

A large, blue, low-poly iceberg floats in a dark blue ocean. The top of the iceberg is visible above the water line, while the much larger, jagged base is submerged. The background is a light blue sky.

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

Øystein Godøy, Stein Sandven,
Markus Fiebig and Trond Thorbjørnsen

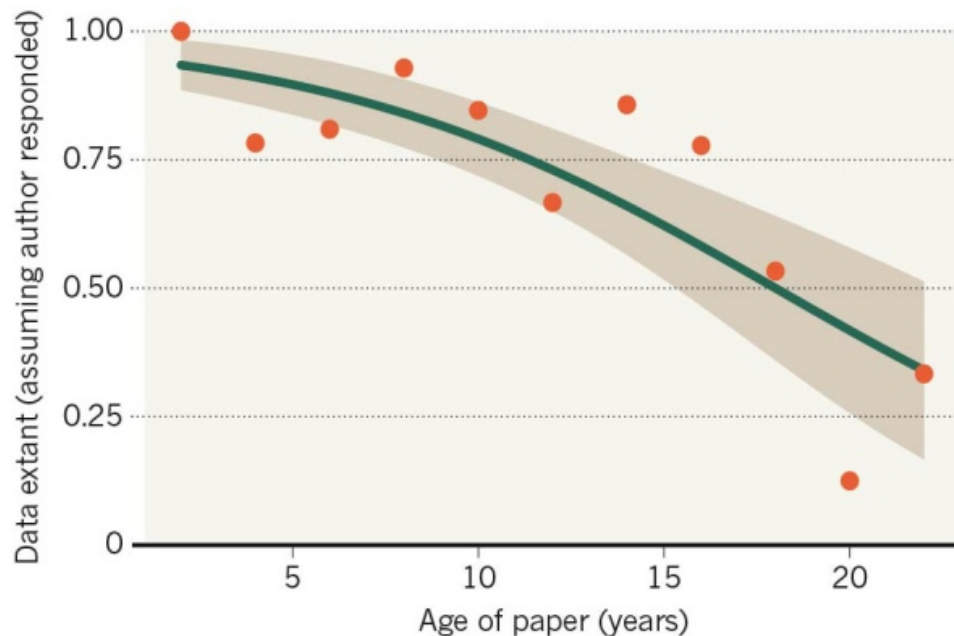
Loosing scientific data

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

–Gibney and Van Noorden (2013), Nature

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



Why bother with structured data management?

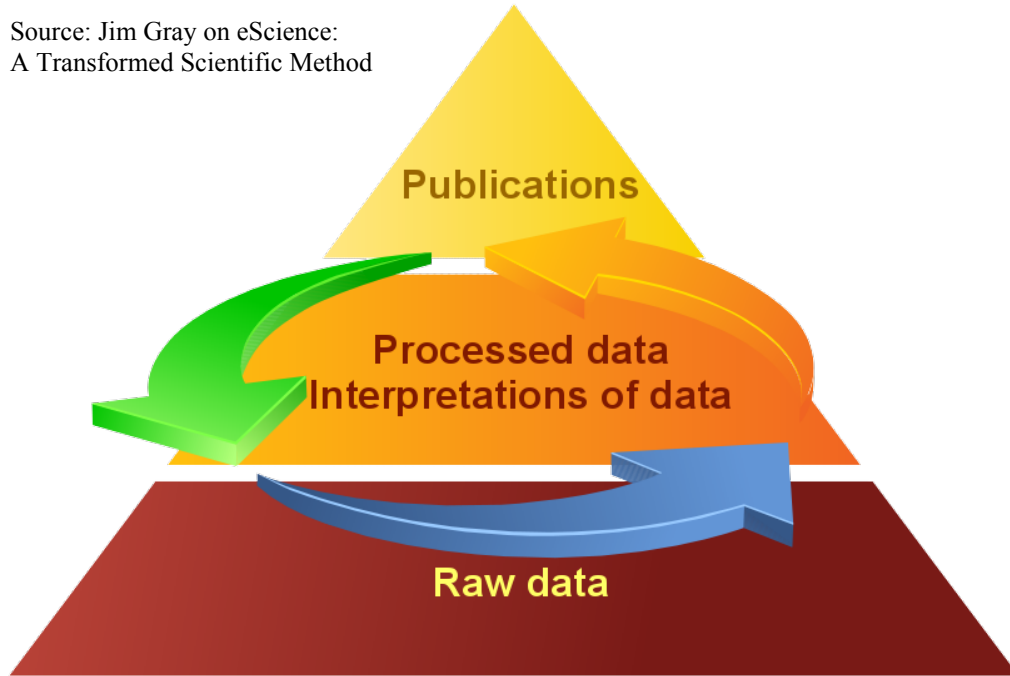
- Maximise public investment in data collection and production
- Promote scientific collaboration
- Promote interdisciplinary science
- Promote scientific transparency
- Leave a legacy



