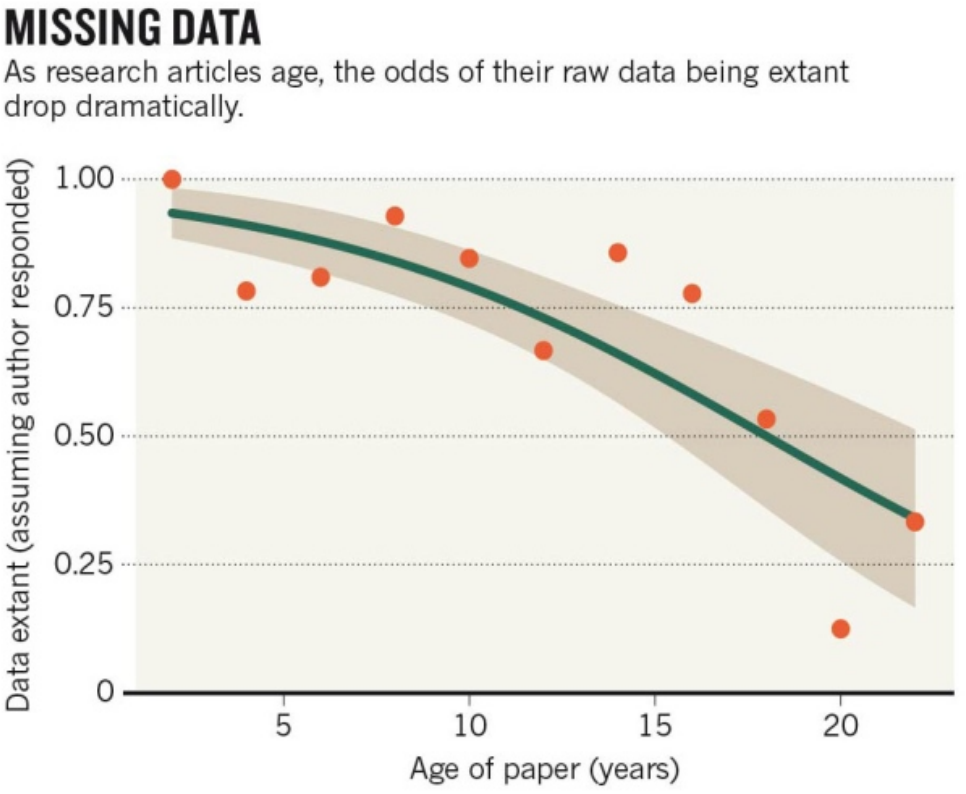
Norwegian and international data infrastructures
Services for providers and consumers
Focusing on the Arctic

Øystein Godøy, Stein Sandven,
Markus Fiebig and Trond Thorbjørnsen

Loosing scientific data

- Decline can mean 80% of data are unavailable after 20 years.

–Gibney and Van Noorden (2013), Nature



NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

Why bother with structured data management?

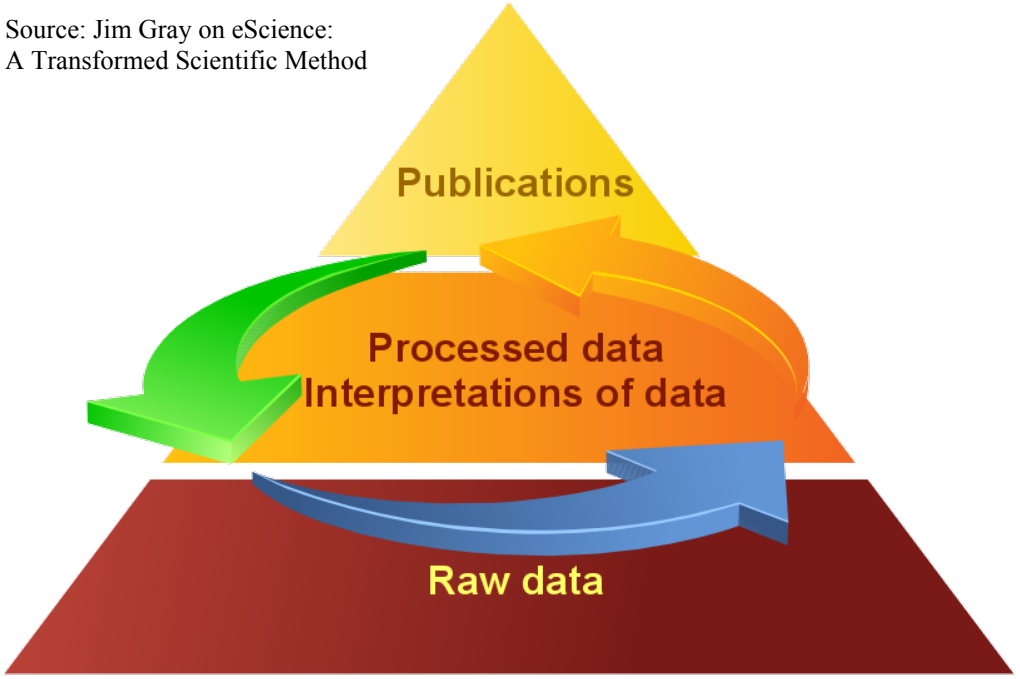
- Maximise public investment in data collection and production
- Promote scientific collaboration
- Promote interdisciplinary science
- Promote scientific transparency
- Leave a legacy

