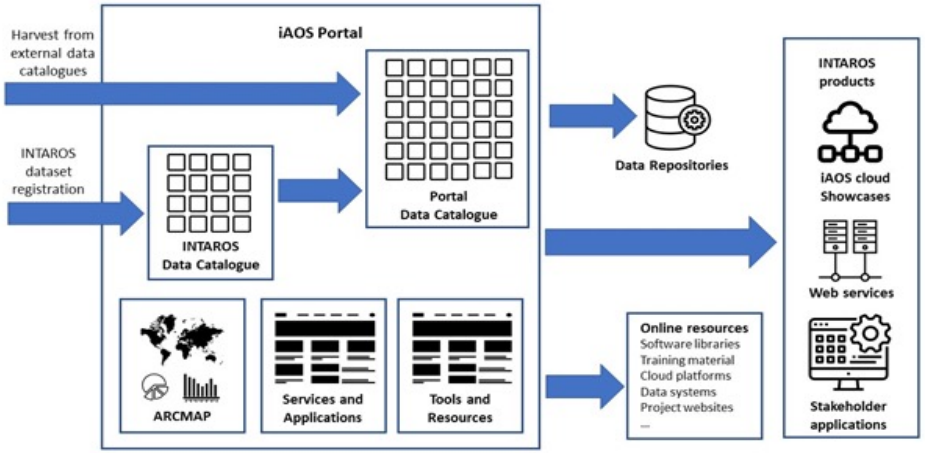
Figure 1. Data value chains for integrating INTAROS data into the various iAOS subsystems.

INTAROS

European Commission

iAOS Portal

- Data catalogue and portal platform established
- Generic portal design
  - INTAROS data catalogue
  - Portal data catalogue
  - ARCMAP
  - Services & Applications
  - Tools and resources
- Data stored in sustained repositories
- Open standard interfaces
- Portal data catalogue
  - >500 datasets harvested
  - Leveraging open APIs
  - Reusing community plugins



Major components of the iAOS Portal and their interconnections.

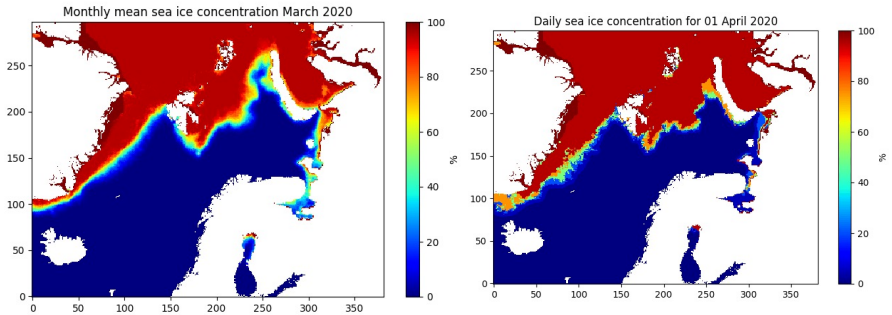
https://portal-intaros.nersc.no/

INTAROS

European Commission

iAOS Portal

- Promotion spaces for
  - Showcases/Applications
  - Cloud services
  - WPS services
  - Geostatistics libraries
- Open for new entries
- Marketplace for future iAOS developments
- Will be maintained by NERSC



INTAROS

INTAROS Datasets Organizations Topics About Services Tools Search

Services

INTAROS develops services for multiple user segments, including science, environmental and ecosystem management, natural hazards monitoring, risk assessment and support for m itigation planning. Services combine in situ, remote sensing and model data from a multitude of providers to provide a data product that user can apply in their daily work.

Services are developed using the iAOS Cloud Platform and Jupyter Notebook. Some examples of INTAROS services are shown below.

Geostatistics for gridding in situ oceanographic data

This service was developed to generate ocean temperature and salinity fields for validation of climate model projections. Marine in situ observations are typically scattered in space and time, while models generate gridded data. The service uses geostatistic methods to interpolate a dispersed set of in situ point measurements to a regular grid, allowing comparison with model projections.

