Data sharing ethics & culture, and how NorDataNet services help

Øystein Godøy



Outline

- Data sharing ethics, certainly before publishing
- Data Life Cycle and its relation to the scientific workflow, revisited from a scientists point of view
- Data sharing in a cultural perspective and relations to the scientific workflow
- NorDataNet service overview



Data sharing ethics prior to publishing

- Identify the PI(s) of a dataset of interest
- Present your intentions to the PI(s) of a dataset of interest
- Make an agreement on how to use the data
- Acknowledge and respect agreements made
 - Shared data must only be used for the agreed purpose.
 - Use in other form than agreed upon has to be negotiated with the data PI(s).
 - If a publication or presentation is planned, clarify the mutual expectations on (co-)authorship.
- Never share unpublished data
- Do not present or publish without consent
- Publish according to international standard
 - ICMJ "Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals".
- Acknowledge the project

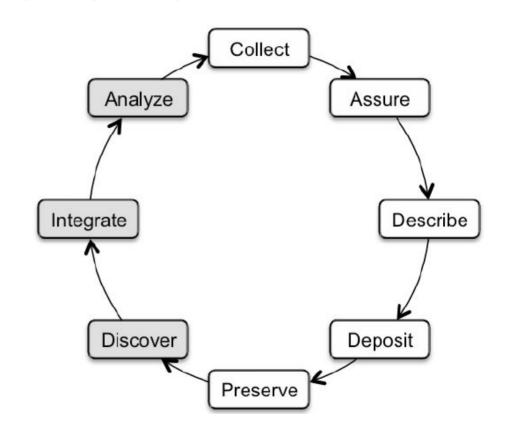


Always cite data used!!



Data Life Cycle and its relation to the researchers workflow

- Generate or collect data
- Quality control and document data
- Analyse data
- Prepare and publish data through a mandated archive
 - Remember to get a DOI!!
- Write a data paper
 - Cite your datasets using the provided DOI
- Write your scientific paper
 - Cite your data paper describing the context of the data





Data sharing in a cultural perspective and relations to the researchers workflow

- Within some communities data sharing is the normal situation
 - In particular within remote sensing communities
- Other communities are more restrictive and consider data a personal item
 - In particular those "suffering" to get data
- Ownership to data are regulated through research grants contracts
 - Usually between the funding agency and the institution of the researcher
- While researchers often only use their own data, the intention of stakeholders (ministries and funding agencies) are to maximise the usage of data
 - Increased value for money
- Research programmes are increasingly focused on interdisciplinary science and societal benefit (impact) of science



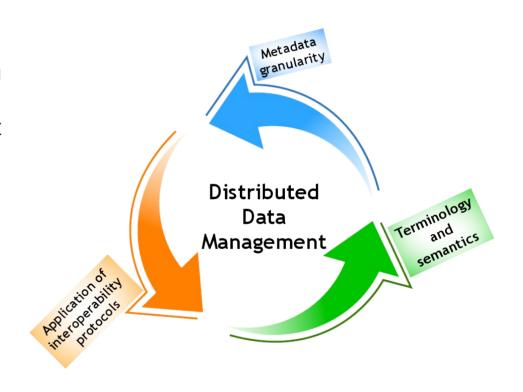
Good data sharing practises

- Reduces the amount of duplicate data records
- Reduces the amount of duplicate field activities collecting data
- Reduces the environmental footprint of science
- Maintains the visibility of institutions and individuals
- Supports the intention of ministries and funding agencies



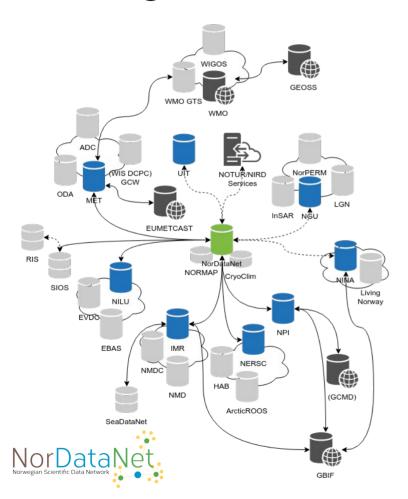
Bottlenecks

- Individual thresholds for sharing data are too high
 - Training, toolboxes and web services reducing the threshold is required
- Education lacks focus on data management
 - Equally important as measurement practise
 - Need to get this into the education
- Data managers and scientists has to work together, identifying gaps and closing them
 - Communication and willingness are essential
- Roles of scientists and data centres are changing, but mutually depending of each other