- Science paradigms
- according to Jim Gray
  - empirical science
  - theoretical science
  - computational science
  - data exploration science

NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

All scientific data online

Source: Jim Gray on eScience:
A Transformed Scientific Method



- Many disciplines overlap and use data from other sciences
- Science, government agencies and companies get a broader data background
- Internet can unify data, software and literature
- Go from literature to computation to data back to literature
- Information is at your fingertips for everyone and everywhere
- Potentially Increased Scientific Information Velocity
- Potentially Huge increase in Science Productivity

NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

The FAIR Guiding Principles for scientific data management and stewardship

- To be Findable:
  - F1. (meta)data are assigned a globally unique and persistent identifier
  - F2. data are described with rich metadata (defined by R1 below)
  - F3. metadata clearly and explicitly include the identifier of the data it describes
  - F4. (meta)data are registered or indexed in a searchable resource

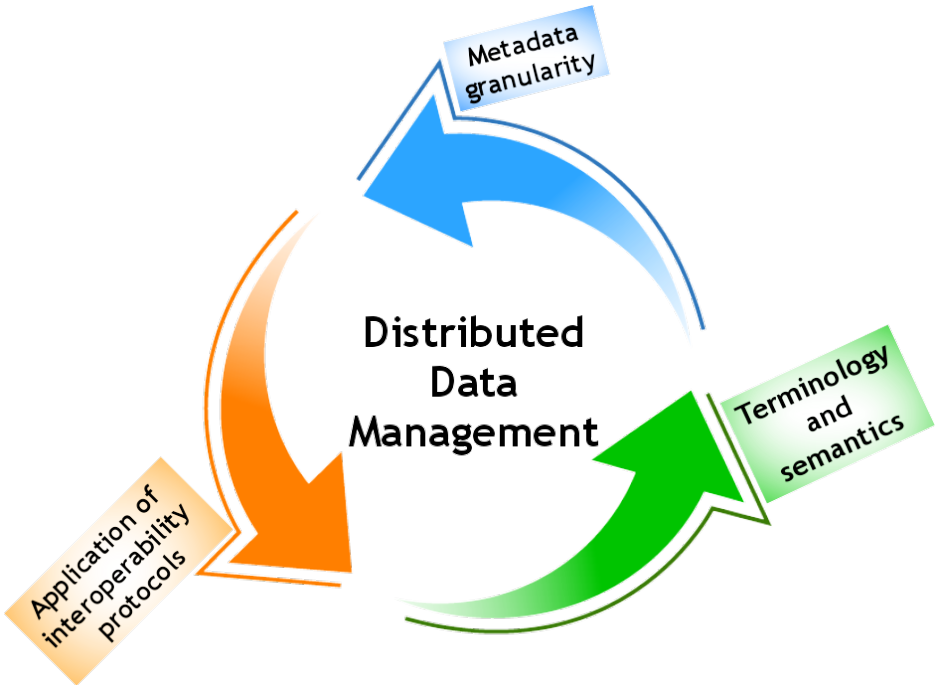
- To be Interoperable:
  - I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
  - I2. (meta)data use vocabularies that follow FAIR principles
  - I3. (meta)data include qualified references to other (meta)data

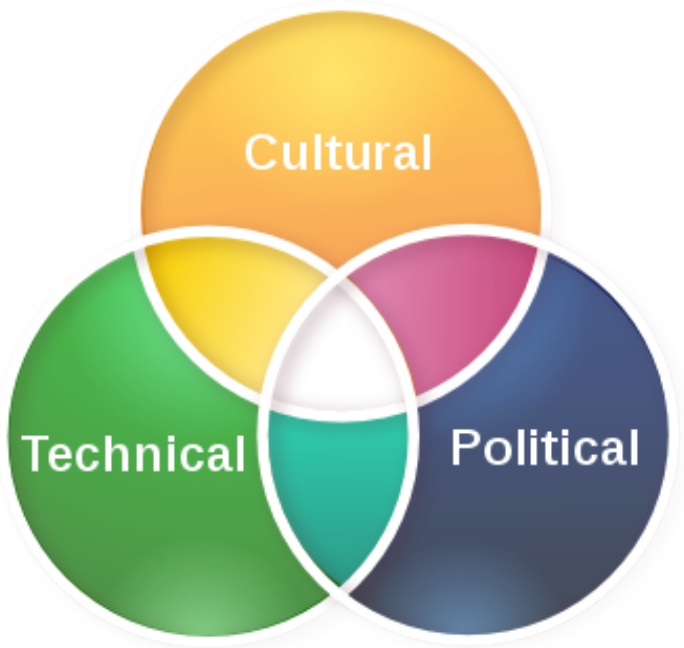
- To be Reusable:
  - R1. meta(data) are richly described

Types of metadata

- Discovery metadata
  - ***who*** measured, simulated or analysed ***what***, ***where***, and ***when*** as well as ***conditions for reuse*** and ***access mechanisms*** for the data
  - to enable users to find appropriate data for the task
- Use metadata
  - ***identification of the variables/parameters*** generated, ***units of variables/parameters***, how ***missing values*** are encoded, definition of ***grid and map projections*** for gridded data, ***methodology applied in space or time*** to achieve the values in a dataset etc
  - to enable users to properly understand the data found