The service was applied to a 22 year long time series of CTD data held by the Norwegian Marine Data Centre. In total the input data amounted to 5.5 billion samples measured over 63500 positions (vertical profiles). Figure 1 shows one of the outputs from the service, a gridded field of ocean temperature for the whole time period. Read more about the geostatistis service in INTAROS Deliverable D5.6.

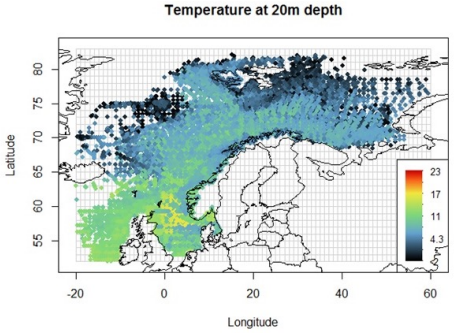


Figure 1. Base map of the whole IMR dataset – Temperature (°C) at 20m depth.

Jupyter Notebook files: RGeostats workshop
Software packages needed:

- Latest Conda package build for RGeostats (build recipe)
- Latest Conda package build for RIntaros (build recipe)

Developer: ARMINES

Analysis of passive acoustic data

This service processes and characterizes passive acoustic data, and produces spectrograms and noise statistics plots that can be used for analysis in combination with time serie s of satellite remote sensing derived parameters. It is implemented using the R version of the open source PAMGuide software package, and has extended s upport for new data formats (NetCDF) and data access through the OPeNDAP protocol.

The service has been tested with datasets from several sources (NERSC, CNRS, PANGAEA). Figure 2 shows an example of passive acoustic collected by CNRS in Kongsfjorden, Svalbard, as part of the INTA ROS field campaigns. The spectrum is dominated by low-frequency noise below 10 Hz. Local peaks around 10 Hz and 80 Hz are also seen, which could be mammal vocalization. Intermit tent broad-band signals are also seen in the spectrogram. Read more about the passive acoustic service in INTAROS Deliverable D5.7.

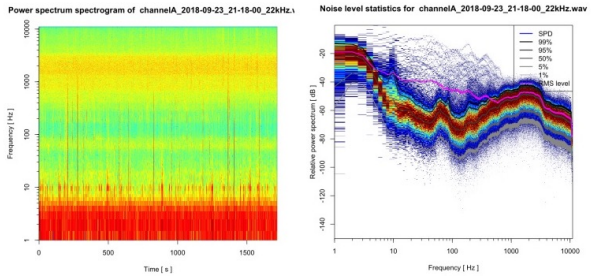


Figure 2. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the passive acoustic service when analysing acoustic data collected in Kon gsfjorden, Svalbard, during the INTAROS project.

Jupyter Notebook files: PAMGuide-R-Tutorial
Software packages needed:

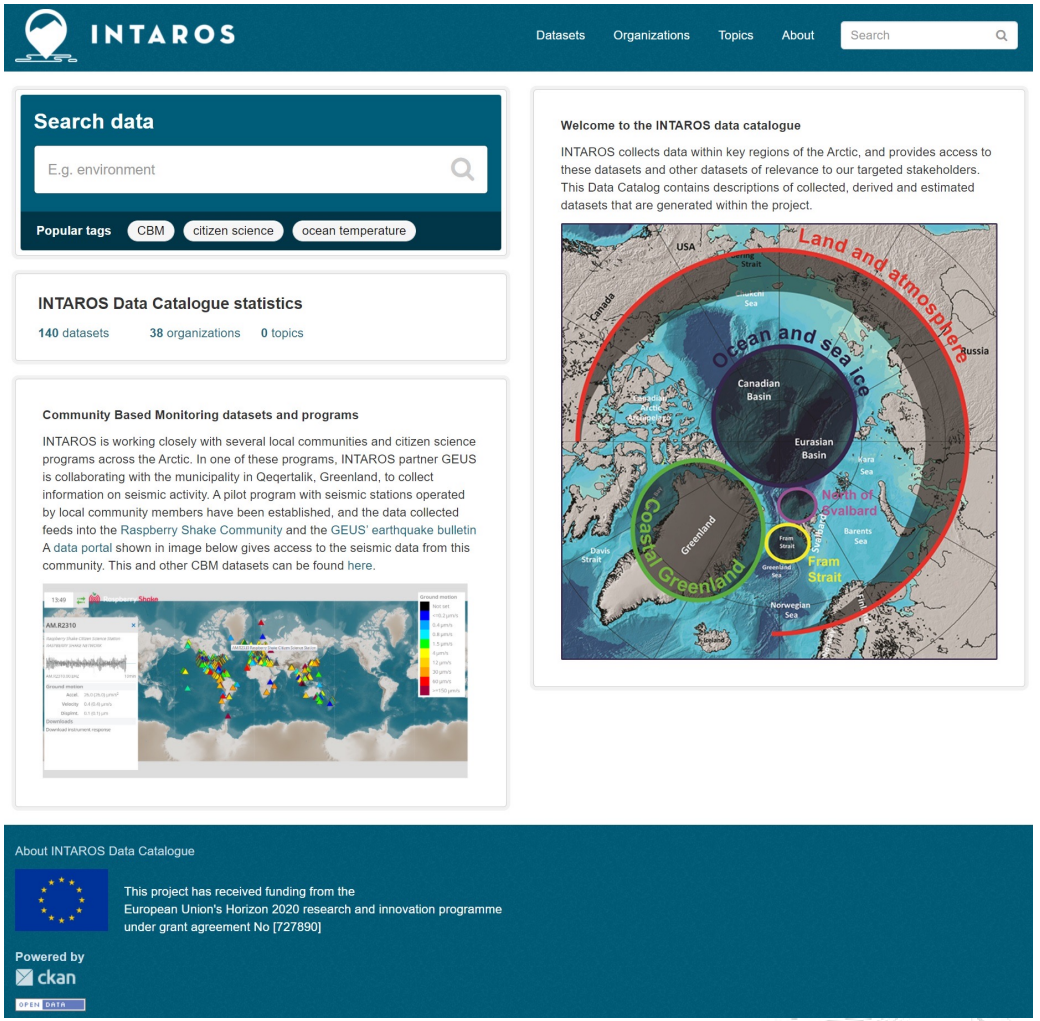
- R
- PAMGuide
- Jupyter notebook

Developer: NERSC

INTAROS Data Catalogue

- Window to INTAROS datasets
- Currently: 140 datasets
  - WP2: 55
  - WP3: 26
  - WP4: 15
  - WP5: 12
  - WP6: 9
  - Russian partners: 23

  From 38 organisations
- Same platform as iAOS portal
- Share metadata schema
- Multi-faceted search
- Will be maintained by NERSC



https://catalog-intaros.nersc.no/

INTAROS

European Commission

INTAROS data catalogue

- Statistics for INTAROS Data Catalog:
  - 140 datasets registered (vs 129 at GA 2021, 48 at GA 2020)
  - In progress (metadata private): 8 datasets
  - Planned (D5.9): 6 (2 already registered)
  - Planned from UAK 2021 cruise: 5+
- Some recent datasets:
  - GCOS_Reference_Upper_Air_Network (NUIM)
  - Marine surface observations from ICOADS (NUIM)
  - Global land meteorological observations from NOAA and C3S (NUIM)
  - Atmospheric data from eddy-covariance measurements of turbulent fluxes from icebreaker Oden during the Ryder 2019 expedition (MISU)
  - Atmospheric data from weather station on icebreaker Oden during the Ryder 2019 expedition (MISU)

