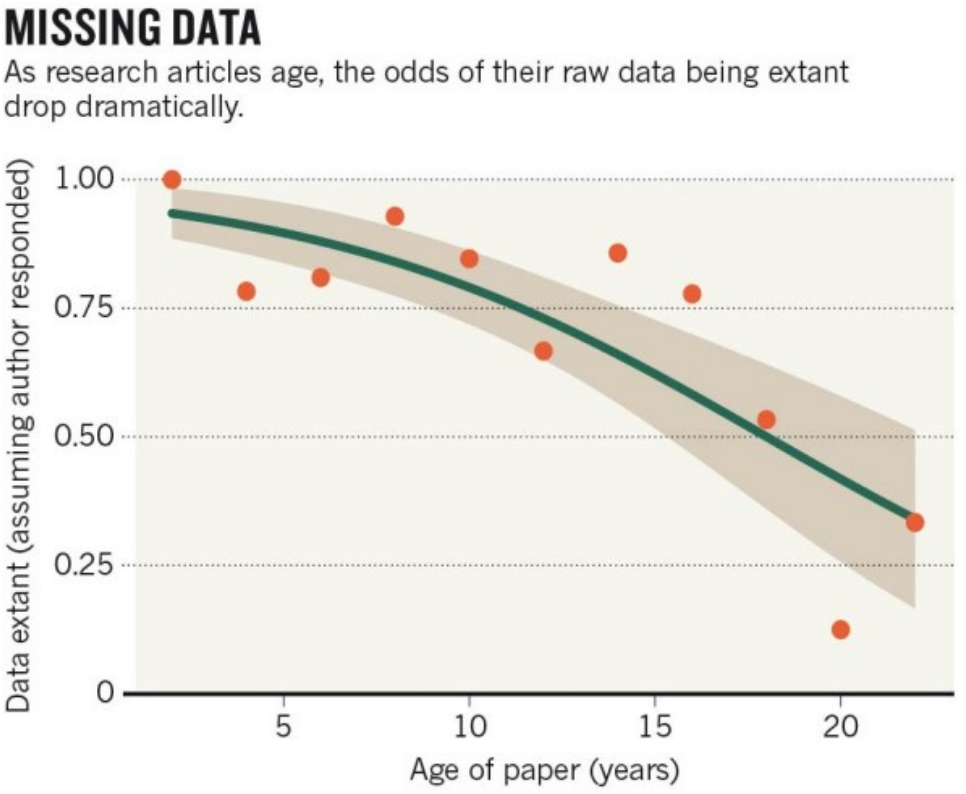
Norwegian and international data infrastructures Services for providers and consumers Focusing on the Arctic

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?

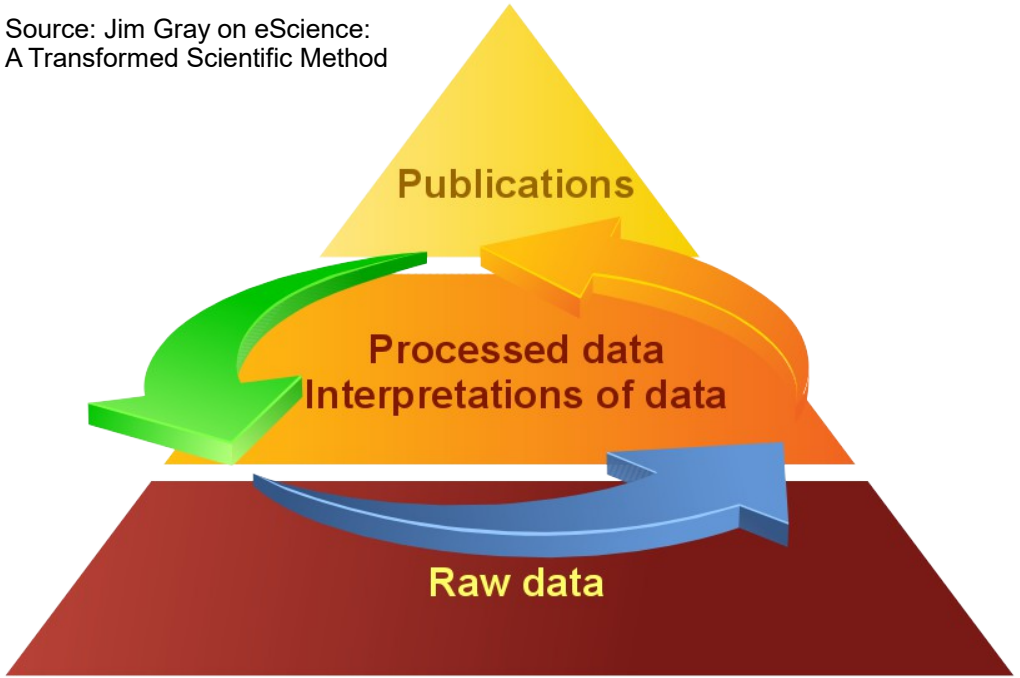
- Maximise public investment in data collection and production
- Promote scientific collaboration
- Promote interdisciplinary science
- Promote scientific transparency
- Leave a legacy