Challenges

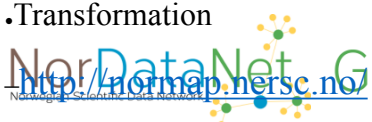




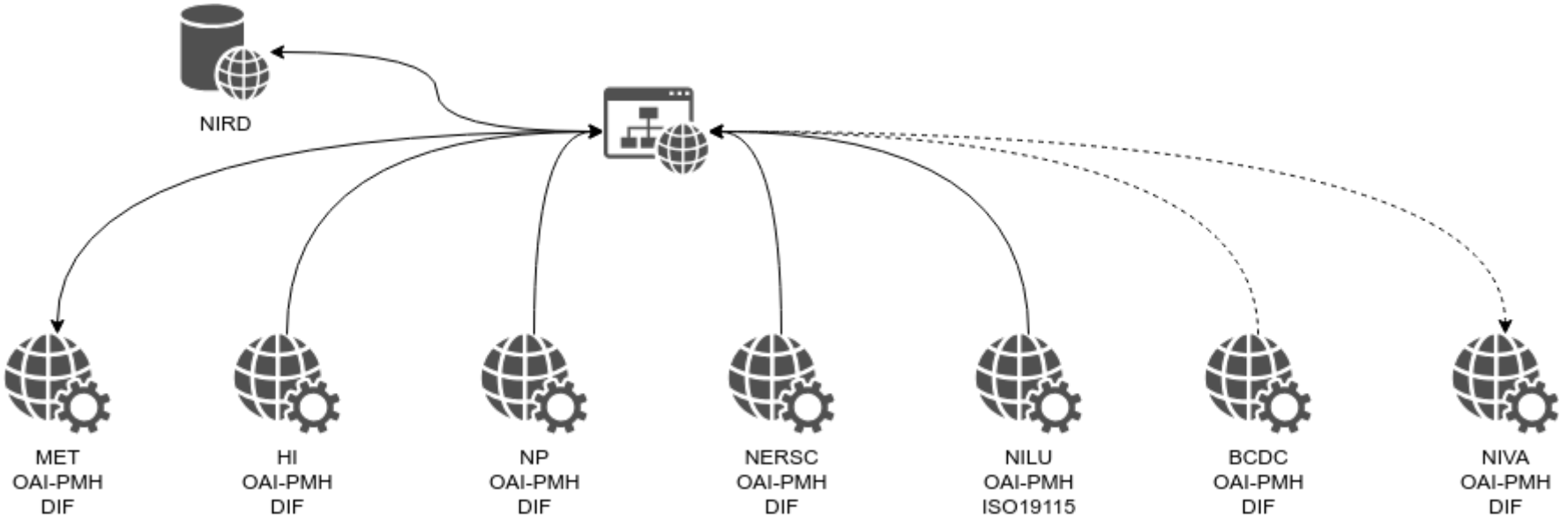
NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

Relevant national activities

- Norwegian Marine Data Centre
  - Data from the marine domain
  - Unified search
  - Download (individual and bundled)
  - http://www.nmdc.no/
- Norwegian Satellite Earth Observation Database for Marine and Polar Research
  - Remote sensing products for ocean and polar regions
  - Unified search
  - Download
  - Visualisation
  - Transformation
  - http://normap.nersc.no/
- Norwegian Scientific Data Network
  - Interdisciplinary data management
  - Unified search
  - Download (individual and bundled)
  - Visualisation
  - Transformation (if available over OPeNDAP)
  - https://www.nordatanet.no/
- GeoAccessNO
  - Pilot project examining combination of physical and digital data, software and literature
  - Not a service
  - https://www.geoaccessno.no/




National nodes



NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

NorDataNet
Norwegian Scientific Data Network

METADATA Search

View | Edit | Track | Node export

Full text search

Search words

Data collection period

Start date
1967-01-01
End date
yyyy-mm-dd

Bounding box

Institutions

Investigator

Search

Geographical search

File Edit View History Bookmarks Tools Help

https://www.nordatanet.no/results/?page=1

Most Visited | Getting Started | NorDataNet | Arctic D...