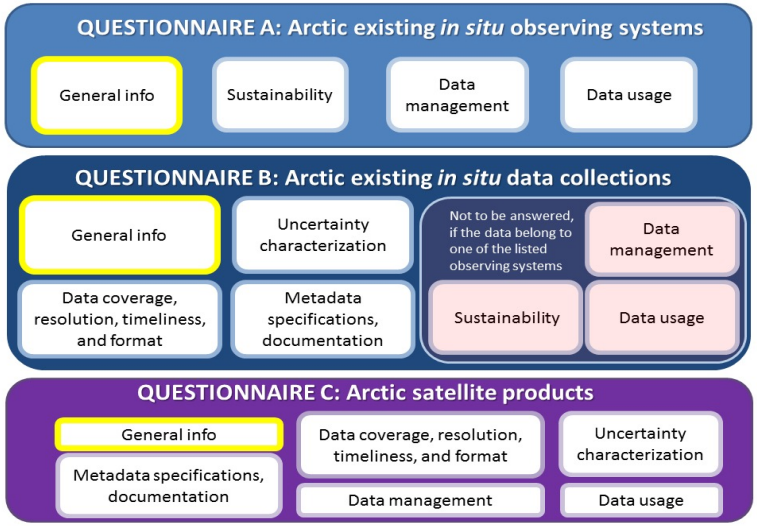
  ...and numerous updates (e.g. DOIs or other data links)

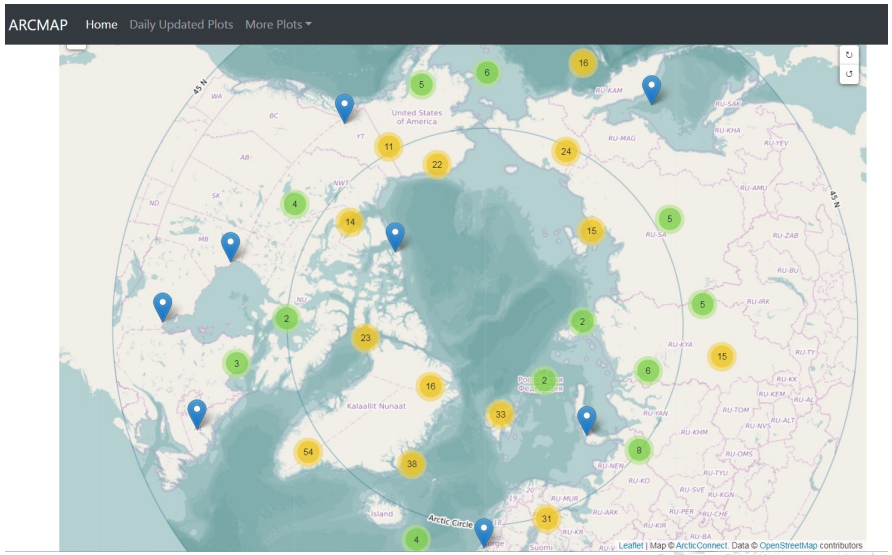
INTAROS

European Commission

ARCMAP

- INTAROS survey in WP2
- Foundation for the assessment of in situ observing capacity
- Well received by EC, SAON, ...
- Spin off project from INTAROS funded by the Norwegian Ministry for Climate and Environment (Arctic Mapping)
- Builds on and extends the INTAROS survey
- Develops methodology and tools for keeping survey information updated and analyzing evolution over time
- Additional support from NERSC Basic Funding



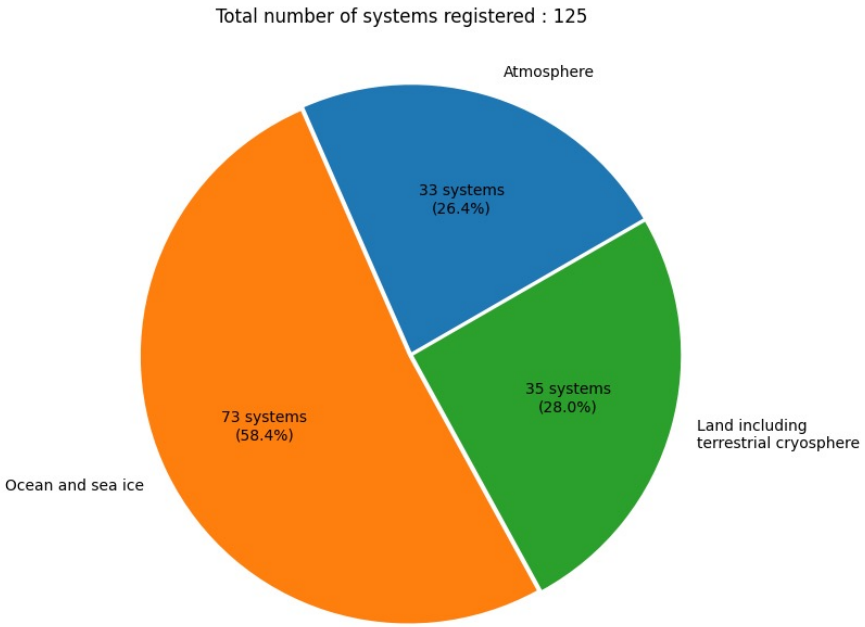


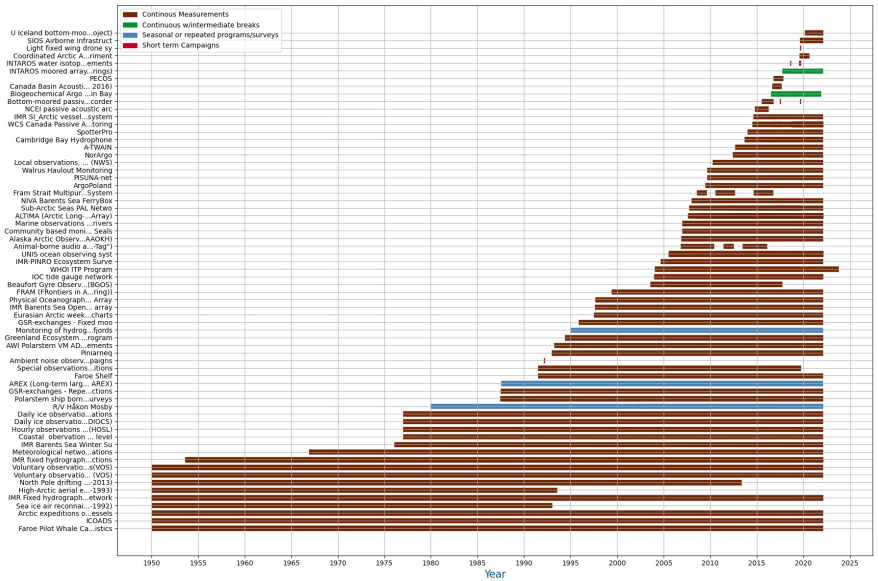
INTAROS

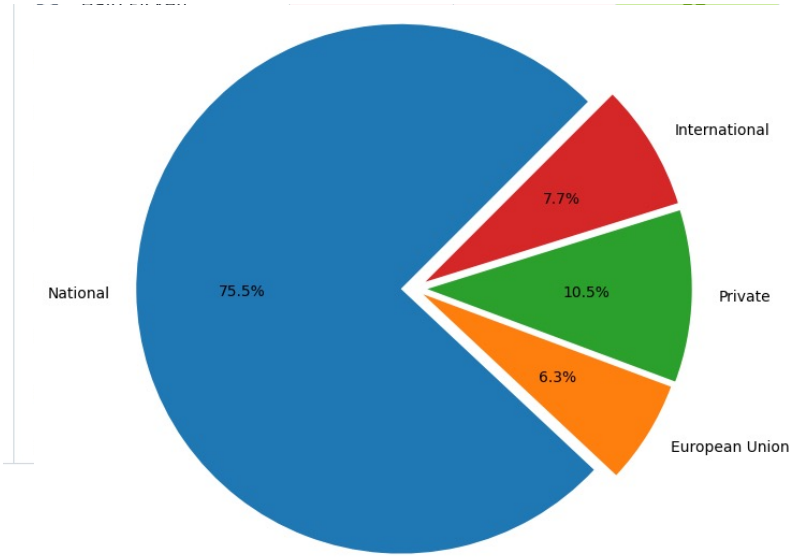
European Commission

ARCMAP

- Number of observing systems is steadily increasing
  - 2018: 49
  - 2020: 105
  - 2022: 125
- Thanks to all respondents!
- And to the ARCMAP support team!
- Daily plots at https://ci.nersc.no/client/plots.html



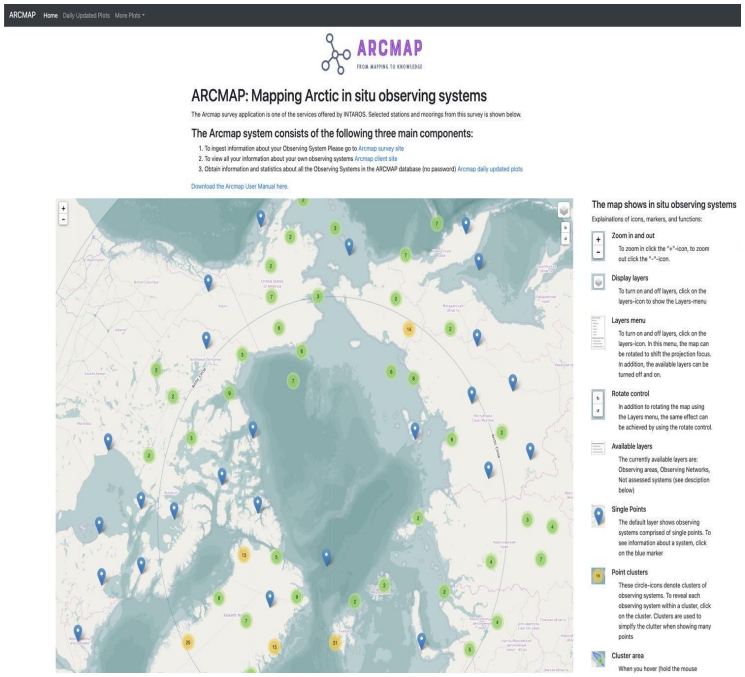


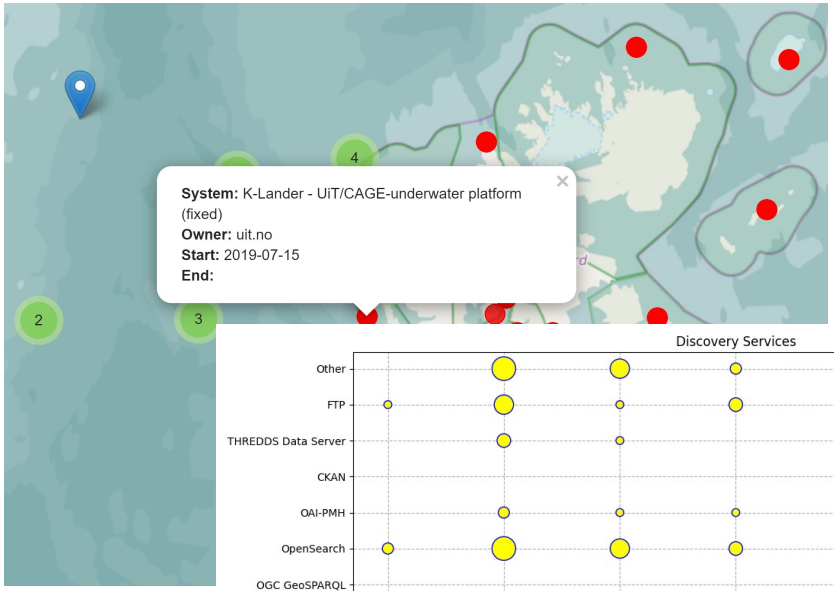


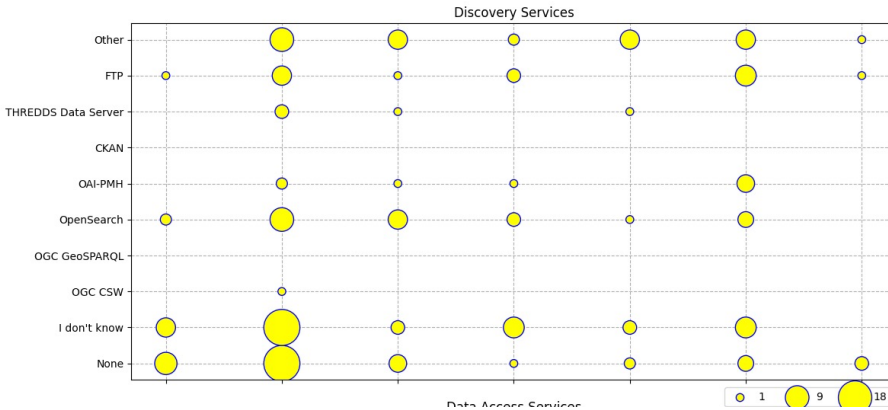
Funding sources

ARCMAP

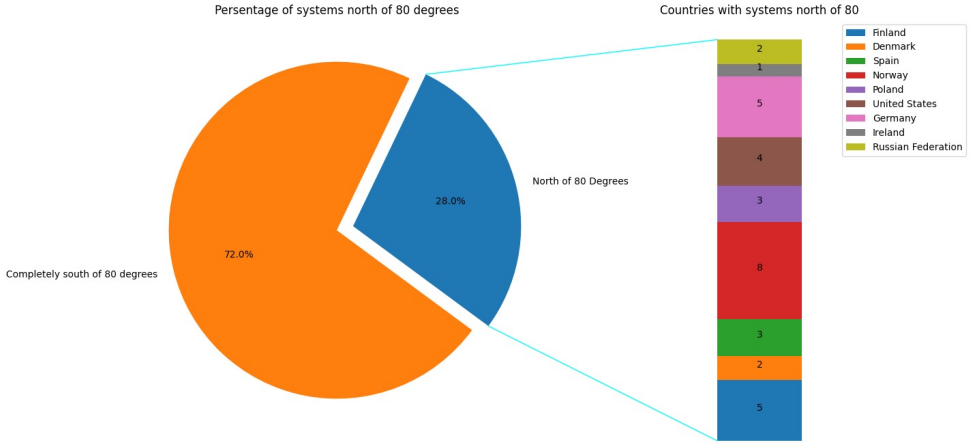
- Latest enhancements
- Automation & Robustness







- Continued interest from SAON, Arctic PASSION
- Will be maintained by NERSC



INTAROS

Expected impact

Expected impact

- iAOS portal and INTAROS data catalogue offers a way to promote datasets, services, applications and other resources
  - Open, accessible through a common web browser
  - Standard API for metadata harvesting
- Portal and data catalog platform can be used to sustain the iAOS portal and INTAROS Data Catalog
  - Open-source framework with large user community
  - Modular architecture allow for extension
  - Many plugins ready for use (e.g. metadata harvesting)
- ARCMAP provides a unique system for assessing Arctic in situ observing capacity
  - Supports open APIs for data sharing (according to FAIR, licensed)
  - Adaptable for other areas

INTAROS

European Commission

Challenges

Challenges

- iAOS portal and INTAROS data catalogue development
  - Wide variety of data from different scientific domains, citizen science and community-based monitoring system
  - Differences in standardisation level of metadata and data
  - Data stored in many Arctic data systems, with differing APIS for metadata search as well as data access
- ARCMAP
  - No standards for this kind of survey (with 'advanced' geographic information)
  - Need for extension of WQ.IO (w/'simple' GIS capabilities)
  - Optimal organisation of survey, formulation of questions

INTAROS

European Commission

Recommendations

Recommendations

- iAOS Portal and INTAROS Data Catalog
  - Further develop and operate to promote datasets and other relevant resoruces for Arctic data projects and initiatives
  - Improve interoperability between (selected) data repositories
  - Enhance search and plotting capabilities, extend harvesting
  - Extend functionality for web and cloud service
- ARCMAP
  - Work with SAON, POAwg and other observing assets systems to define a joint metadata standard and license for data sharing
  - Extend the survey in new projects to gather more information filling gaps (regions, scientific domains)
  - Develop tailored plots and assessment reports using the accumulated information

INTAROS

European Commission

Deliverables

- D5.4 iAOS portal with user manual
  (Lead NERSC, M24)
  Delivered 31 Dec 2018 – Approved 20 Apr 2020
- D5.12 iAOS portal with user manual
  (Lead NERSC, M54
  Delivered 2 Jul 2021 – Under evaluation

INTAROS

European Commission

Comments and discussion (please use the chat)