- 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

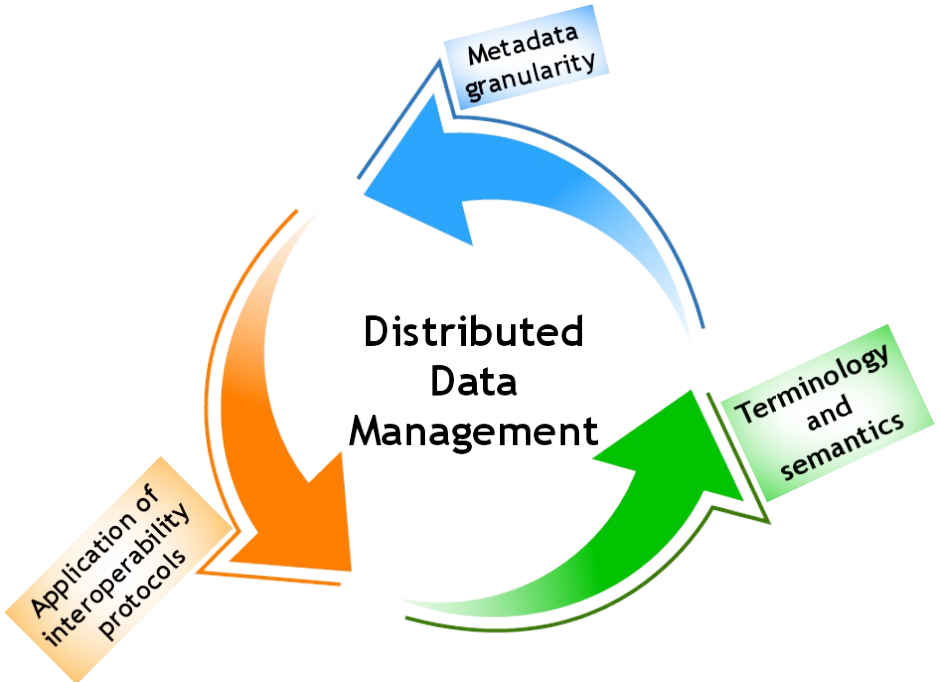
NorDataNet GeoAccessNO NMDC norsk marint datasenter NORMAP

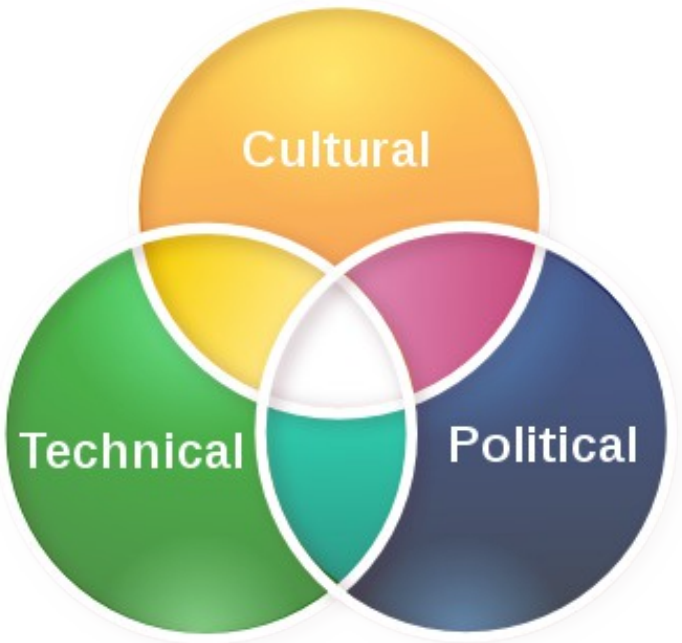
Types of metadata

- 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

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

Challenges





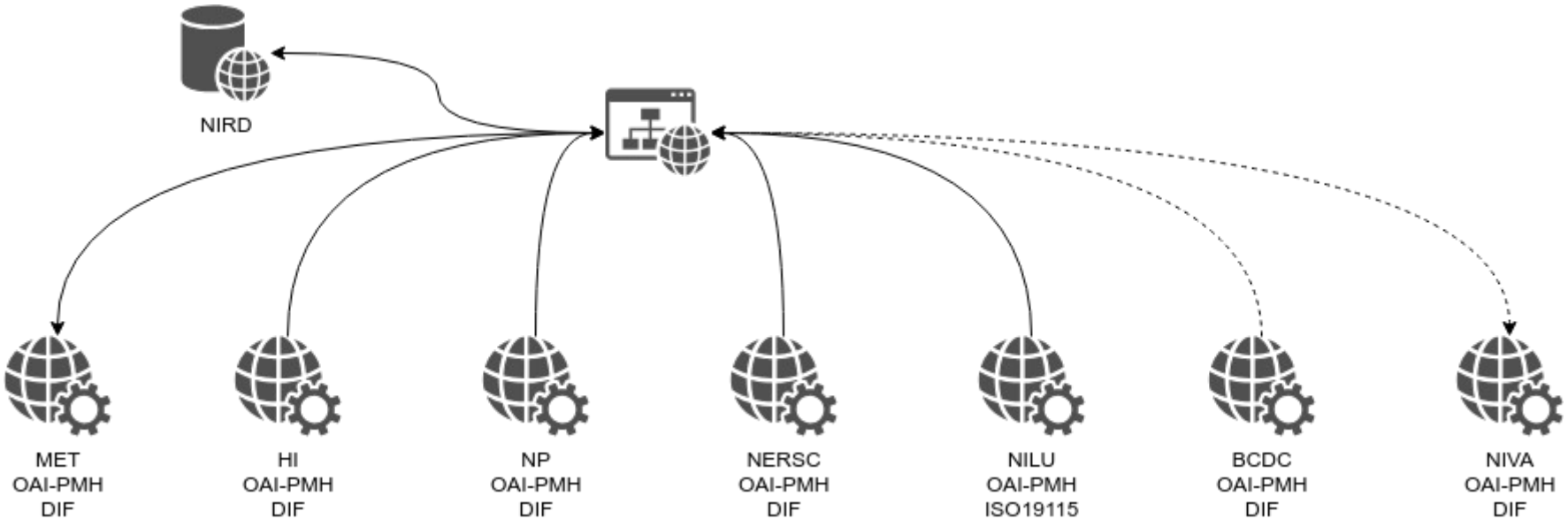
NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