Home | About | Search for data | Submit data | Support | Login

NorDataNet
Norwegian Scientific Data Network

Number of datasets found: 8

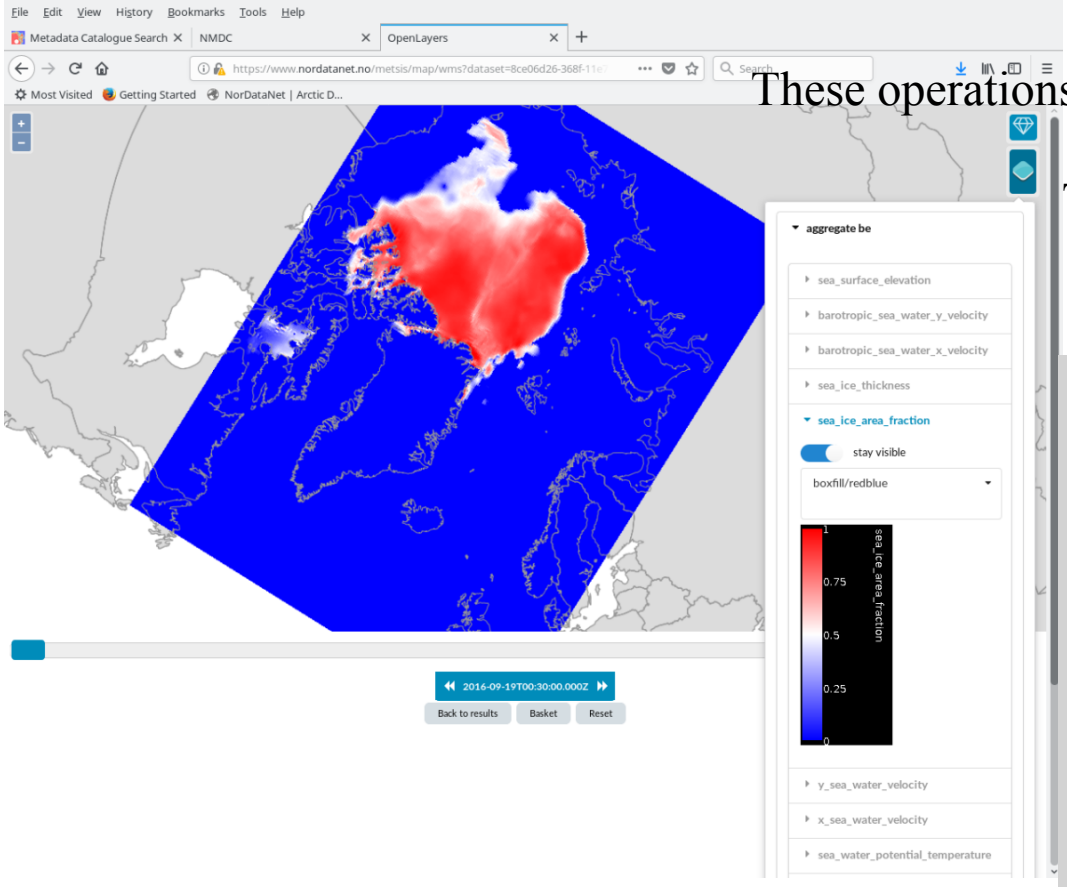
Search Results

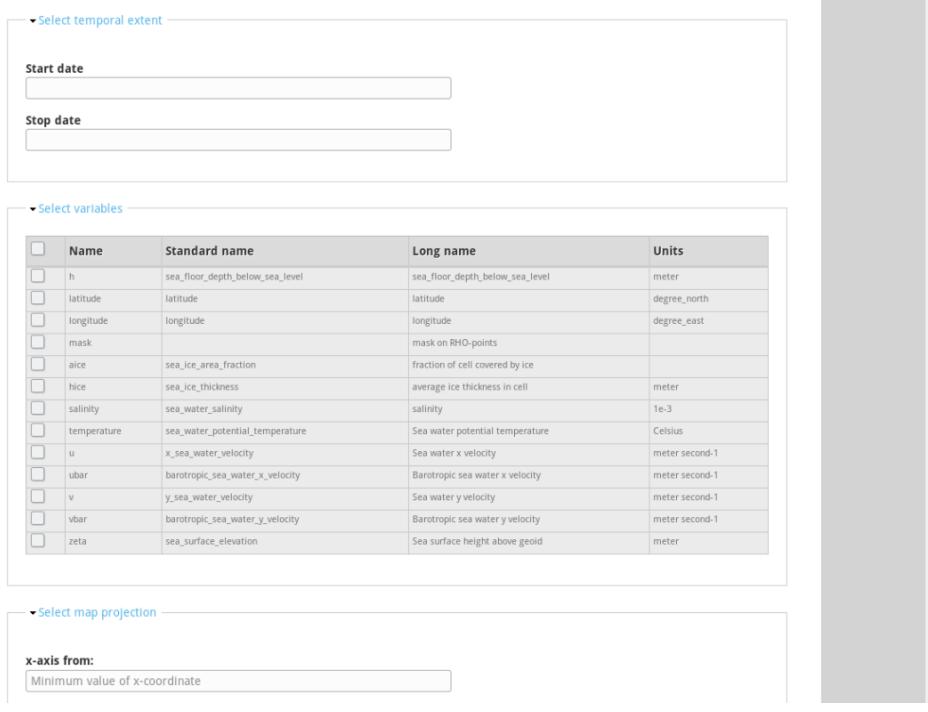
Dataset name	Institutions	Abstract	Collection period
met-arome-arctic-2p5km-extracted Download data Metadata Transform	Norwegian Meteorological Institute	Extracted variables based on the latest run of the AROME-Arctic model, without additional post-processing. Data on surface, and selected model and pressure levels. Horizontal data resolution is 2,5km. The forecast is updated 4 times per day. For historical runs see http://thredds.met.no/thredds/catalog/aromearcticraw/catalog.html	2016-02-01T12:00:00Z to
met-arome-arctic-2p5km-	Norwegian Meteorological	Post processed forecasts based on the latest run of the AROME-Arctic model. Parameters like temperature, cloud cover, precipitation and wind have gone through additional post-processing. Horizontal data resolution is 2,5km.	2016-02-01T12:00:00Z to

NorDataNet Norwegian Scientific Data Network | GeoAccessNO | NMDC norsk marint datasenter | NORMAP | INTAROS