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

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

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

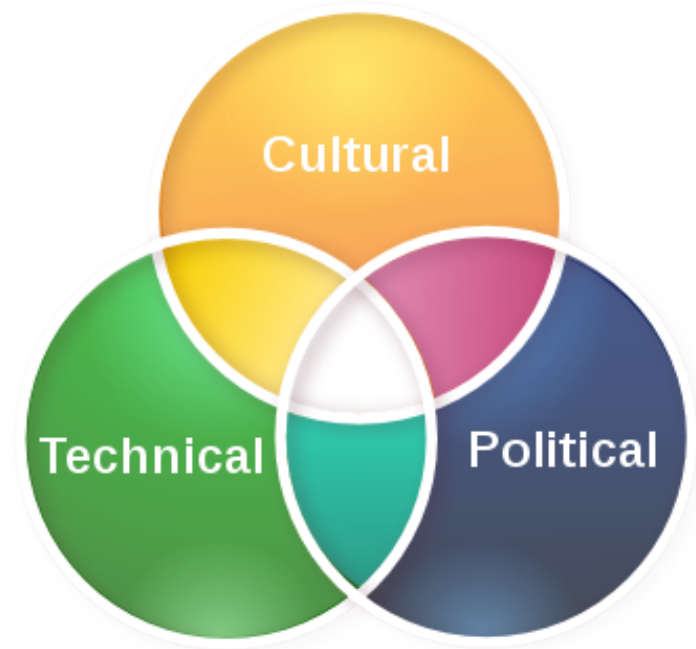
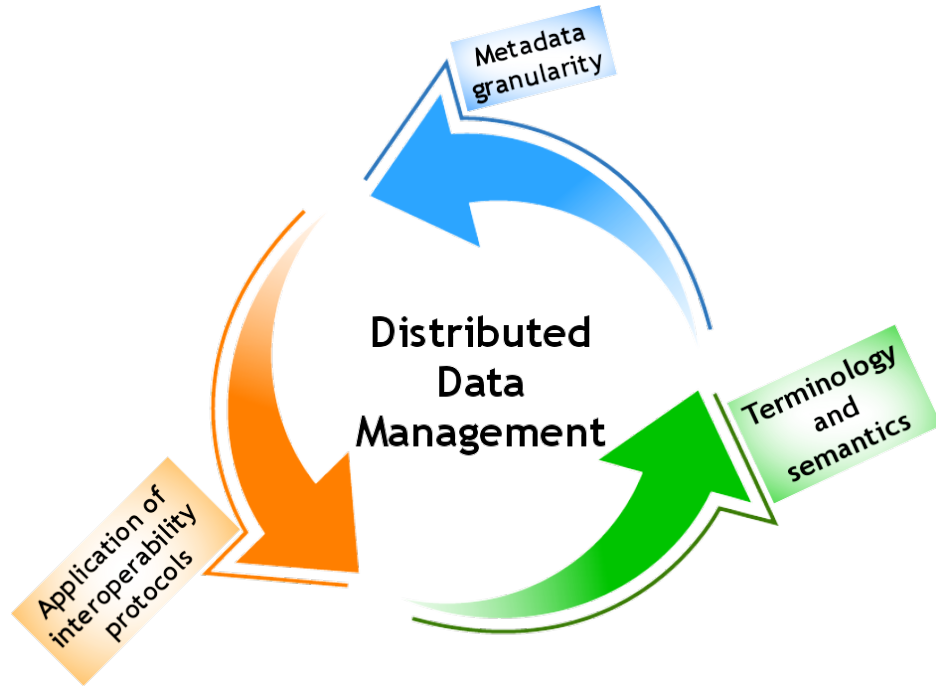
- who* measured, simulated or analysed *what*, *where*, and *when* as well as *conditions for reuse* and *access mechanisms* for the data
- to enable users to find appropriate data for the task

•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



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://nordatanet.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/>

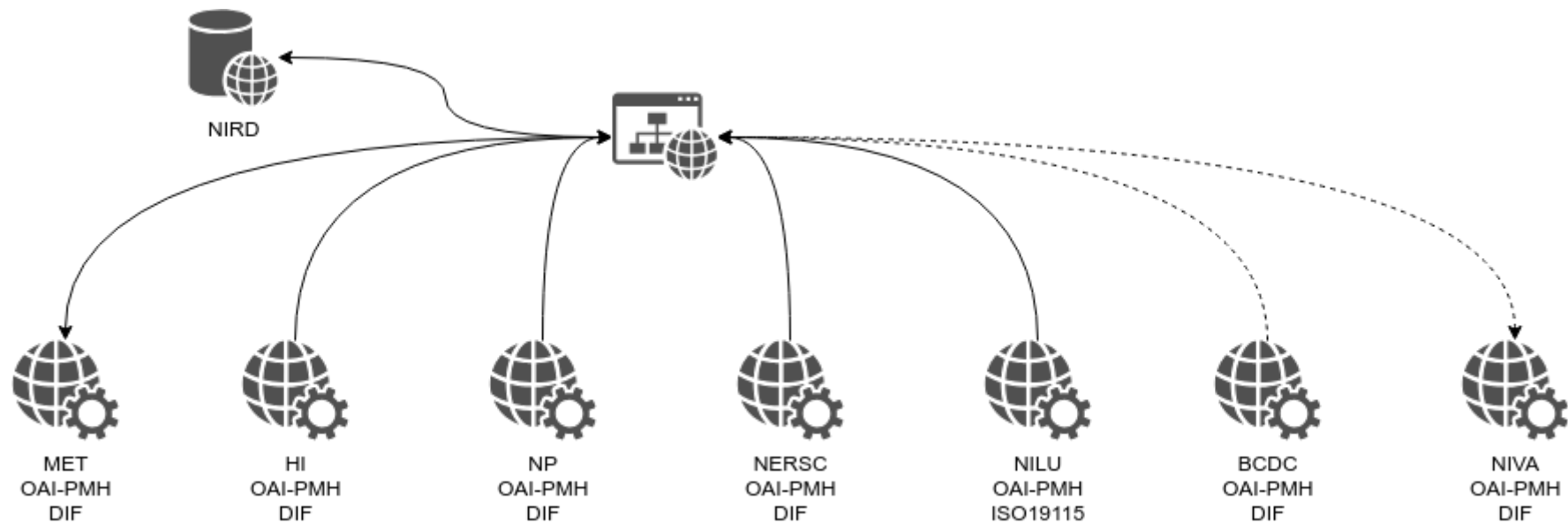
•GeoAccessNO

–Pilot project examining combination of physical and digital data, software and literature

–Not a service

–<https://www.geoaccessno.no/>

National nodes



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

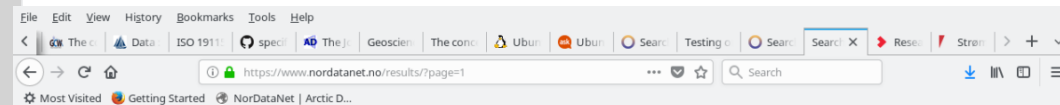
Bounding box

Institutions

Investigator

Search

Geographical search



Home About Search for data Submit data Support Login


NorDataNet

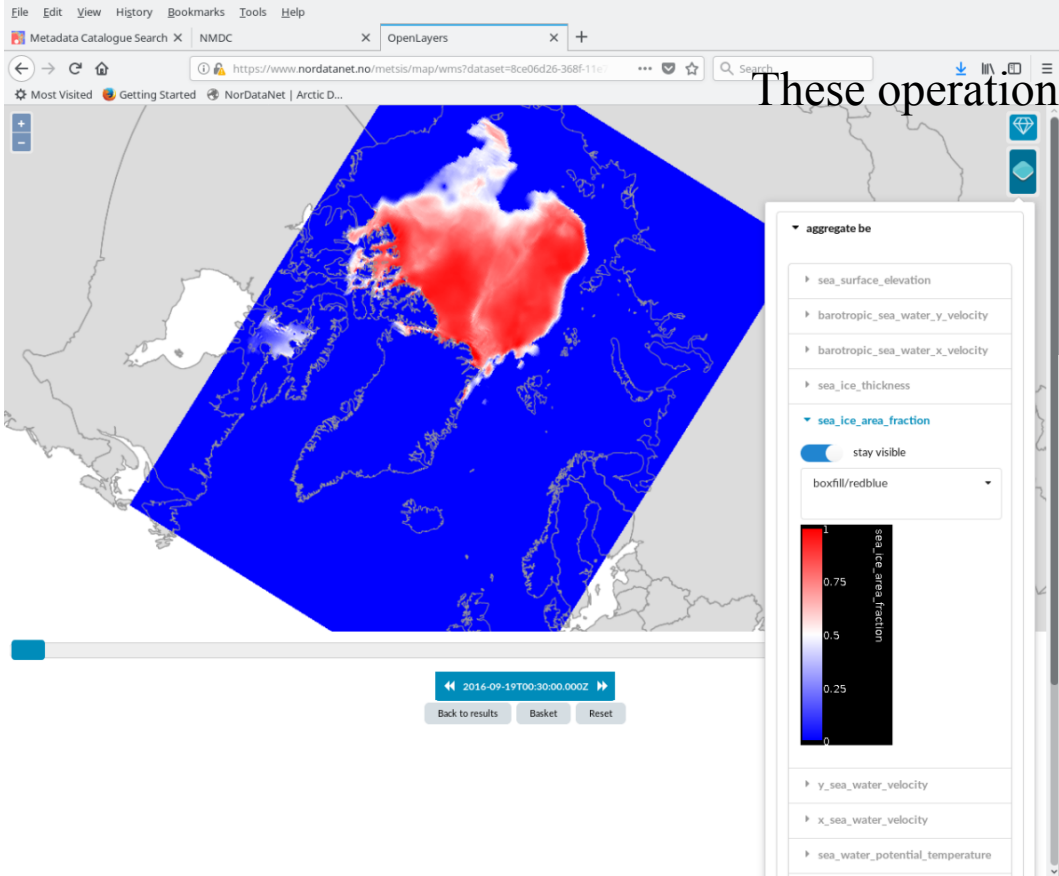
Norwegian Scientific Data Network



Number of datasets found: 8

Search Results

<input type="checkbox"/>	Dataset name	Institutions	Abstract	Collection period
<input type="checkbox"/>	met-arome-arctic-2p5km-extracted	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
			Download data	
			Metadata	
			Transform	
				
<input type="checkbox"/>	met-arome-arctic-2p5km-	Norwegian Meteorological Institute	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. 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



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

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

• Select temporal extent

Start date

Stop date

• Select variables

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

• Select map projection

x-axis from:

Basket

METSIS Basket

Operations

- Choose an operation -

- Choose an operation -

Delete

Download

☐ Transform

☒ Use Geoaccessno workflow

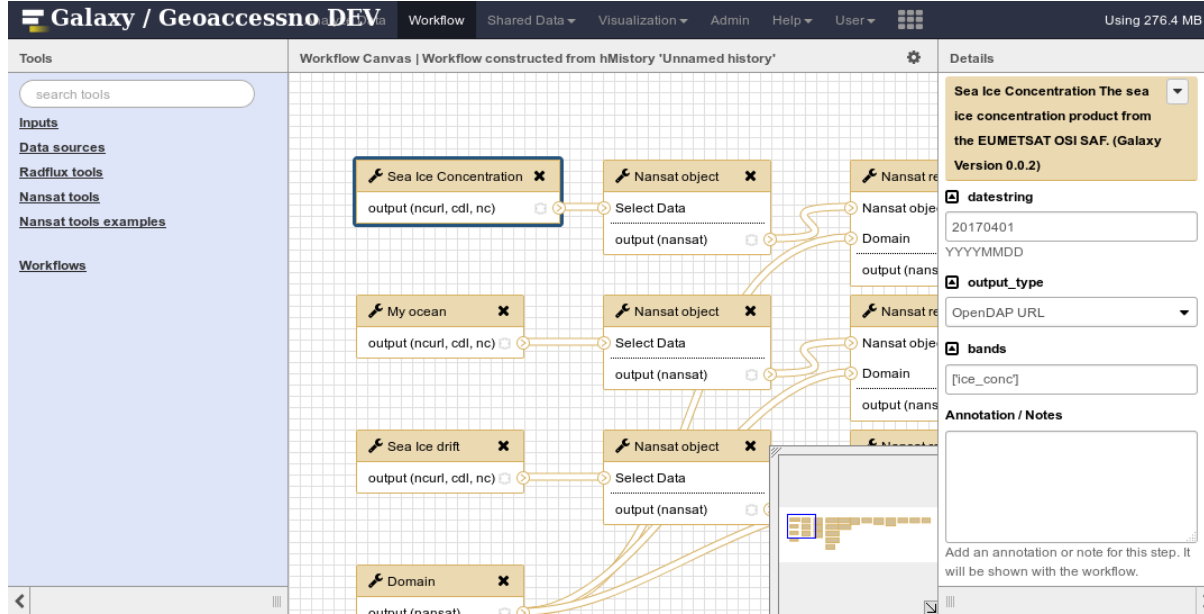
		Item added
	f6f63a2	25 sec

•API access

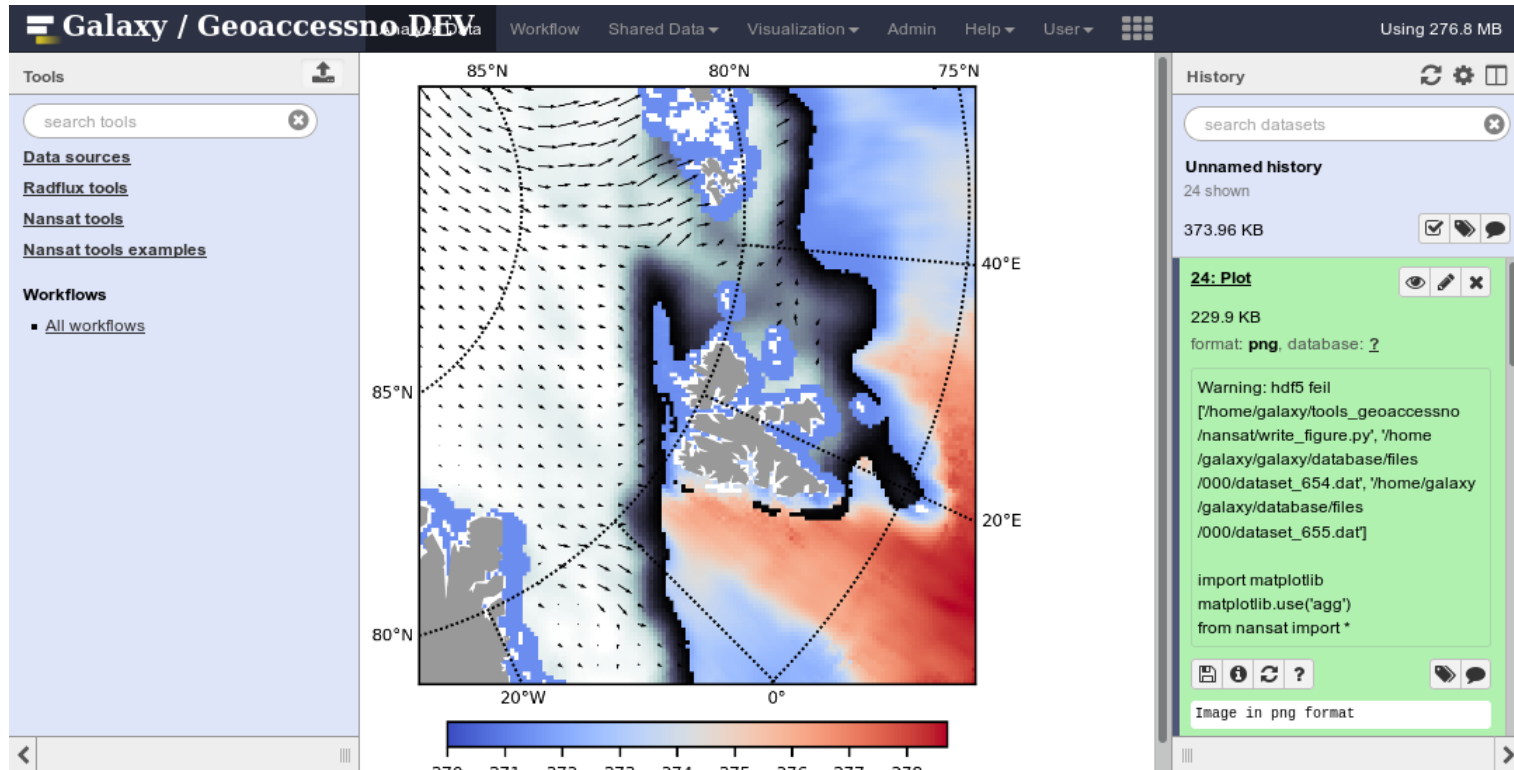
•User can run tools
and workflows
externally

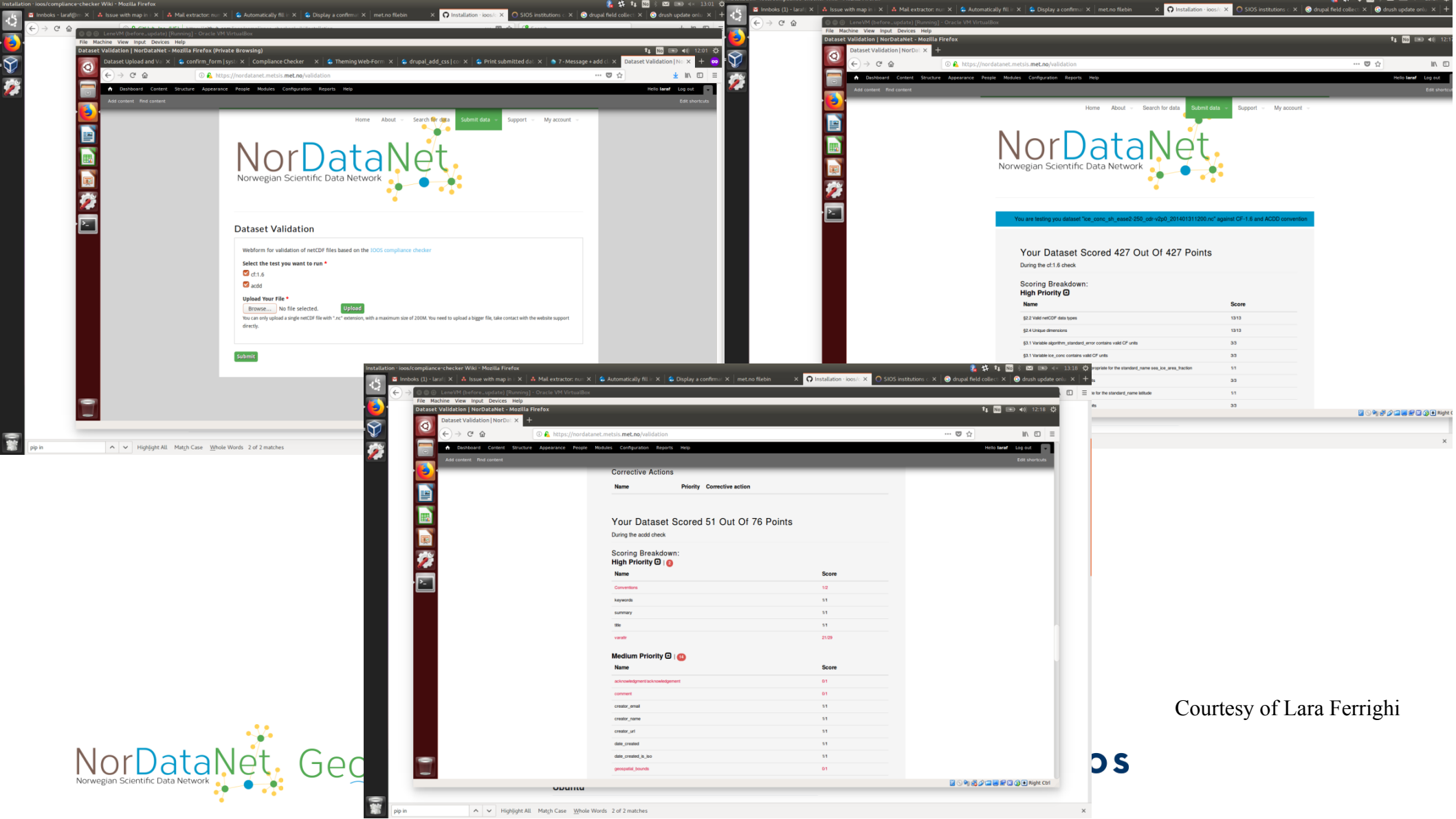
•Enables integration
in web services

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





Dataset Validation

Webform for validation of netCDF files based on the [ISO95 compliance checker](#)

Select the test you want to run

- ☒ cf-1.6
- ☒ addc

Upload Your File

No file selected.

You can only upload a single netCDF file with ".nc" extension, with a maximum size of 200M. You need to upload a bigger file, take contact with the website support directly.



You are testing your dataset "ice_core_sh_ess02-290_v0-p0_201401311200.nc" against CF-1.6 and ACDD convention

Your Dataset Scored 427 Out Of 427 Points

During the cf-1.6 check

Scoring Breakdown: High Priority

Name	Score
\$2.2 Valid netCDF data types	13/13
\$2.4 Unique dimensions	13/13
\$3.1 Variable algorithm, standard_error contains valid CF units	3/3
\$3.1 Variable ice_core contains valid CF units	3/3
prepare for the standard_name var, ice_core, fraction	1/1
is for the standard_name latitude	1/1
	3/3

Corrective Actions

Name Priority Corrective action

Your Dataset Scored 51 Out Of 76 Points

During the addc check

Scoring Breakdown: High Priority

Name	Score
Conventions	1/2
keywords	1/1
summary	1/1
title	1/1
variable	21/29

Medium Priority

Name	Score
acknowledgment/acknowledgment	0/1
comment	0/1
creator_email	1/1
creator_name	1/1
creator_url	1/1
date_created	1/1
date_created_iso	1/1
geospatial_bounds	0/1



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



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



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



INTAROS

INTAROS

survey of
Arctic
observing
systems

QUESTIONNAIRE A: Arctic existing *in situ* observing systems

General info

Sustainability

Data
management

Data usage

QUESTIONNAIRE B: Arctic existing *in situ* data collections

General info

Uncertainty
characterization

Not to be answered,
if the data belong to
one of the listed
observing systems

Data
management

Data coverage,
resolution, timeliness,
and format

Metadata
specifications,
documentation

Sustainability

Data usage

QUESTIONNAIRE C: Arctic satellite products

General info

Data coverage, resolution,
timeliness, and format

Uncertainty
characterization

Metadata specifications,
documentation

Data management

Data usage

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

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



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



INTAROS