Norwegian Scientific Data Network - NorDataNet



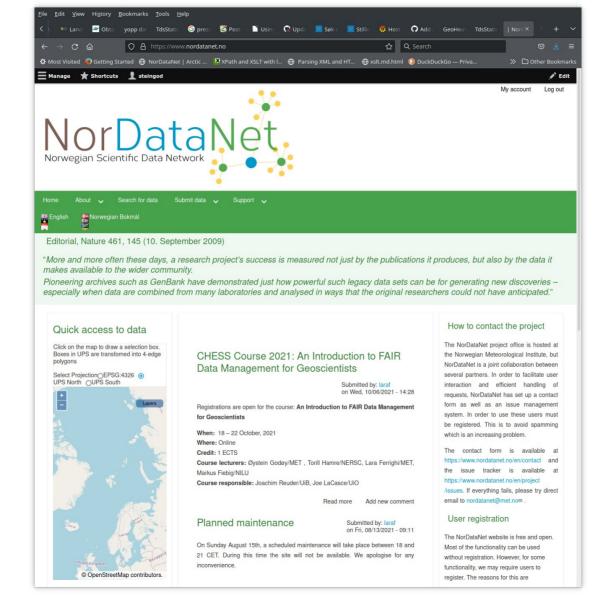
- Distributed data management network
- Development funded by the Research Council of Norway
- Activities are supporting other efforts like SIOS, GCW, NBS, YOPP, ...
- Building on existing data centres nationally and the legacy of IPY
 - Originally geoscientific data
 - But requested by RCN to be interdisciplinary
- Adding efforts like NORMAP and CryoClim as collections
 - Preparing integrations with NBS
- Discovery metadata are harvested into a unified catalogue
 - OAI-PMH, OGC CSW, (OpenSearch)
- Moving towards standardised data hosted by core partners
 - Actionable data as basis for user oriented services
 - Externally harvested information may be of any kind
- Relying on technologies like OPeNDAP, OGC WMS and OGC WPS

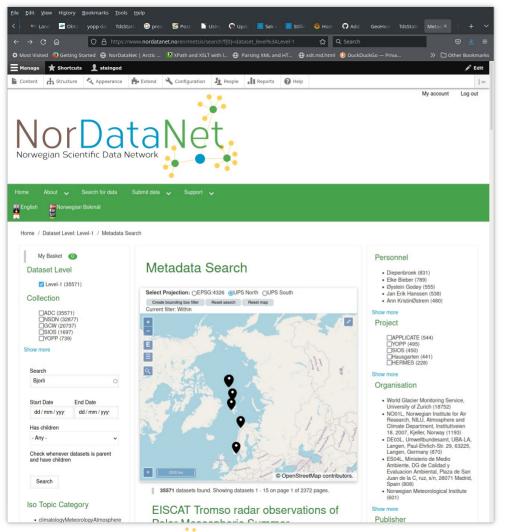
NorDataNet User Services

- Need to identify which services to offer to users (data providers and data consumers)
 - Simplifying their work
 - Showing the benefit of structured data management
 - Slow uptake in user communities
 - Need to interact better with user communities
 - Exploring interactions between research and operations
- Primary focus on Findable, Accessible, Interoperable, Reuseable
 - Data for services
- Modular approach

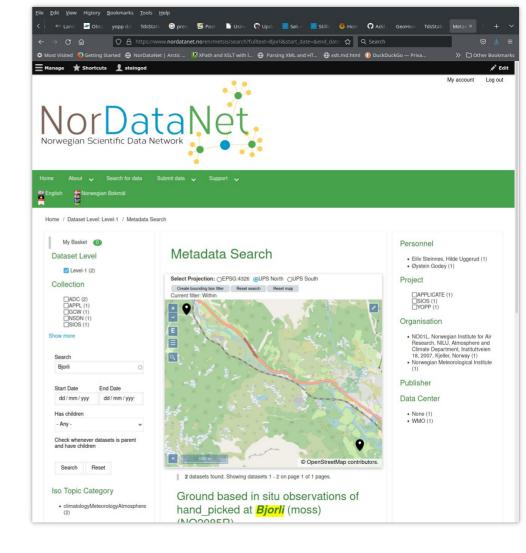
- Currently offering
 - A unified search interface
 - On the fly visualisation and transformation of data depending on FAIRness and protocols
 - Compliance checker for NetCDF-CF
 - Joint with SIOS and Nansen Legacy
 - Conversion tool for NetCDF-CF
 - Joint with SIOS and NMDC
- Ongoing work
 - Integration with NIRD
 - Joint with Nansen Legacy and SIOS InfraNOR
 - Improving visualisation and transformation
 - Improving data documentation support to data providers

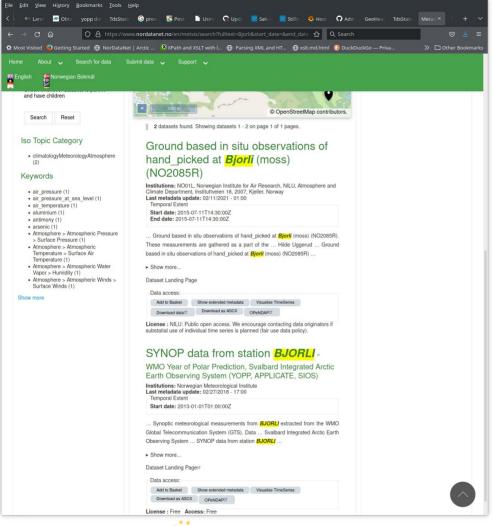




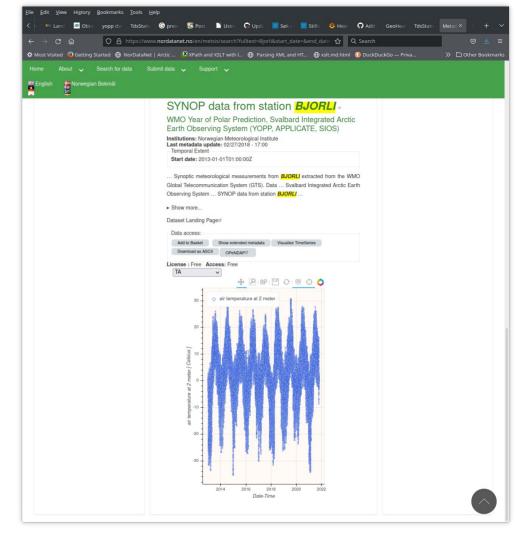


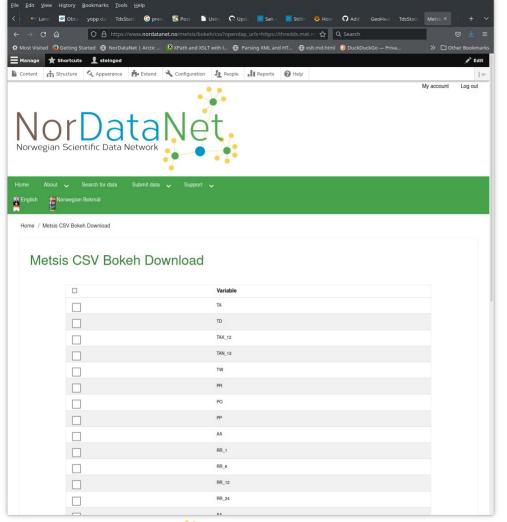






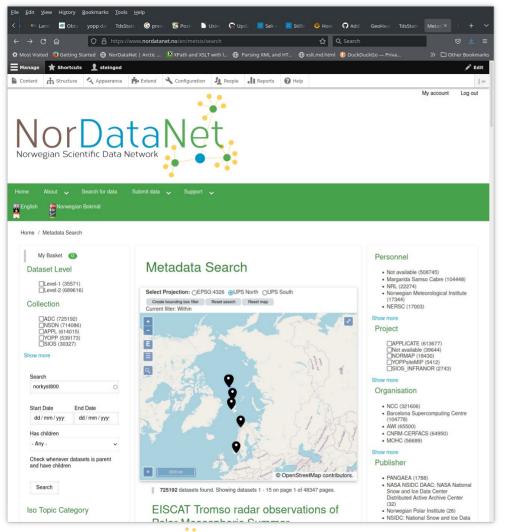




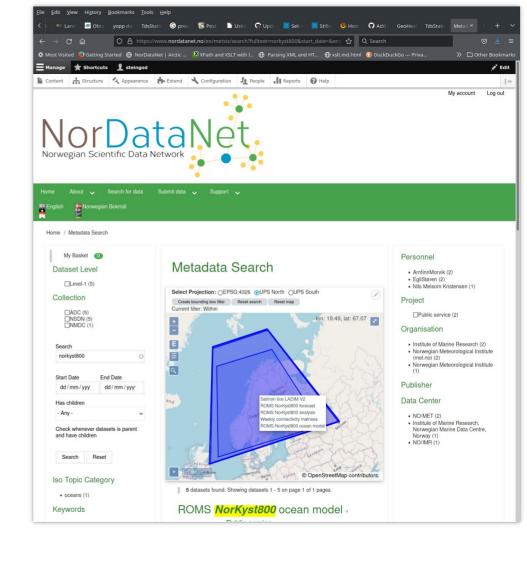


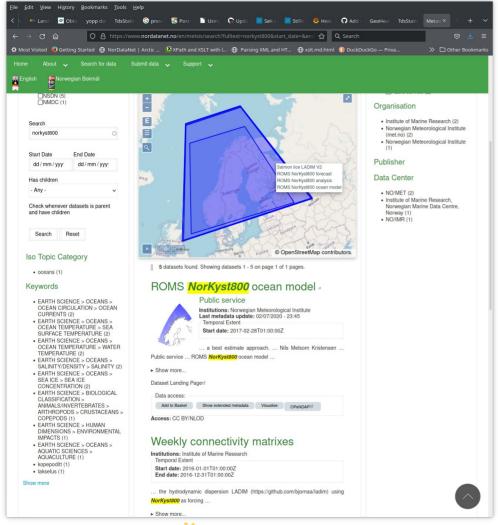


Eile Edit View History Bookmarks Tools Help			
< *** Land ■ Obta: yopp dai TdsStatic ⊕ presi ⑤ Post ▶ Usin; ♠ Updi Sok o			
			Other Bookmarks
	Search for data Submit data Support		
English Norw			
. 💆	П		
		RR_12	
		RR_24	
		SA	
		EE	
		FF	
		DD	
		FG_010	
		FG	
		w	
		NN	
		HL	
		CL	
		CM	
		CH	
		ww	
		W1	
		HW HW	
		PW	
		HWA	
		PWA	
	CSV	v	
	Submit Go back to search		

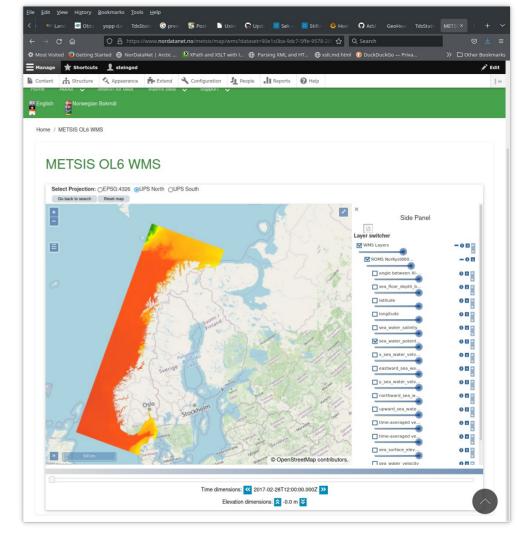


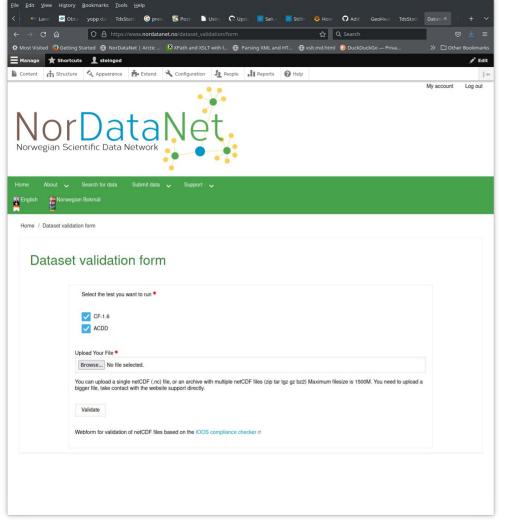




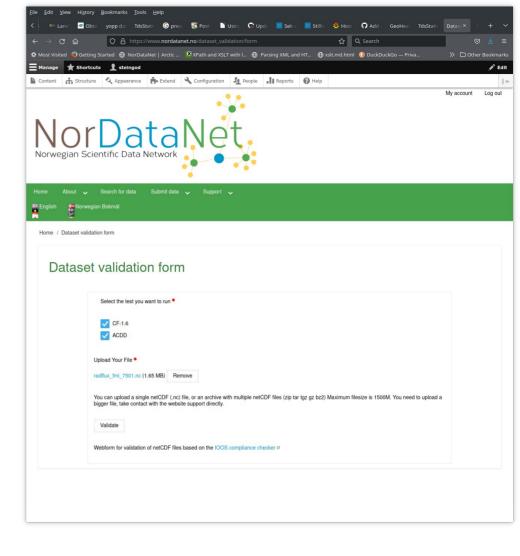


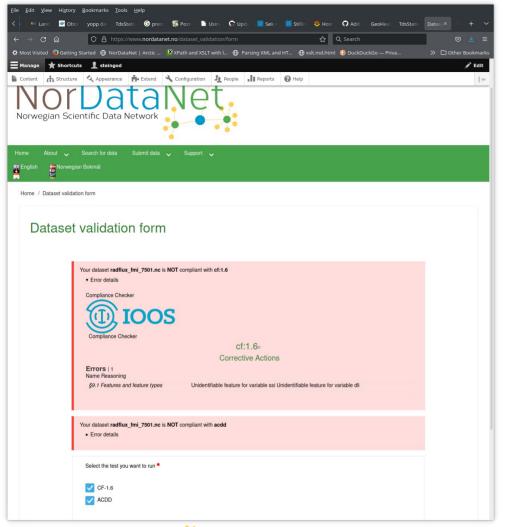




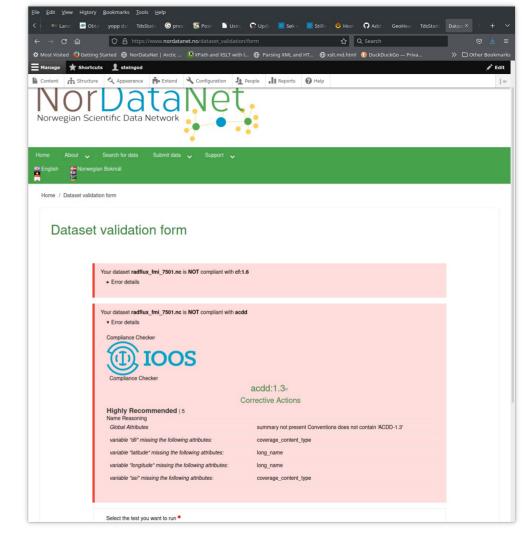


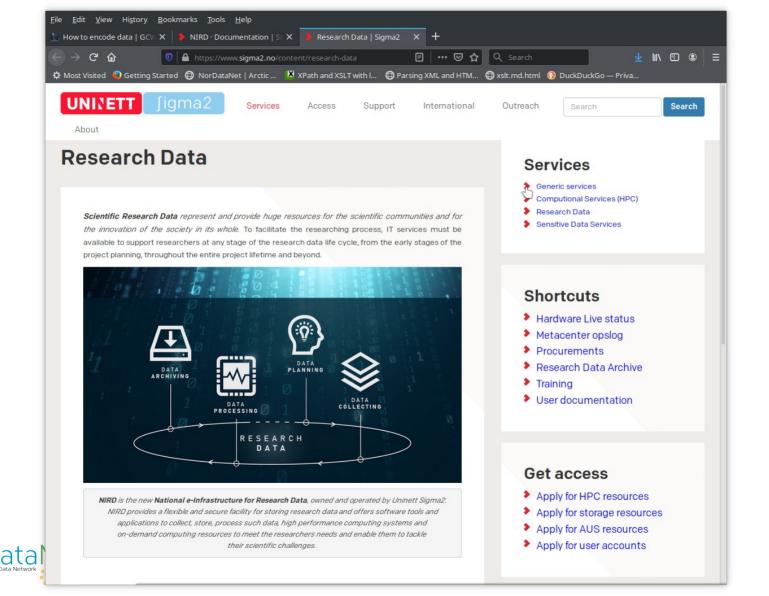




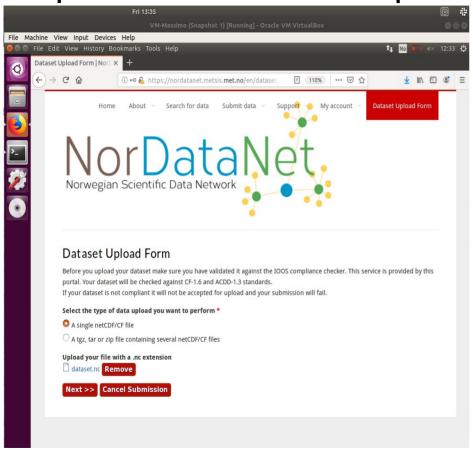








Upload Interface to expose data from NIRD archive (1)



- The user has the possibility to upload a single dataset in NetCDF/CF format.
- A group of files (tgz/zip file) containing NetCDF/CF files.
- The NetCDF files must follow the CF/ACDD conventions in order to be able to extract the correct metadata from the dataset.
- Validation tools for NetCDF/CF are integrated in the work flow.

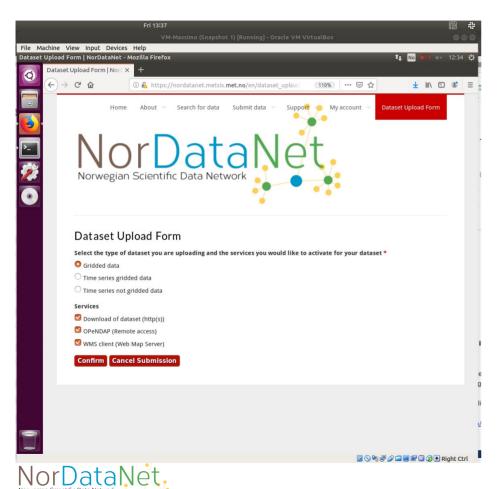


Upload Interface to expose data from NIRD archive (2)

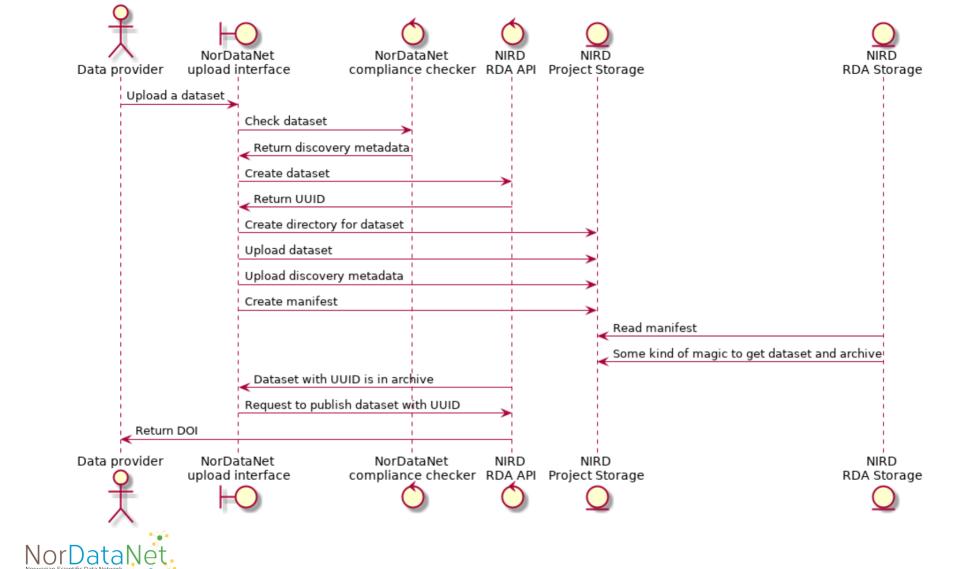


- When submitting the dataset, it will be checked with respect to be CF/ACDD conventions.
- If these tests are passed, metadata are extracted and prepared to be sent to NIRD.
- The user can also check that the metadata extracted are correct and if not cancel the submission.

Upload Interface to expose data from NIRD archive (3)



- Depending on the type of data submitted the user can decide with type of services needed by NIRD to expose the dataset
- Data download (HTTP)
- OPeNDAP (remote access)
- OGC WMS (map visualization)



Data centres

- Affiliated with NorDataNet
 - Institute of Marine Research / Norwegian Marine data Centre
 - Nansen Environmental and Remote Sensing Center
 - Norwegian Institute for Air Research / EBAS
 - Norwegian Meteorological Institute / Arctic Data Centre
 - Norwegian Polar Institute
 - Norwegian Infrastructure for Research Data (NIRD)
- Other relevant data centers
 - Bjerknes Climate Data Centre