These operations can be used through the basket on multiple products s

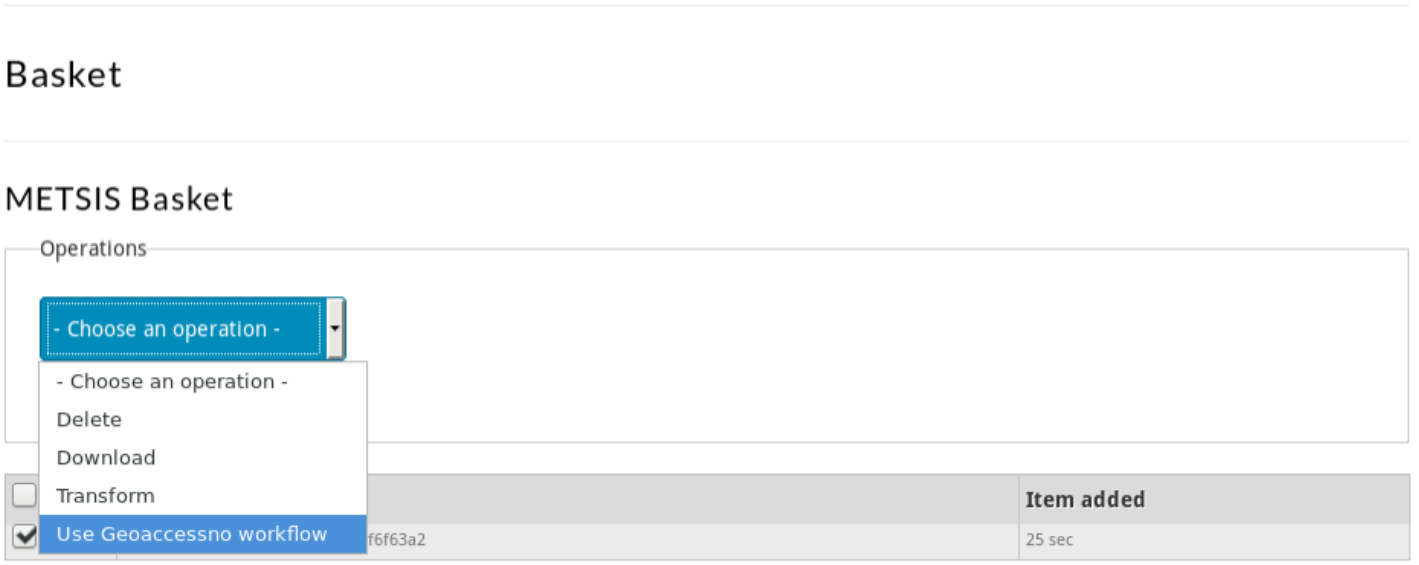
The basket allows bundling of data as served as well..





NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

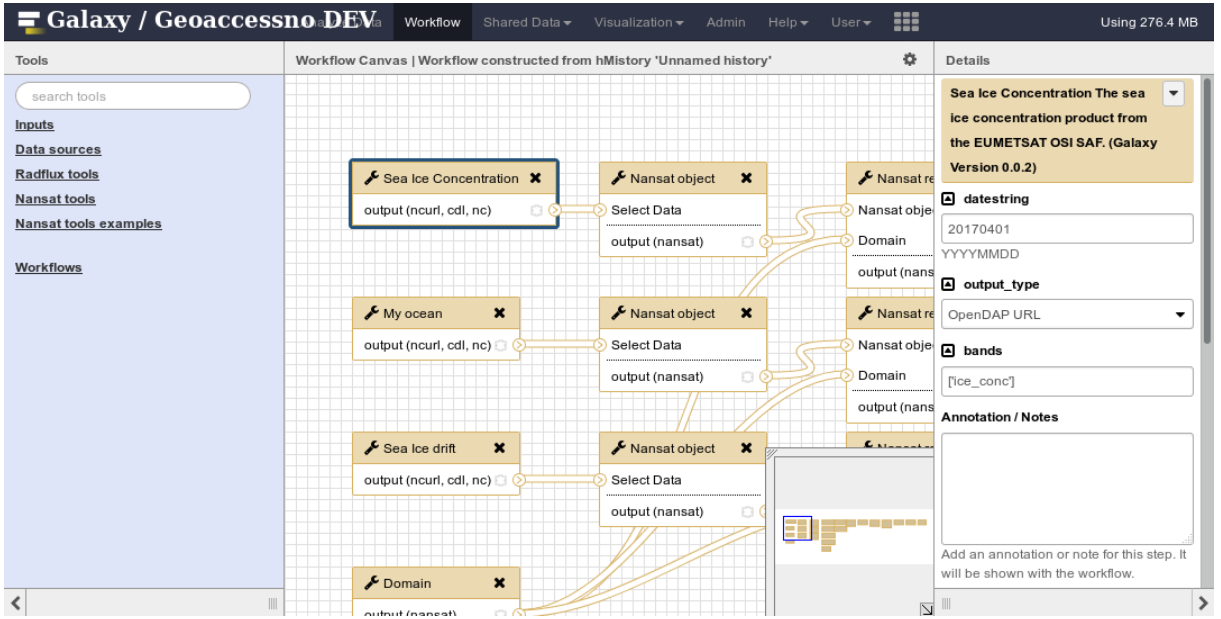
Norwegian Scientific Data Network



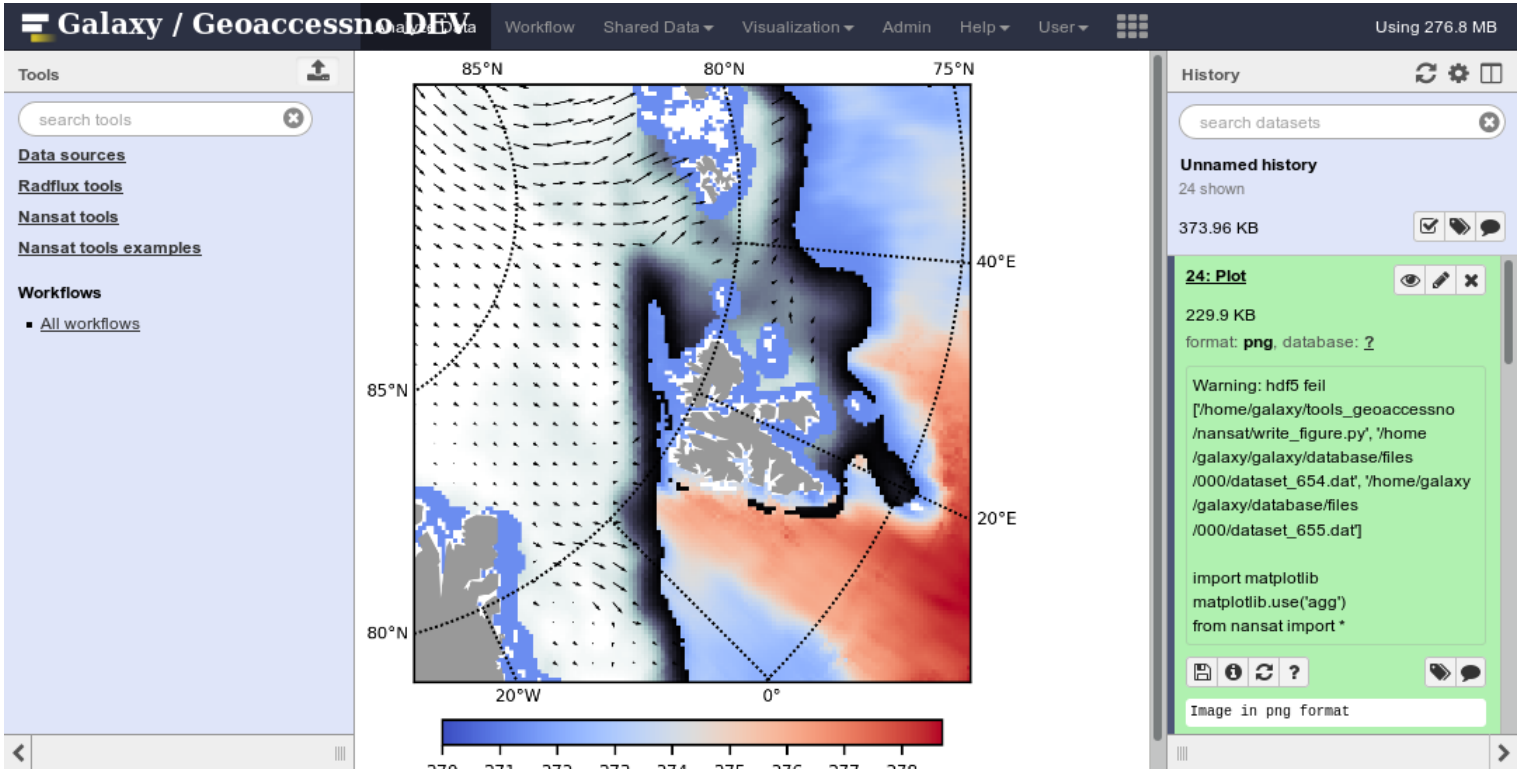
- API access
- User can run tools and workflows externally
- Enables integration in web services

NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

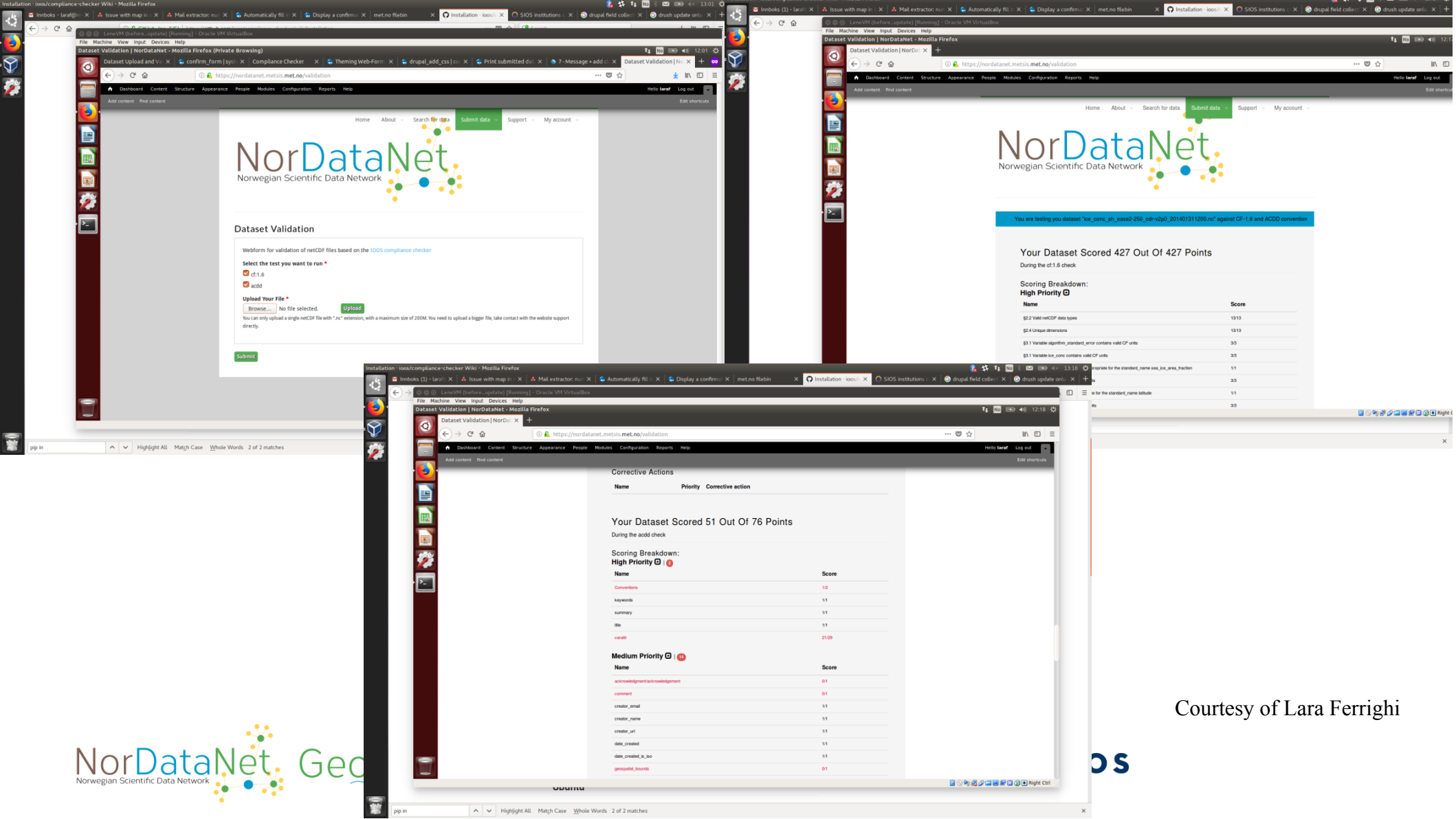
Virtual research environment



- Based on the Galaxy framework
  - Visualise and edit the workflow graphically
  - Workflow can be shared with other users
- Integrates data and software developed by the scientific community
- Creates workflow from the



NorDataNet Norwegian Scientific Data Network
GeoAccessNO
NMDC norsk marint datasenter
NORMAP
INTAROS



Courtesy of Lara Ferrighi

Relevant international activities

- INTAROS (Integrated Arctic Observation System)
  - Data from different domains
- Land and cryosphere
- Ocean and sea ice
- Atmosphere
- Community-based monitoring
- Natural Hazards
  - Distributed data centres, but unified search
  - Data visualisation and download
  - Processing services (tools)
  - https://intaros.nersc.no/

- SIOS-KC (SIOS Knowledge Centre)
  - Interdisciplinary data management
- Open data space
- Unified search across distributed data centres
- Download (individual and bundled)
- Visualisation
  - Maps
  - Timeseries in progress
  - Profiles planned
- Transformation
  - if data available over OPeNDAP
  - Working towards VRE, starting with simple transformation services
  - https://sios-svalbard.org/

INTAROS – Integrated Arctic Observation System

A project funded by EC - H2020-BG-09-2016
Coordinator: Stein Sandven, Nansen Environmental and Remote Sensing Center, Norway

Overall objective: to develop an efficient integrated Arctic Observation System by extending, improving and unifying existing and evolving systems in different regions of the Arctic