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/

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

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

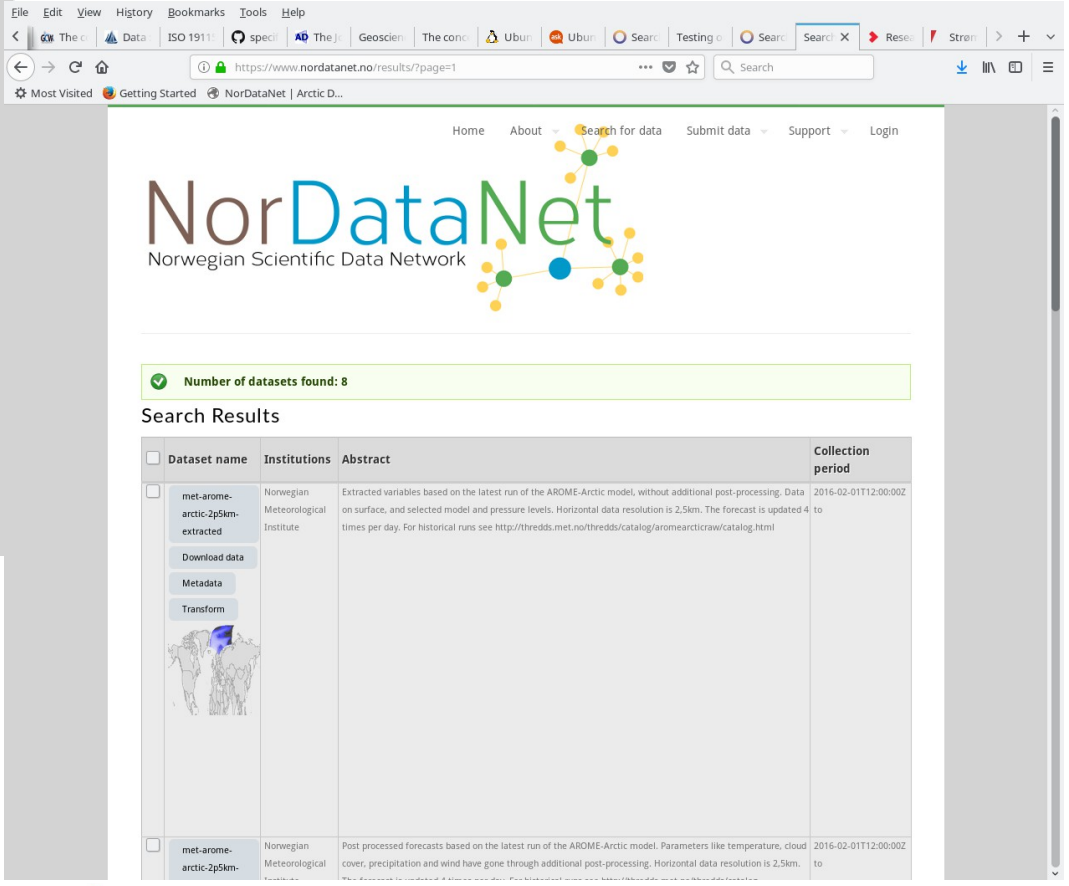
Bounding box

Institutions

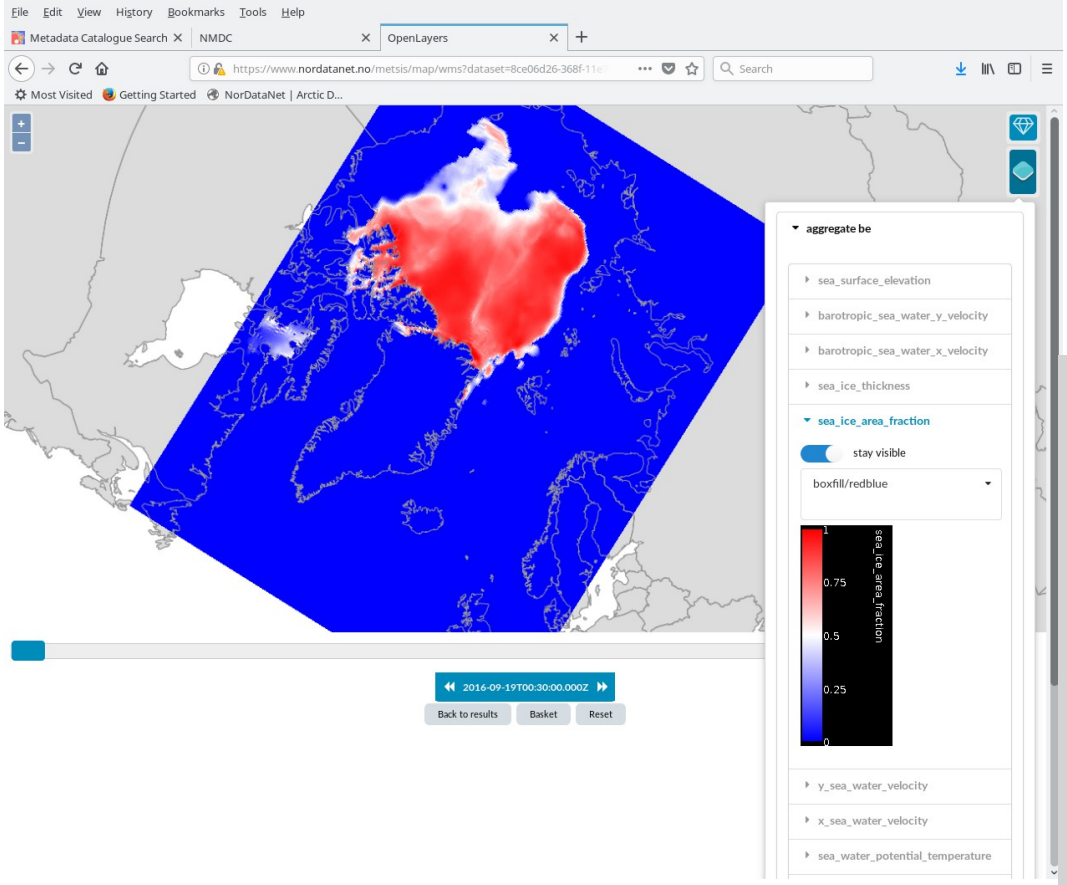
Investigator

Search

Geographical search



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



These operations can be used through the basket on multiple products sharing features

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

Select temporal extent

Start date

Stop date

Select variables

	Name	Standard name	Long name	Units
☐	h	sea_floor_depth_below_sea_level	sea_floor_depth_below_sea_level	meter
☐	latitude	latitude	latitude	degree_north
☐	longitude	longitude	longitude	degree_east
☐	mask		mask on RHO-points	
☐	aice	sea_ice_area_fraction	fraction of cell covered by ice	
☐	hice	sea_ice_thickness	average ice thickness in cell	meter
☐	salinity	sea_water_salinity	salinity	1e-3
☐	temperature	sea_water_potential_temperature	Sea water potential temperature	Celsius
☐	u	x_sea_water_velocity	Sea water x velocity	meter second-1
☐	ubar	barotropic_sea_water_x_velocity	Barotropic sea water x velocity	meter second-1
☐	v	y_sea_water_velocity	Sea water y velocity	meter second-1
☐	vbar	barotropic_sea_water_y_velocity	Barotropic sea water y velocity	meter second-1
☐	zeta	sea_surface_elevation	Sea surface height above geoid	meter

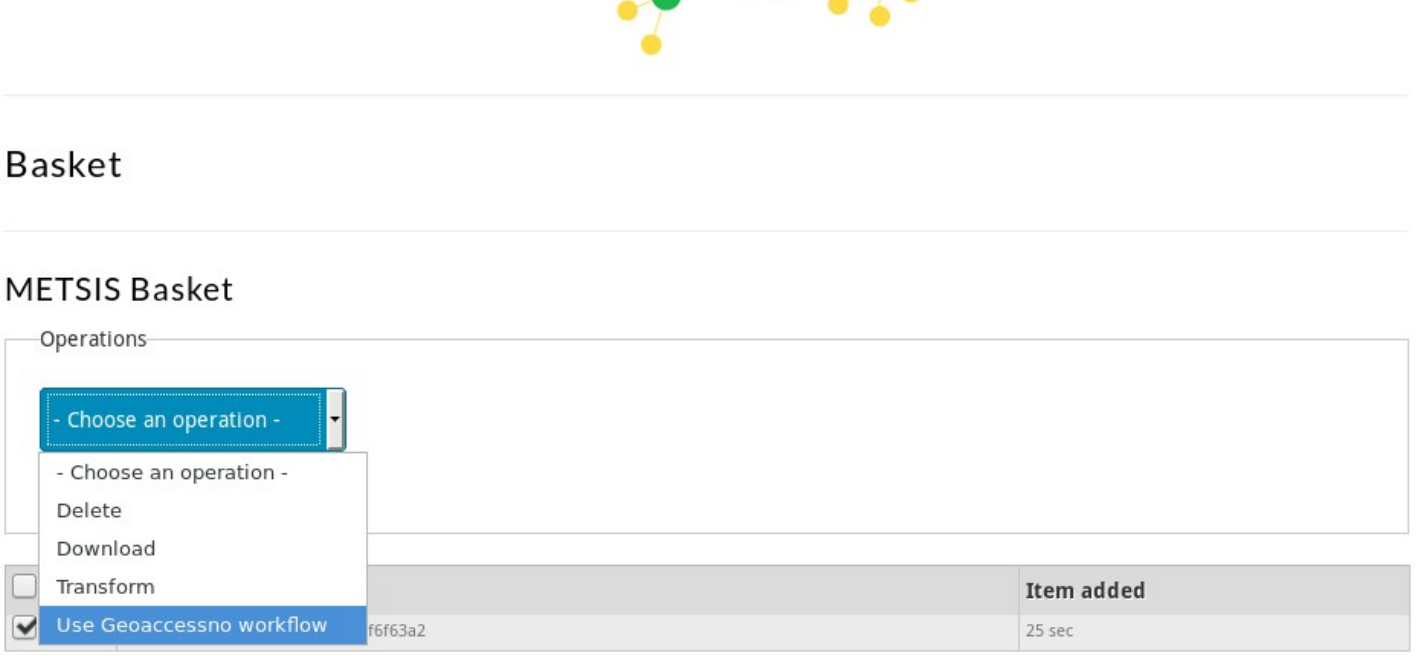
Select map projection

x-axis from:

Minimum value of x-coordinate

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

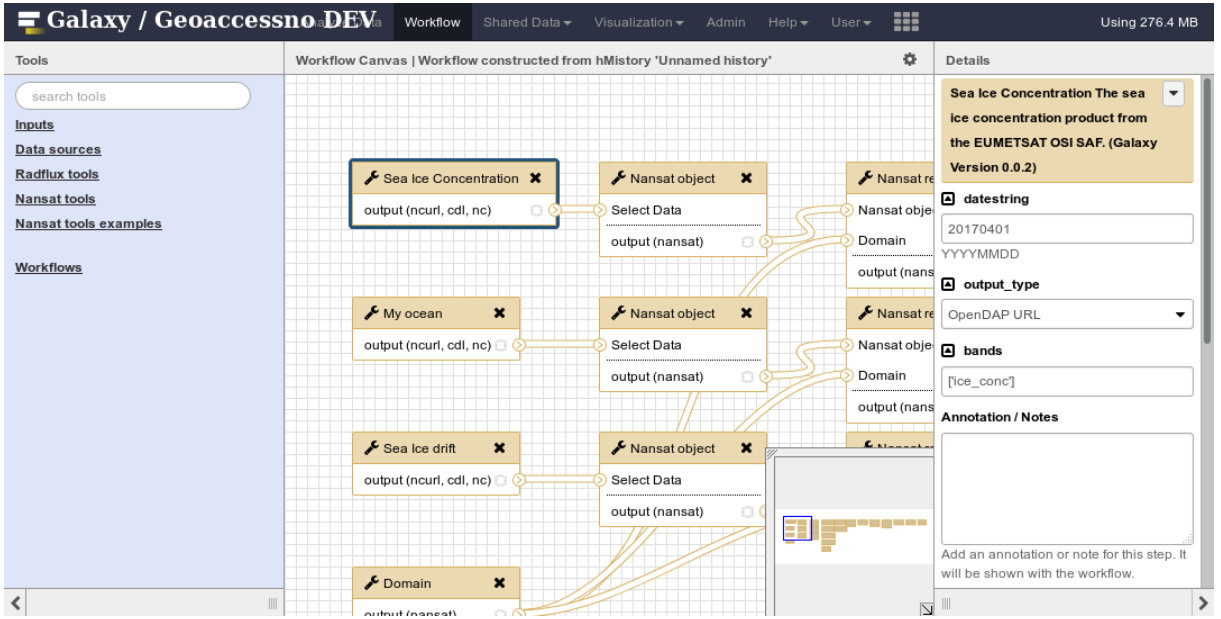
Norwegian Scientific Data Network



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

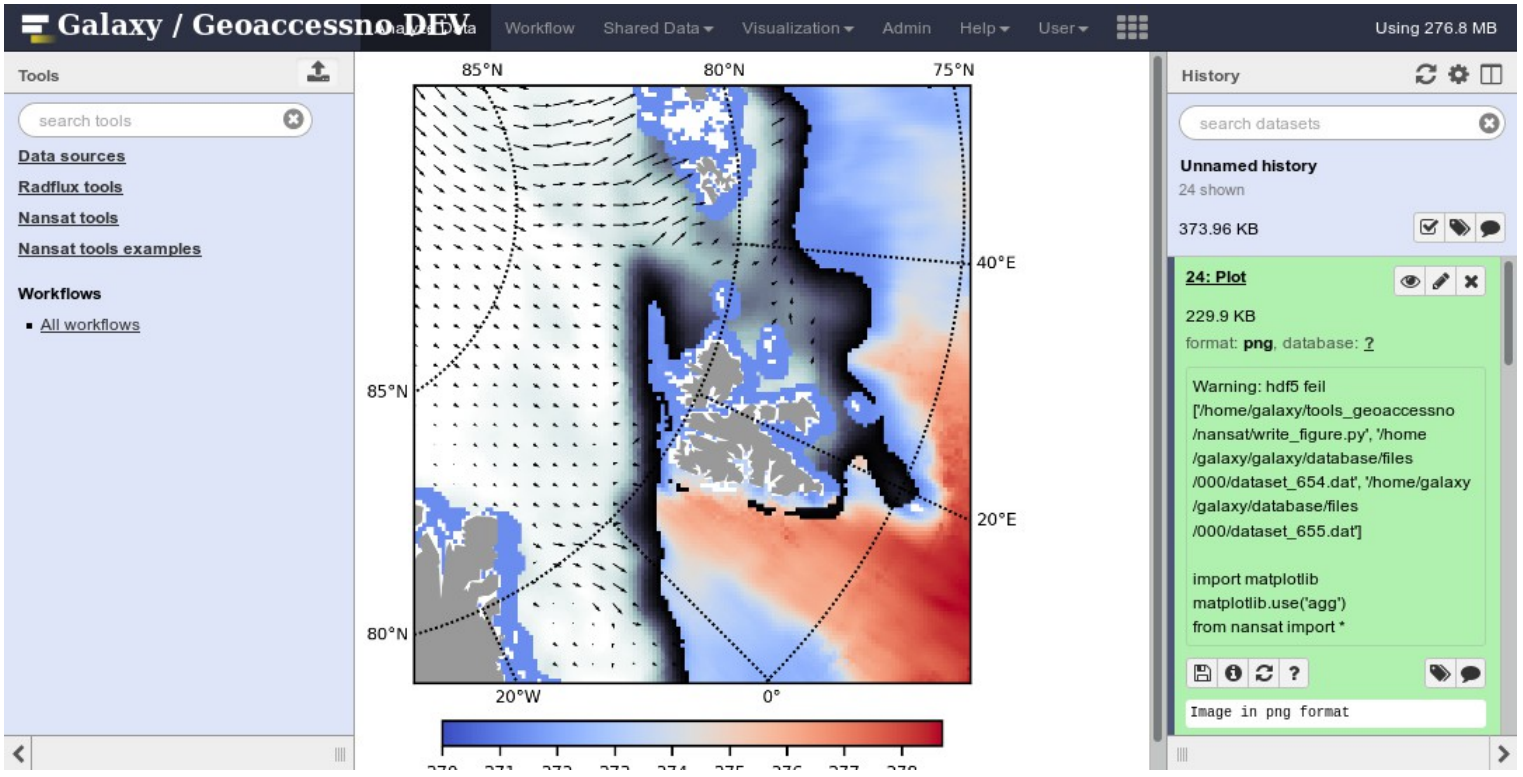
NorDataNet Norwegian Scientific Data Network GeoAccessNO NMDC norsk marint datasenter NORMAP INTAROS

Virtual research environment

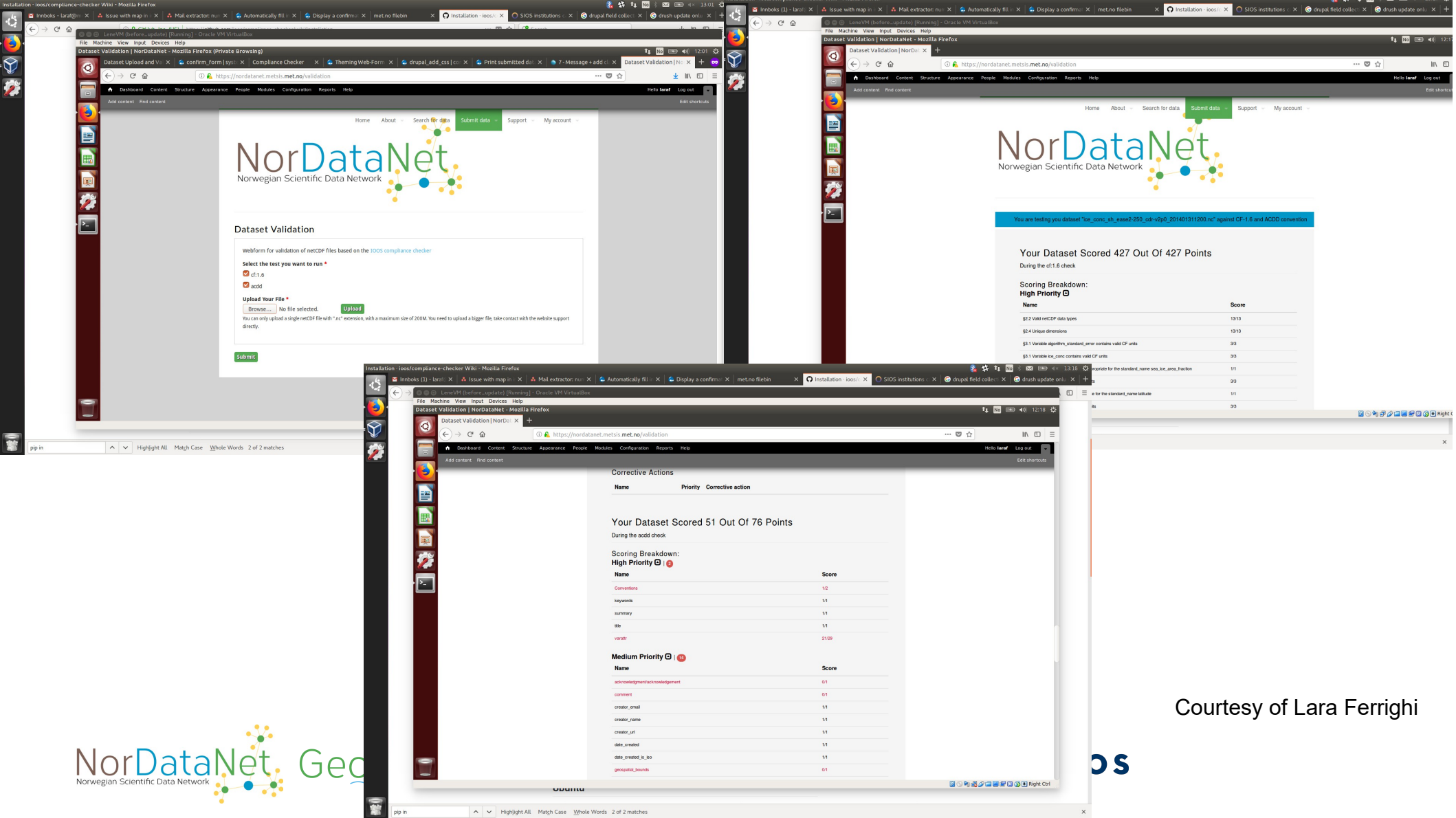


- Based on the Galaxy framework
  - Visualise and edit the workflow graphically
  - Workflow can be shared with other users
- Integrates data and software developed by the scientific community
- Creates workflow from

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



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



Courtesy of Lara Ferrighi

NorDataNet Norwegian Scientific Data Network

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/
  - Strong linkage between NorDataNet and SIOS

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

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

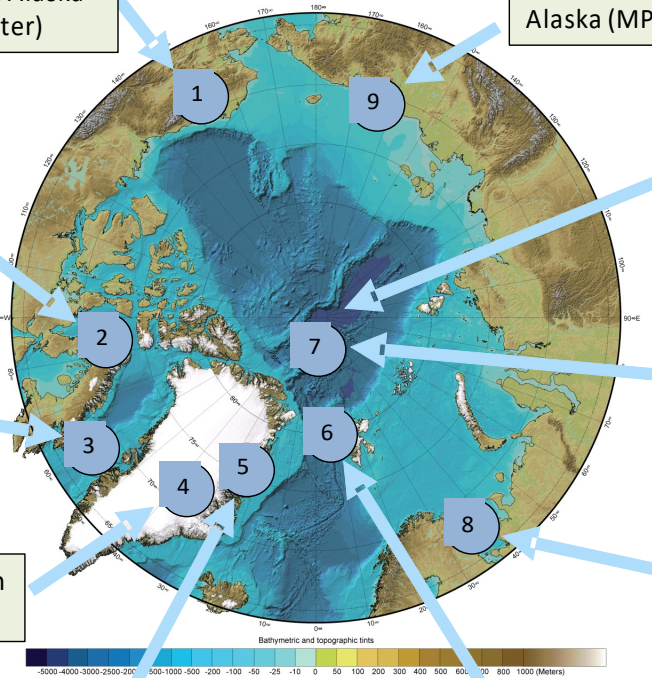
at appropriate temporal and spatial scales and resolution according to user requirements (e.g. climate research, operational service, etc.

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 field activities in 2017-2018



Temperature profiles of the soil to observe thaw depth in Alaska sites (USFD and Univ. Exeter)

Observations of atmospheric C02 and CH4 in Siberia and Alaska (MPG)

Observing meteorological, snow and soil data from 4 location in Eastern Canadian Arctic by CNRS Takuvik

Oceanographic, sea ice and snow measurements in central Arctic from automated ice buoys (ITPs and IMBs) by FMI and IOPAN

Oceanographic and marine ecosystem data from bio-Argo floats in Baffin Bay by CNRS Takuvik

Atmospheric profiles and surface measurements in the central Arctic from icebreaker ODEN (Stockholm University)

PROMICE weather station data from Greenland Ice Sheet by GEUS

Soil temperature and snow measurements from stations in northern Finland (FMI)

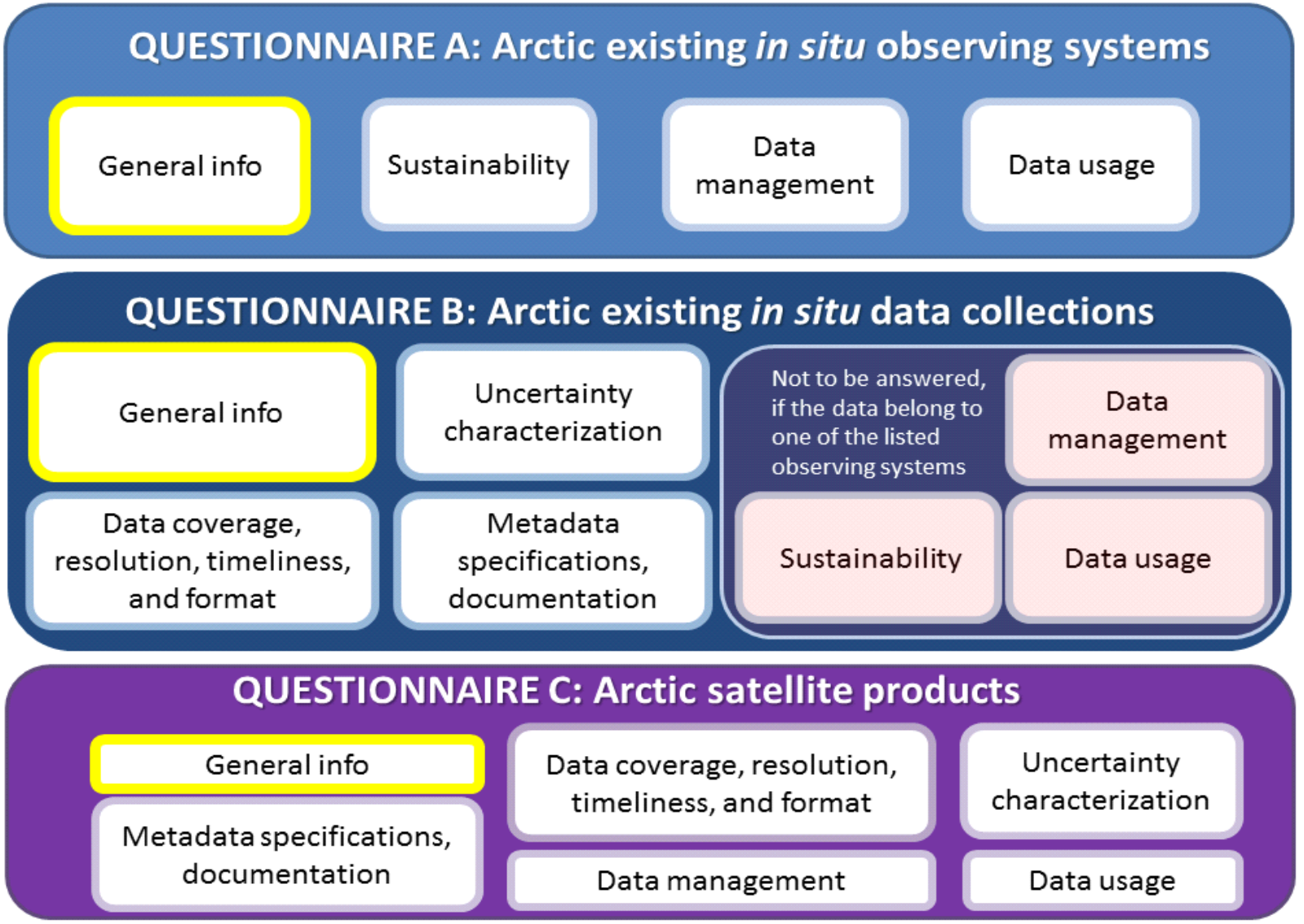
Oceanographical CTD sections In Young Sound fjord, as part of the Greenland Ecosystem Monitoring program (Aarhus University)

Satellite and aircraft observations of the Greenland Ice Sheet (GEUS, DTU, UPM)

- Biogeochemical observations in the Fram Strait (Hausgarten) and Kongsfjorden (AWI, CNRS IUEM)
- Oceanographic moorings north of Svalbard incl. biogeochemistry (IOPAN, UiB, IMR, NERSC, CNRS)
- Glider experiment by CNRS LOCEAN

NorDataNet
Norwegian Scientific Data Network

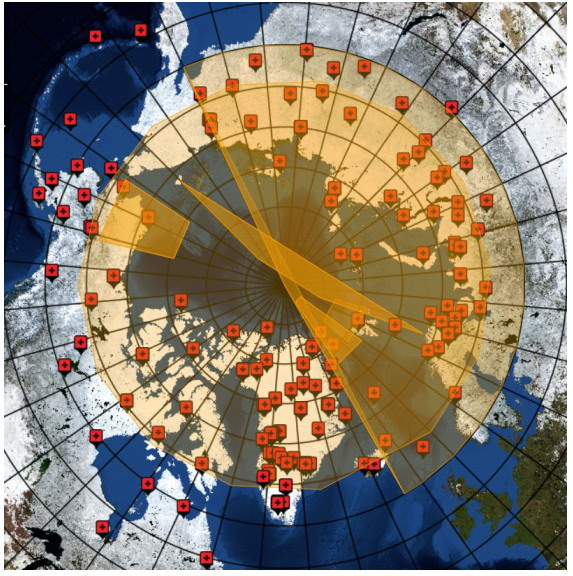
INTAROS survey of Arctic observing systems

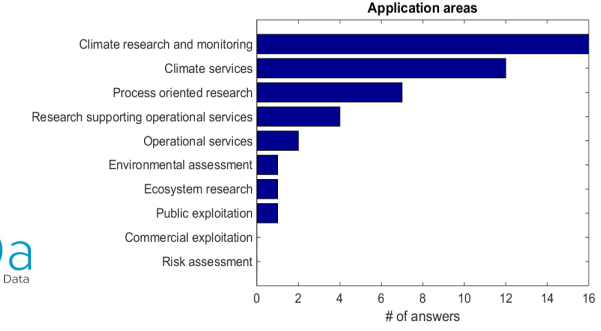


NorDataNet Norwegian Scientific Data Network

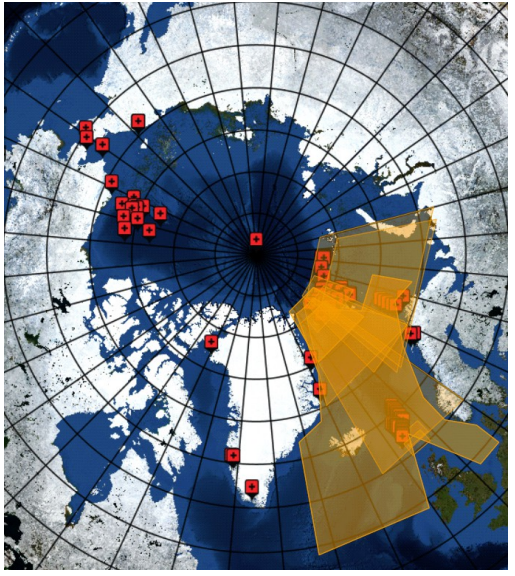
Overview of surveyed data

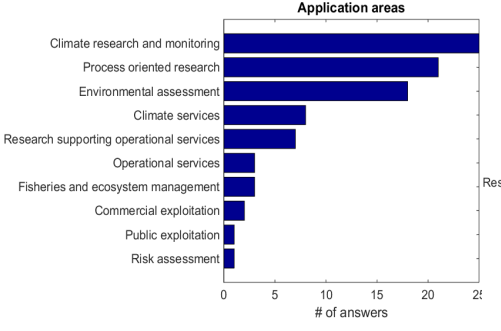
Atmosphere



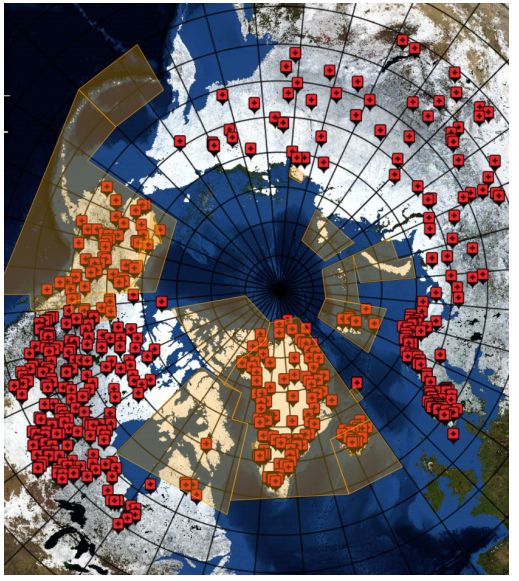


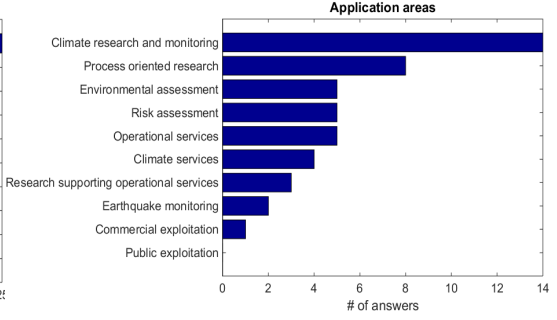
Ocean and sea ice





Land and terrestrial cryosphere





NorDa
Norwegian Scientific Data

Results: Sustainability and data management

Ocean and sea ice (25)

CONCLUDING REMARKS:

- It is a major problem that in-situ observing systems **lack sustainability**.
- We recommend development of multi-disciplinary observatories using **well proven and robust instrumentation mounted in sea floor installations**, bottom anchored oceanographic moorings, and drifting ice-tethered platforms.
- Need to **develop and adapt technologies and sensors** to make biogeochemical and biological observations feasible.
- There are many gaps in the data coverage in the Arctic, but the **gaps in biogeochemical observations** are particularly important.
- In the Arctic there are **limiting factors in accessing data** in the same way as in other regions.

Observation System	Platform	Sustainability			Data Management						Data repository
R/V Håkon Mosby	Vessels	5	5	N/A	3	4	2	3	2	3	NMDC
SAVN (Faerose National History Museum)	Community Based	missing			missing						
SIOS Airborne Infrastructure	Airborn Sensors	3	4	N/A	2	2	2	2	2	3	
UNIS ocean observing System	Fixed Moorings	4	4	N/A	2	2	2	2	2	3	

*(for terrestrial stations only)

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
- 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 implemtn applications servers offering OPeNDAP and OGC WMS for data stored in the NIRD archive

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