http://intaros.eu/

47 partners from 20 countries
Start date: 01 December 2016 - Duration: 5 year

NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

An integrated Arctic Observing System needs to cover

1. Atmosphere themes
2. Ocean themes
3. Terrestrial themes

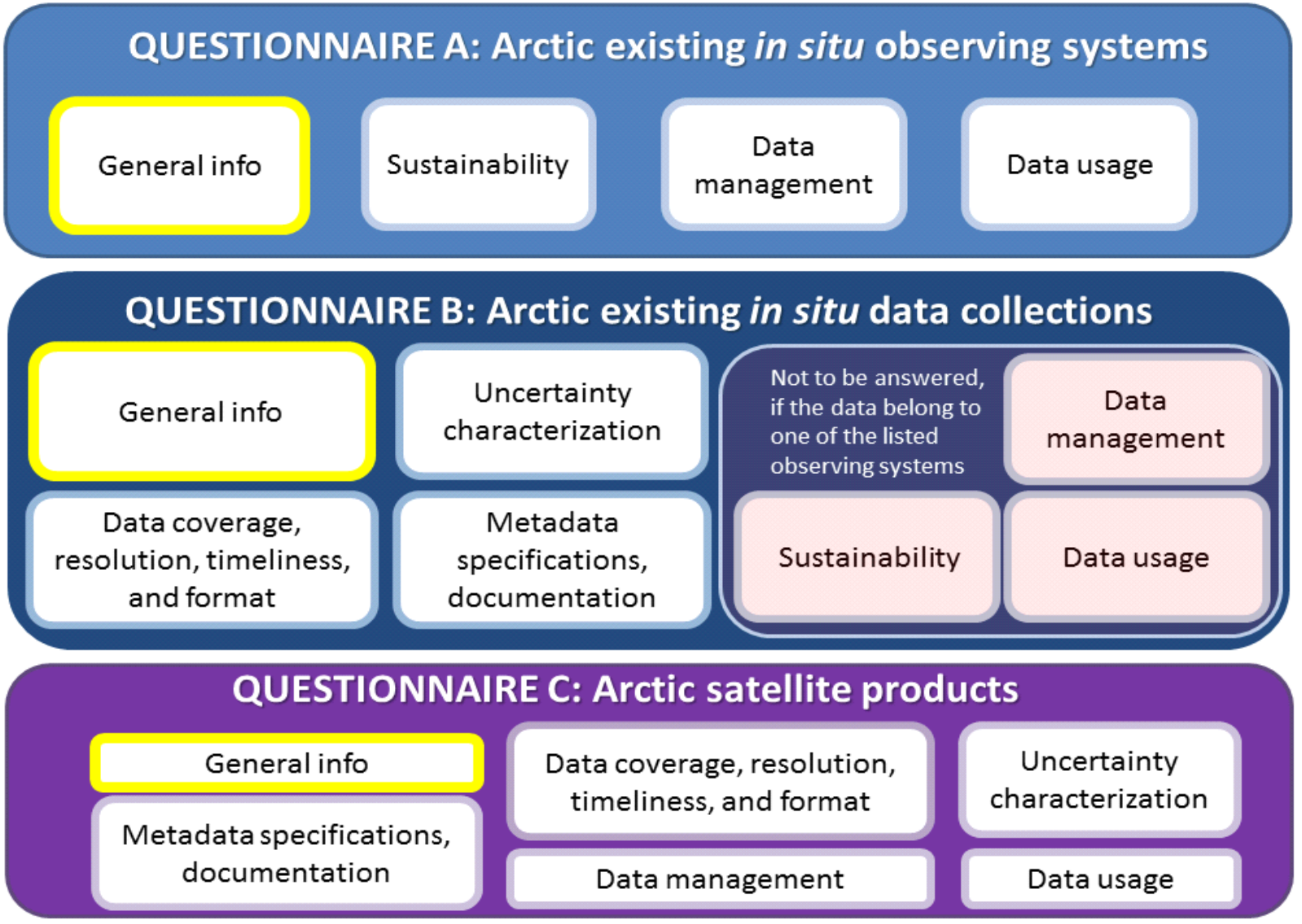
at appropriate temporal and spatial scales and resolution according to user requirements (e.g. climate research, operational service, etc.

Copernicus is a major driver to develop satellite-based observing and modelling of many variables. The largest gaps are in the in-situ observation network, which should provide

- data not obtained from remote sensing
- data needed for validation of remote sensing and numerical models

NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

INTAROS survey of Arctic observing systems



NorDataNet Norwegian Scientific Data Network

Status

- Data discovery
  - Supported by all systems
- Data submission
  - Supported by all systems, mostly manual
  - Work in progress for upload and conformance check interface in NorDataNet
- Data access
  - Simple download working for all
  - Bundling of datasets working for some
  - Standardised encoding and access interfaces starting to gain momentum
- Data transformation/processing
  - Subsetting, reformatting and reprojection of data partly implemented in NorDataNet and NORMAP, to be extended
  - Work flow management planned for NorDataNet based on experience gained in GeoAccessNO
- Too fragmented data encoding
  - Need harmonisation to make usage of data more efficient
  - Simplest solution through generic interface between provider and consumer
  - OPeNDAP, CDM, CF etc
  - Disconnect from underlying format
  - Connect directly into analysis tools like Python, Matlab, R
- Most data centres support or will support DOIs in the near future
  - Enables traceability of data consumption
  - Visibility for scientists and data centres
- NorDataNet is starting a discipline specific Advanced User Support activity with NIRD to implement applications servers offering OPeNDAP and OGC WMS for data stored in the NIRD archive
- NorDataNet is participating in semantic coordination efforts in the polar region
- NorDataNet central services are running on a private cloud

NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

Future (1)

- NorDataNet
  - Increase the number of data centres offering OPeNDAP access to datasets
  - Integration of work flow management evaluated through GeoAccessNO
    - Allows connection to various processing resources (including HPC) and accounting towards NOTUR accounts
    - Which processing service to use would rely on the task
  - Visualisation of datasets on the fly through web services
    - Time series - in progress as an OGC WPS
    - Profiles - planned
  - Linkage of data to publications under consideration
    - Several issues to resolve
    - Need to make other functionality work properly first
  - Primary focus in the short term is consolidation of existing setup and ingestion of data to make it more useful

NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

Future (2)

- INTAROS
  - Extend the mapping of in situ observing systems
    - Questionnaire opened to the public (https://intaros.nersc.no/node/651)
    - Continued in Arctic 2030 project for MD (Miljødirektoratet)
  - Start linking to data repositories
    - Harvest metadata from established repositories (e.g. NMDC)
    - Standard protocols (e.g. OpenSearch, OAI-PHM)
  - Continue the development of processing tools for development of stakeholder services
    - Cloud-based platform, using open standards
  - Extending network to Norwegian organisations and stakeholders
    - Better communication between INTAROS and Norwegian institutions working in the Arctic
- Coordinate with NorDataNet and other Norwegian infrastructures

NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS