ARICE Webinar
Data Management

Organized by APECS and AP

Moderation: Josefine Lenz (AWI, APECS & ARICE)






Halldór Jóhannson (Arctic Portal & ARICE)
Introduction on Data Management: Policy, Best practices and Engagement with stakeholders

Øystein Godøy (Norwegian Meteorol. Inst.) and Stein Sandven (NERSC)
INTAROS data management practices: Meta Data and federated Search

Anseok Joo (Arctic Portal)
Data presentation and the ARICE 3D Icebreaker

Q&A

Introduction on Data Management:

Policy, Best practices & Engagement with stakeholders

Halldór Jóhannsson (ARICE & Arctic Portal)






www.arice.eu

Grant agreement No 730965

ARICE Horizon 2020 – Objectives

Making the Arctic accessible for excellent science

Providing better capacities for marine-based research in the ice-covered Arctic Ocean

Networking

- Optimal use of existing polar research vessels
- Arctic Research Icebreaker Consortium
- Which shares and jointly funds operational ship time

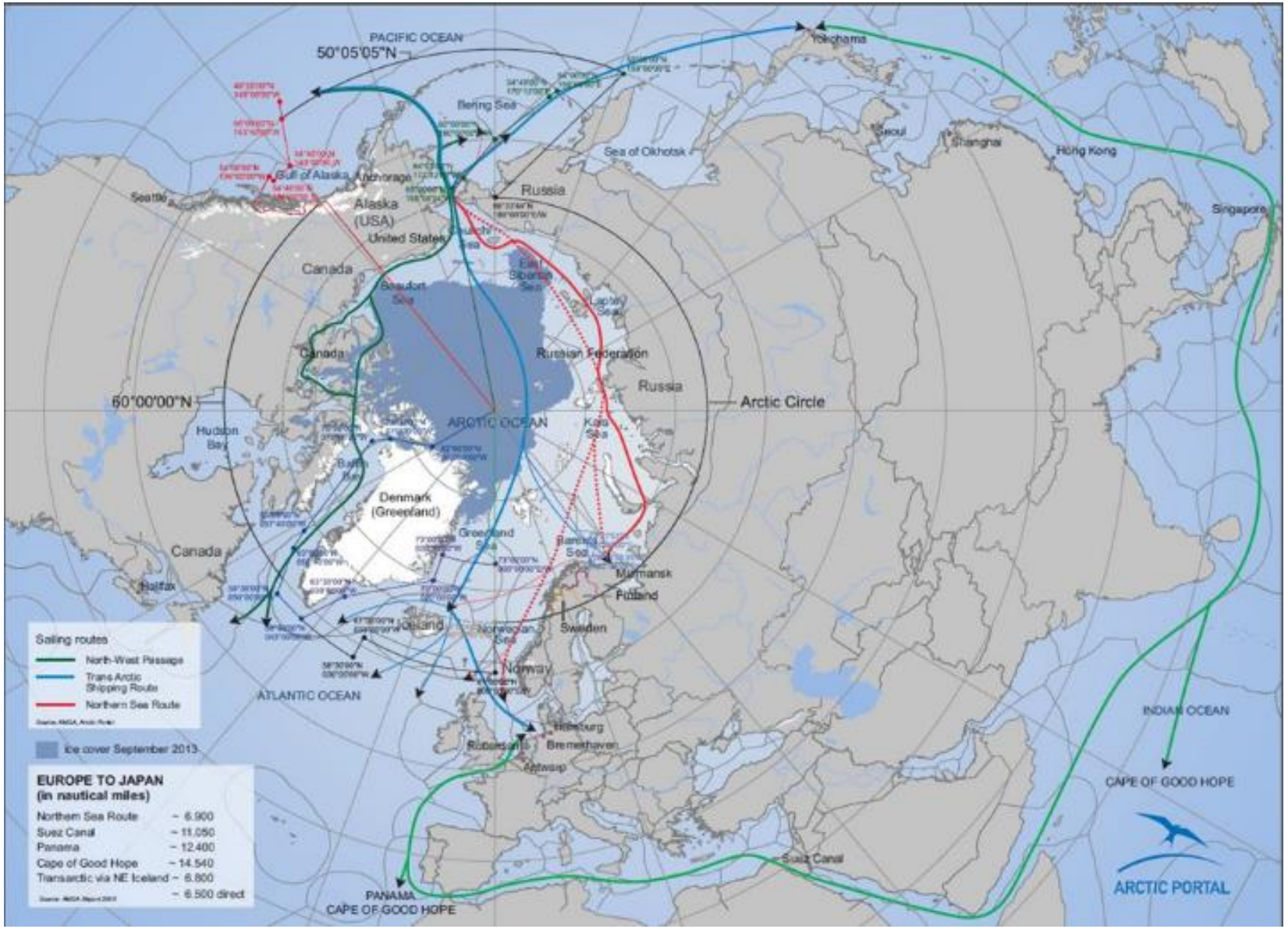
Joint Research Activities

- Partnering with the maritime industry
- Exploring into new key technologies
- Innovative 3D Virtual Icebreaker

The Arctic

- ✓ Interest and awareness about the Arctic and its global importance economically, environmentally, politically and strategically is rising.
- ✓ Our Climate and consequently our Environment is undergoing change.
- ✓ The Arctic is estimated to hold as much as 30% of the world's unused resources.
- ✓ Shipping is increasing rapidly in the region. Year-round Trans-Arctic shipping is a real possibility in the very near future.
- ✓ Tourism is increasing rapidly.
- ✓ The Arctic and the North - An area of opportunities and threats!
- ✓ Need for knowledge-based and responsible development, for the benefit of the regions Indigenous and local people and the Global community.

Shipping in the Arctic



www.arice.eu

Tourism and Shipping in the Arctic

The increasing interest in the Arctic region – how can we raise awareness, increase knowledge and benefit stakeholders?

Policy and Arctic Data

- Arctic Council
  - Working groups and projects
- International Arctic Science Com - IASC
- Sustaining Arctic Observations Networks - SAON (AC, IASC, WMO
  - Arctic Data Committee and Committee on Observations and Networks
- National and regional bodies – EU Horison, NSF, etc.
- Organisations – EPB, Northern Forum, Uarctic, AEC…..

Need For Information, Consulting and Data Services

- ✓ New regulations and forums have been and are being implemented, such as the IMO Polar Code, in order to regulate shipping in the Arctic to increase safety and reduce the risks of environmental disasters.
- ✓ The Arctic Council has made three binding agreements on search and rescue, oil spill response and science cooperation.
- ✓ The science ministers of the Arctic countries and observing countries, met in Washington September 2016 and again in Berlin October 2018 to discuss the need for better observations and access to quality data services. Third meeting to be held in Japan, co-lead by Iceland, in November 2020. - *The follow-up actions will include the implementation of the observing system and data management in the Arctic - www.asm3.org*
- ✓ Recognition of the need of sound data services to increase knowledge and support policy and responsible development is rapidly increasing.
- ✓ Arctic Portal is a partner in many of the current relevant data and knowledge projects in the Arctic region.

EU Arctic Cluster

- **EU – ARCTIC / POLAR CLUSTER**
- Lead by EU-PolarNet in cooperation with EPB
- Including EU Arctic and Antartic funded projects

- Focusing on:
  - **Data Management**
  - **Communication & Outreach**
  - **User Engagement**
  - **Education and Training**

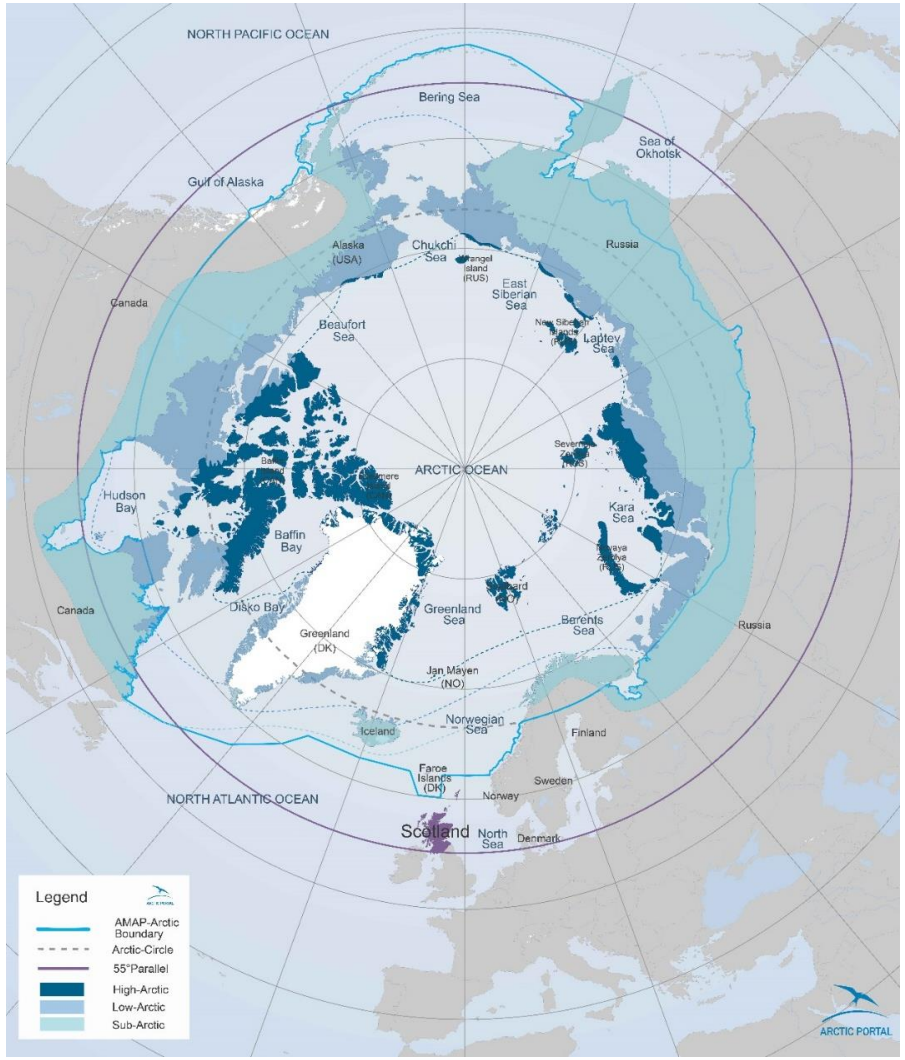
  - **www.eu-polarcluster.eu**

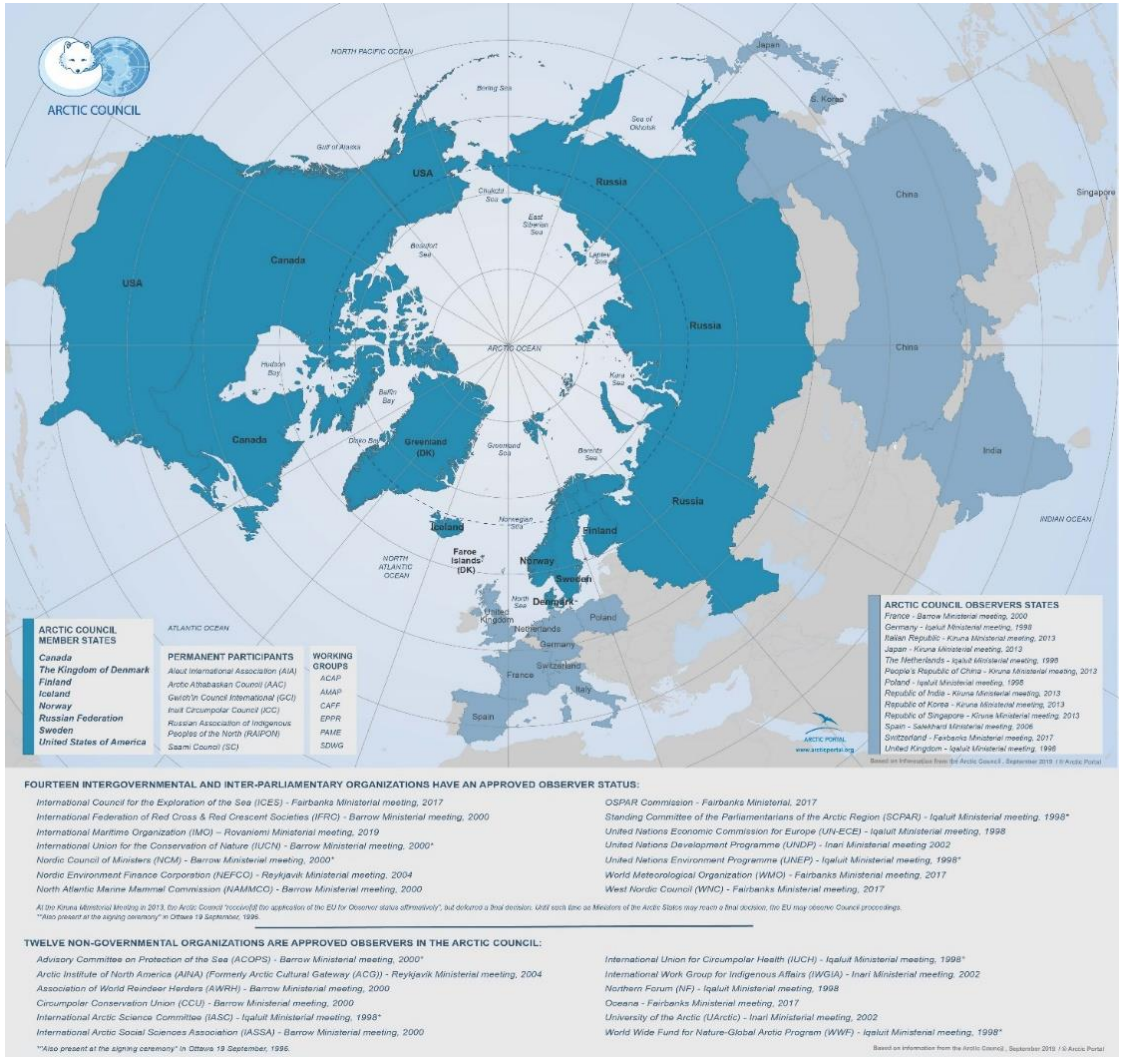
EU Polar Cluster

www.polarcluster.eu

PROVIDING ANSWERS FOR A CHANGING ARCTIC AND ANTARCTIC

Data Communication





Examples of recent mapping and outreach services: Scotland and the Arctic and Arctic Council member states and observers

INTAROS – Integrated Arctic Observation System

Assessment of existing Arctic observing systems and data

Stein Sandven (NERSC)






www.arice.eu

Grant agreement No 730965

Assessment of existing Arctic observing systems and data

Contributing authors:

Stein Sandven, Pirazzini Roberta, Hanne Sagen, Michael Tjernström, Torill Hamre, Andreas Ahlstrøm, Ingo Schewe, Georg Heygster, Carsten Ludwigsen, Mathias Goeckede, David Gustafsson, Agnieszka Beszczynska-Möller, Thomas Soltwedel Georg Heygster, Mikael Sejr, Francisco Navarro, Mathilde Sorensen, Truls Johannessen, Finn Danielsen and other scientist form the INTAROS consortium

INTAROS receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727890

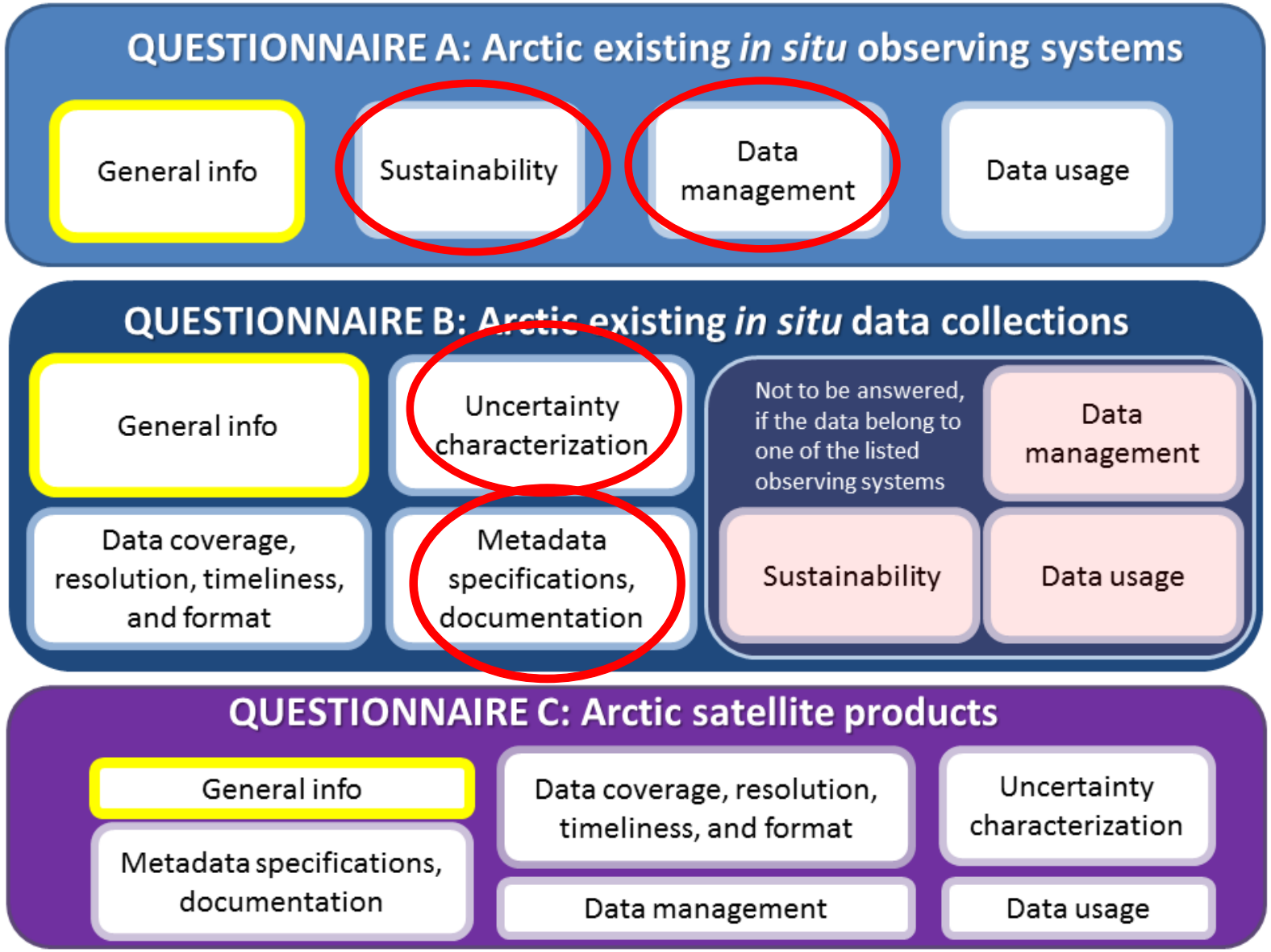
Assessment method

3 QUESTIONNAIRES were defined to collect specific information about:

A. The Arctic existing in situ observing systems (58 replies)
B. The Arctic in situ data collections: existing and exploited (149 replies)
C. The Arctic satellite products: existing and exploited (29 replies)

The questionnaires were web-based, using Google form, available to partners and collaborators through the INTAROS internal web page (https://intaros.nersc.no/)

Content of the survey



Example: Data storage

1. Data are not stored in any institutional repositories
2. Data are stored in institutional repositories
3. Data are stored in distributed repositories shared by several institutions
4. Data are stored in National repositories according to legal obligations
5. Data are stored in National data repositories without legal constraints on their location
6. Data are stored in International data repositories

1: lowest maturity score, 6: highest maturity score

Example: Data access – level 1 - 4

1. Data is available on request to trusted users (Data record is not ready yet to be given to users; it may be available to beta-users for testing.
2. Data is available on supervised request through originator (Data record is now ready to be given to users without any restrictions on academic usage. Users can get the measurement data by requesting it from the data originator)
3. Data is available on automated request through originator (Data record is now ready to be given to users without any restrictions on academic usage.)
4. Data and documentation are available on supervised request through originator (Measurement series and appropriate documentation to understand the measurements are publicly available for academic use through the data provider. Academic reuse is permitted)

1: lowest maturity score, 6: highest maturity score

Example: Data access – level 5 – 8

5. Data and documentation are available on automated request through originator (Data record and appropriate documentation to understand the measurements are publicly available for academic use through a publicly accessible site. Academic reuse is permitted)
6. Data and documentation are available through originator and recognized data portal (Data record and documentation are available through a recognized and measurement-appropriate data portal such as the Copernicus Climate Change Services Data Portal, NDACC portal or NOAA's National Centers for Environmental Information)
7. As in (6) + source data, code and metadata available upon request (the source data, metadata and any processing code is also archived by the data provider allowing subsequent reprocessing of the full measurement series if required by a third party).
8. As in (7) + no access restrictions apply (there are no restrictions on use or reuse of the data, metadata, or code and all aspects are made available free of charge

1: lowest maturity score, 8: highest maturity score

Result: Example of ocean observing system: A-TWAIN mooring array north of Svalbard

Maturity Assessment Matrix:

A-TWAIN

Methadata	Documentation	Uncertanty	Data	Sustanability
Standards	Formal Description of Measurement Methodology	Traceability	Storage	Scientific and Expert Support
Collection Level	Formal Validation Report	Comparability	Access	Funding Support
File Level	Formal Measurement Series User Guidence	Standards	User Feedback	Site Representativeness*
		Validation	Updates to Record	
		Uncertanty Quatification	Version Control	
		Routine Quality Management	Data Preservation	

* if applicable

1: lowest maturity score, 6: highest maturity score

1	2	3	4	5	6	Not applicable or no data provided

Results for each Ocean and Sea Ice observing system regarding Sustainability and data management

Observation System	Platform	Sustainability			Data Management						Data repository
		Scientific and expert support	Funding support	Site representativeness*	Data storage	Data access	User feedback	Updates to record	Version control	Long term data preservation	
A-TWAIN	Moorings	5	3	N/A	2	6	1	2	2	4	Norwegian Polar Data Centre
A-TWAIN Poland	Moorings	4	3	N/A	3	5	2	2	2	2	IOPAN database
AREX (Long-term large-scale monitoring program)	Repeated sections	4	4	N/A	6	5	2	3	2	3	IOPAN database
Argo Poland	Buoys	4	3	N/A	6	6	1	3	4	4	ARGO GDAC (Coriolis)
Canada Basin Acoustic Propagation Experiment (C	Moorings	3	2	N/A	2	2	2	2	2	3	Scripps Institution of Oceanography
FRAM - (FRontiers in Arctic marine Monitoring)	Buoys, Moorings, Vess	6	5	N/A	6	5	2	4	4	4	PANGAEA
Fram Strait Multipurpose Acoustic System	Moorings	3	2	N/A	2	2	1	2	2	3	NERSC (harvested by NMDC and NorDataNet)
Greenland Ecosystem Monitoring Programme	Moorings, Vessels	5	6	5	2	4	2	2	2	3	GEM database
IMR Barents Sea Opening mooring array	Moorings	5	4	N/A	4	3	2	2	1	4	NMDC and ICES
IMR Barents Sea Winter Survey	Vessels	4	6	N/A	4	5	2	4	1	4	NMDC and ICES
IMR fixed hydrographic sections	Vessels	5	5	N/A	4	4	2	3	1	4	NMDC and ICES
IMR fixed hydrographic sections (near coast)	Vessels	4	5	N/A	2	3	2	3	1	3	NMDC and ICES
IMR SI_Arctic vessel mounted ADCP system	Vessels	4	3	N/A	4	3	2	3	1	4	NMDC and ICES
IMR-PINRO Ecosystem Survey	Vessels	5	6	N/A	2	6	2	3	1	4	NMDC and ICES
International Arctic Buoy Programme	Ice buoys array	4	4	N/A	2	3	1	4	1	1	
IOC Tide Gauges in Greenland	Tide Gauges	3	3	3	4	3	2	2	2	3	IOC and DTU
IOPAN Long-term Monitoring in Svalbard Fjords	Vessels	4	4	N/A	6	6	5	4	1	5	IOPAN database
NIVA Barents Sea Ferry Box	Vessels	6	4	N/A	1	4	2	2	2	3	NIVA database
NorArgo	Buoys	6	3	N/A	6	6	5	4		5	
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)

1: lowest maturity score, 6: highest maturity score

Overall results for ocean and sea ice

- Maturity level is generally **Medium** (3, 4) for most of the assessed variables
- 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.
- Moving platforms (Argo floats, gliders) are not yet adapted for **under-ice operations**. It is recommended to develop acoustic positioning systems for these platforms
- There are many gaps in the data coverage in the Arctic, but the **gaps in biogeochemical observations** are particularly important. Therefore we need to **develop and adapt technologies and sensors** to make biogeochemical and biological observations feasible.
- Accessibility to data in **near realtime** from several underwater platforms is a severe limitation. Methods to improve this deficiency need to be developed

Data presentation & the ARICE 3D Icebreaker

Anseok Joo (Arctic Portal)






www.arice.eu

Grant agreement No 730965

Overview

- Scope: How do I take various information stored on a WMS-compliant server and visualise it on a Web browser? A concrete case study.
- Client-server communication protocol: Web Map Service (WMS)
- Tools
  - Server-side: GeoServer
  - Client-side: OpenLayers
- Demo & a look at the internals

Web Map Service

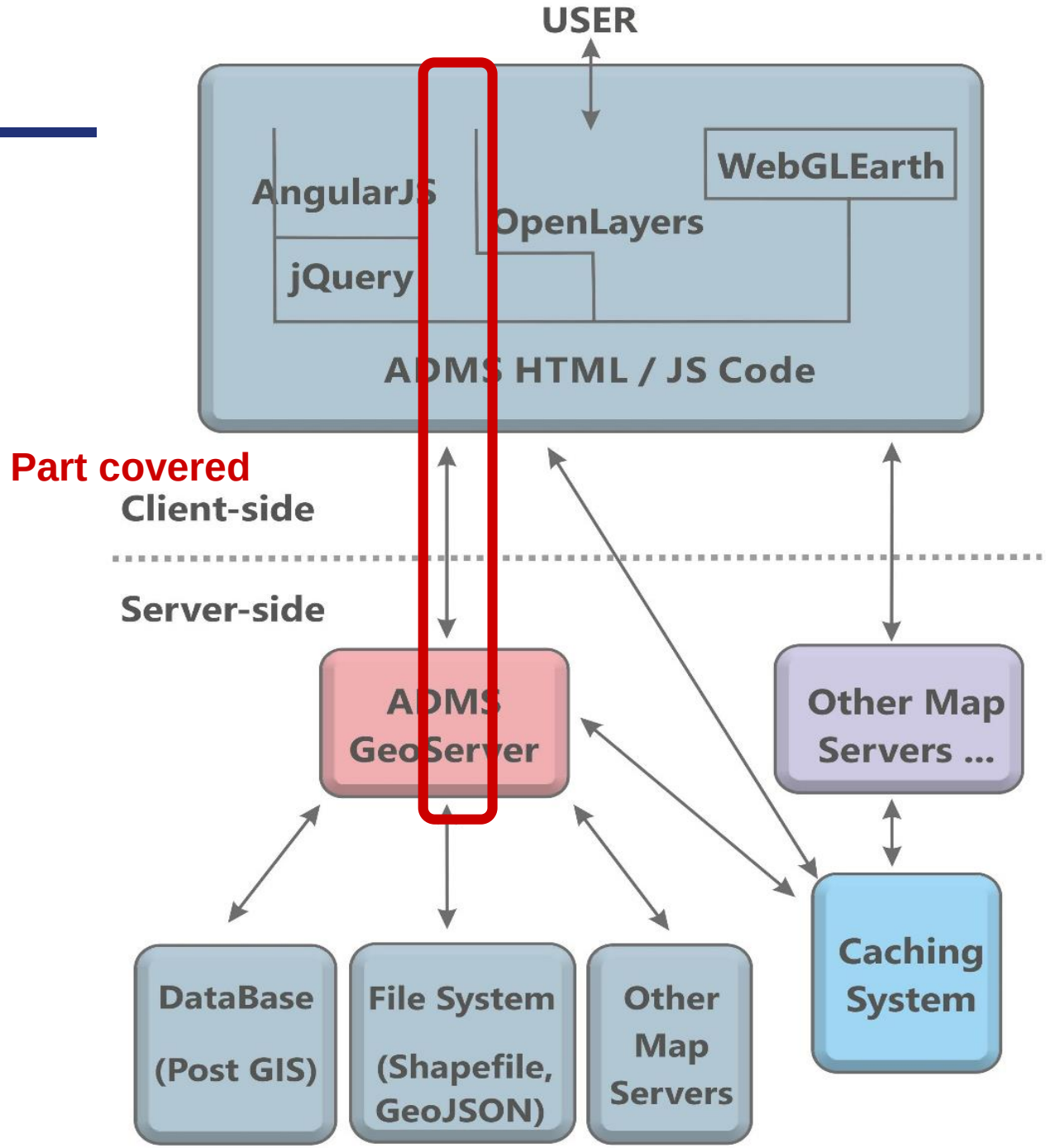
- Open Geospatial Consortium (OGC) defines several standards for communication protocols, languages, and file structures pertaining to geospatial information. www.opengeospatial.org/standards
- Of the standards, Web Map Service (WMS) is concerned with Web visualisation.
- WMS provides the syntax for “Get” requests
  - GetCapabilities: return server info & projections and layers served
  - GetMap: return image file for a part of layer in a projection
  - GetLegendGraphic: return image file for a layer’s legend
  - GetFeatureInfo: return data at a layer at a coordinate

Tools

- GeoServer geoserver.org
  - Reference implementation for OGC standards
  - (Other servers such as deegree, MapServer, QGIS Server, ArcGIS Server comply with WMS too, and so should be functionally identical)
- OpenLayers openlayers.org
  - A JavaScript library for rendering maps on Web browsers
  - Currently the only library handling all WMS “Get” requests and responses

System demo

- Let's see how each of the "Get" requests work.



Arctic Data Management

Perspectives of APPLICATE, SIOS and NorDataNet

Øystein Godøy (Norwegian Meteorol. Inst.)






www.arice.eu

Arctic Research Icebreaker Cons.
ARICE

Grant agreement No 730965

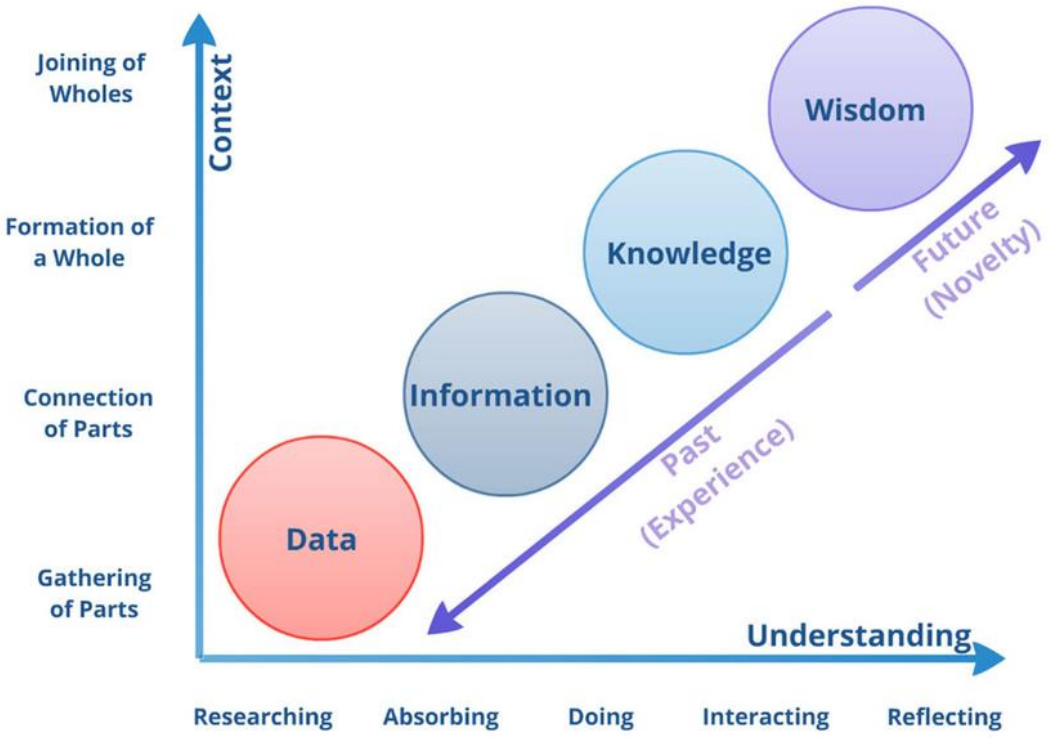
Arctic Data Management

- Nothing different from other data management aproaches
  - But better perspectives on interdisciplinary data management
- Need to bridge between data centres
  - Towards federated search
    - To combine information from different projects, programmes, disciplines etc
- Some examples of federated data management systems
  - Norwegian Scientific Data Network (NorDataNet)
  - Svalbard Integrated Arctic Observing System (SIOS)
  - WMO Global Cryosphere Watch
- For this to work a structured approach on describing datasets at various levels is required
- Activities related to SAON/IASC Arctic Data Committee
  - Working Group on Federated Search
  - Polar Semantics Working
    - Training and hackathon session during Third Polar Data Forum
      - To raise awareness and inform on available resources

The challenge of data

- Need to integrate data across data providers (silos), communities and languages to simplify data consumption.
- Need to combine different types of data.
  - E.g. in situ, remote sensing, numerical simulations and LTK.
  - While minimising the human effort required.
- Need to switch from 80% of human effort on massaging data and 20% on use to the opposite.
- In this context we need to transition from data to knowledge and understanding through connection of the “dots”
  - From a fragmented to consolidated view.
- Need to use semantic web technologies

Russ Ackoff “From Data to Wisdom” -
Journal of Applied Systems Analysis, Volume 16, 1989 p 3-9



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 Accessible:
  - A1. (meta)data are retrievable by their identifier using a *standardized communications protocol*
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
  - A2. metadata are accessible, even when the data are no longer available

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

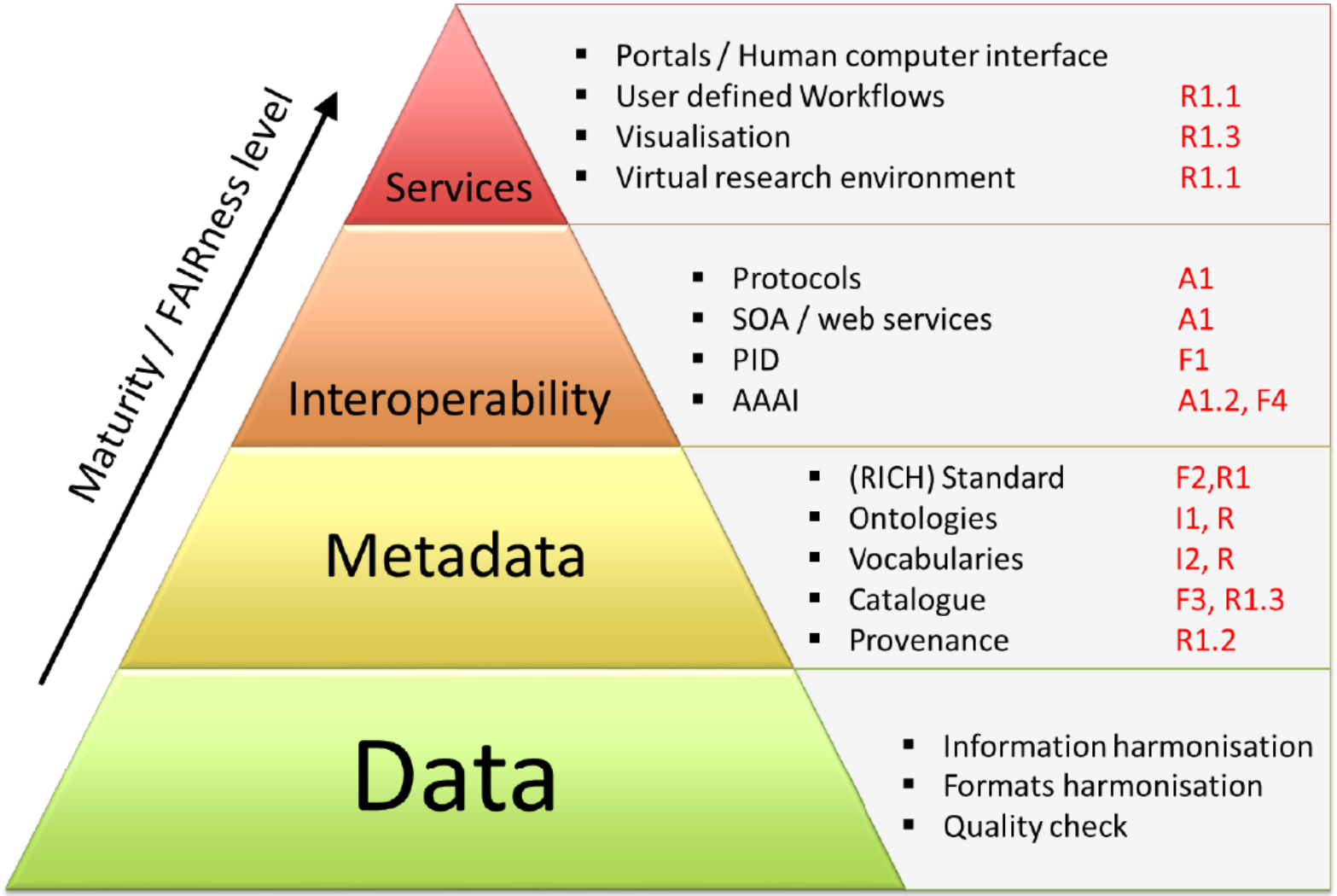
To be Reusable:

R1. meta(data) are richly described with a *plurality of accurate and relevant attributes*

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

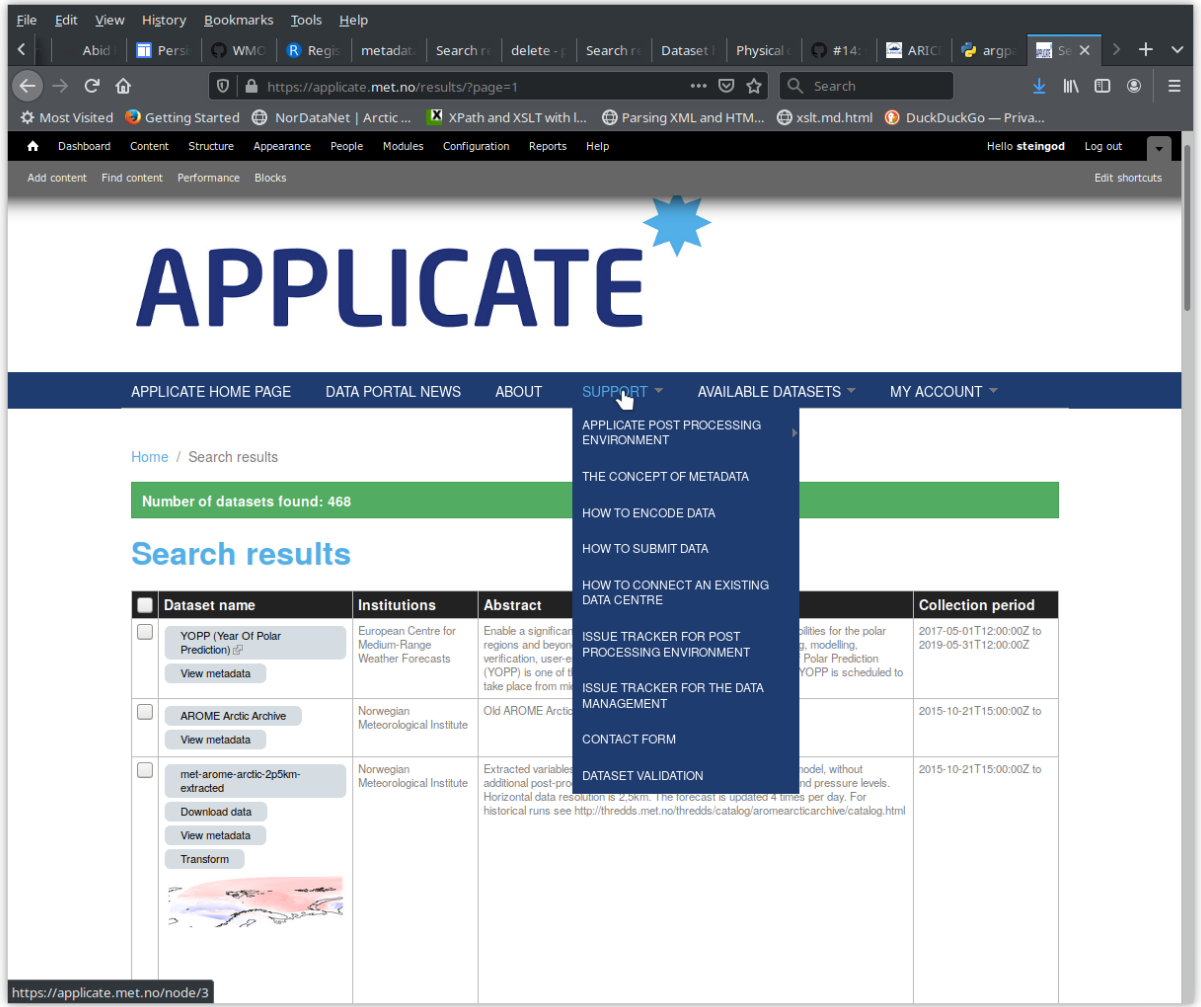

Maturity / FAIRness level
Services
Portals / Human computer interface
User defined Workflows R1.1
Visualisation R1.3
Virtual research environment R1.1
Interoperability
Protocols A1
SOA / web services A1
PID F1
AAAI A1.2, F4
Metadata
(RICH) Standard F2,R1
Ontologies I1, R
Vocabularies I2, R
Catalogue F3, R1.3
Provenance R1.2
Data
Information harmonisation
Formats harmonisation
Quality check

Types of metadata for datasets

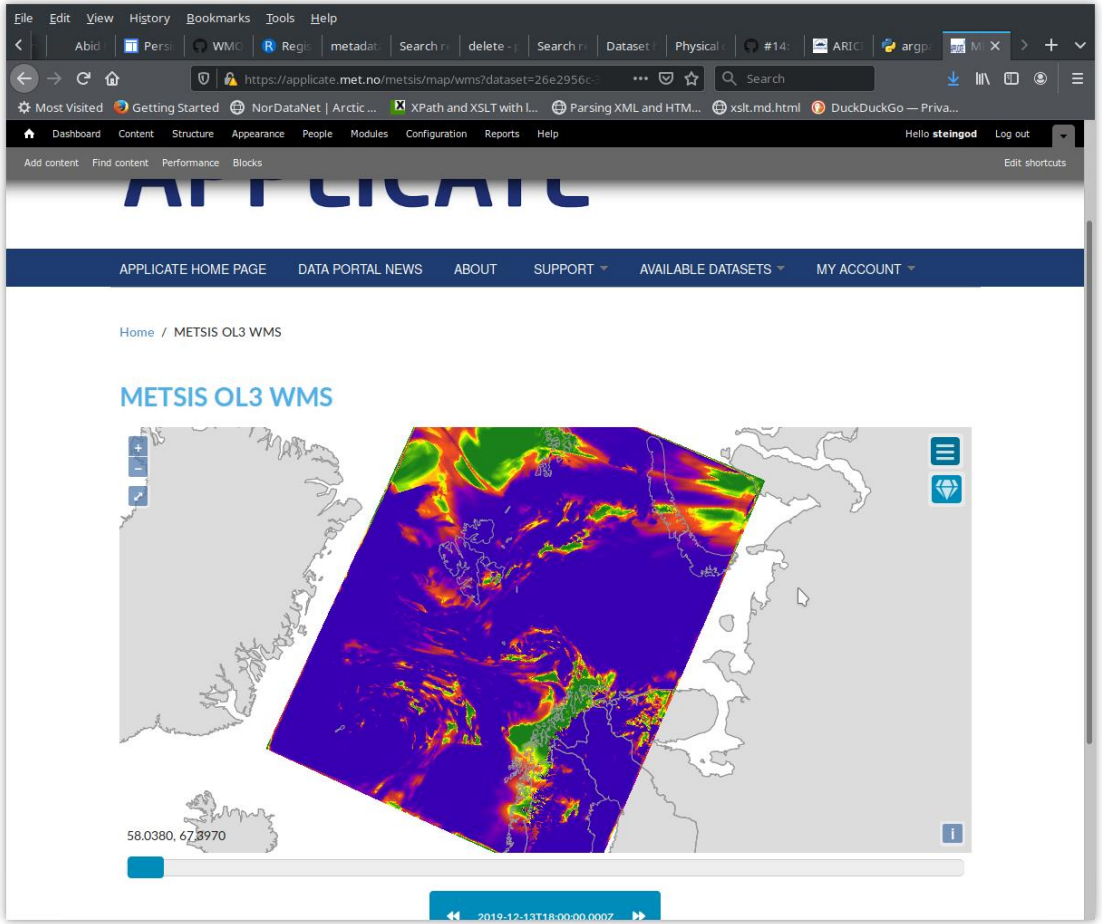
Type	Purpose	Description	Examples
Discovery metadata	Used to find relevant data	Discovery metadata are also called index metadata and are a digital version of the library index card. It describes who did what, where and when, how to access data and potential constraints on the data. It shall also link to further information on the data like site metadata.	ISO19115 GCMD DIF ACDD MMD
Use metadata	Used to understand data found	Use metadata are describing the actual content of a dataset and how it is encoded. The purpose is to enable the user to understand the data without any further communication. It describes content of variables using standardised vocabularies, units of variable, encoding of missing values, map projections etc.	Climate and Forecast Convention BUFR GRIB DwCA
Configuration metadata	Used to tune portal services for datasets for users.	Configuration metadata are used to improve the services offered through a portal to the user community. This can be e.g. how to best visualise a product.	MMD
Site metadata	Used to understand data found	Site metadata are used to describe the context of observational data. It describes the location of an observation, the instrumentation, procedures etc. To a certain extent it overlaps with discovery metadata, but more so it really extends discovery metadata. Site metadata can be used for observation network design.	WIGOS OGC O&M

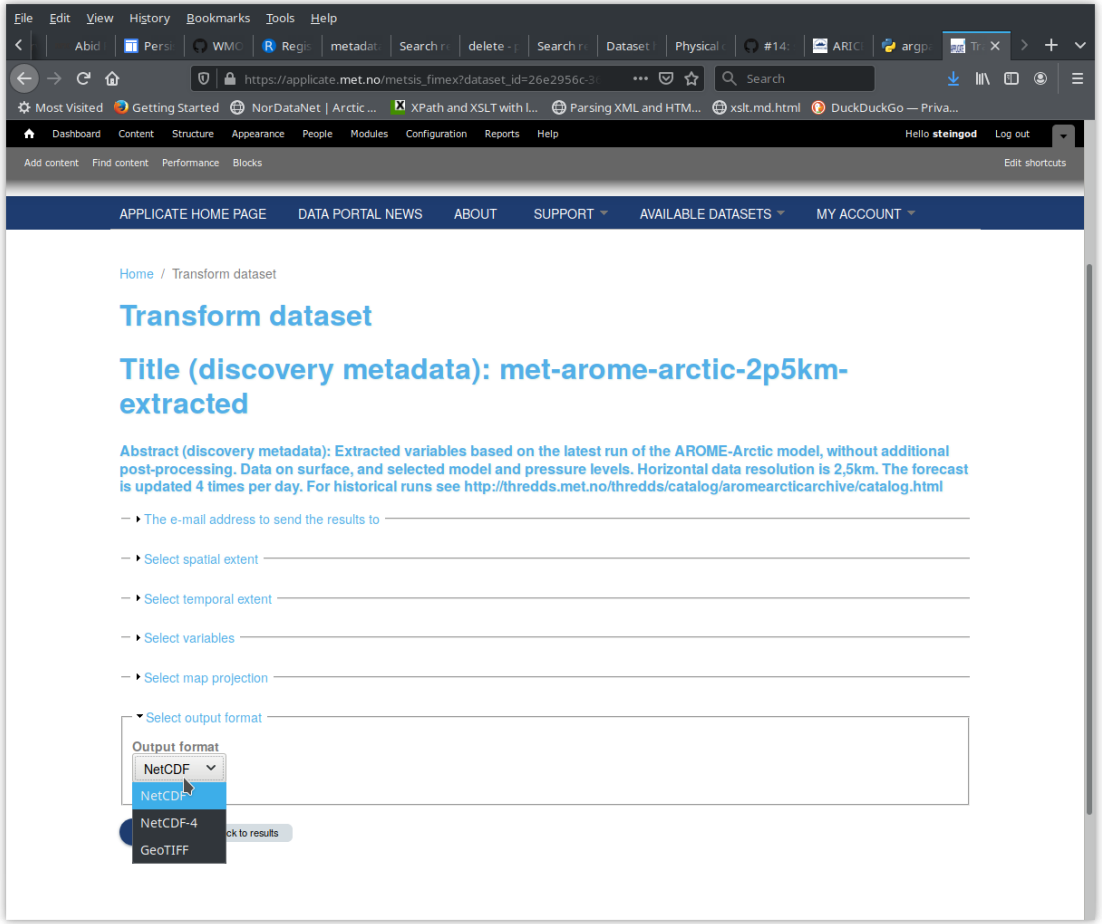
APPLICATE Data Management

- Part of EU-Polar-Cluster
- One catalogue providing access to all data
- A contribution to YOPP
- Relying on NetCDF/CF with discovery metadata embedded as ACDD elements
- Primarily numerical simulations provided by partners
- Data extracted from WMO GTS
- ECMWF YOPP Reference dataset
- Data hosted by MET and ECMWF

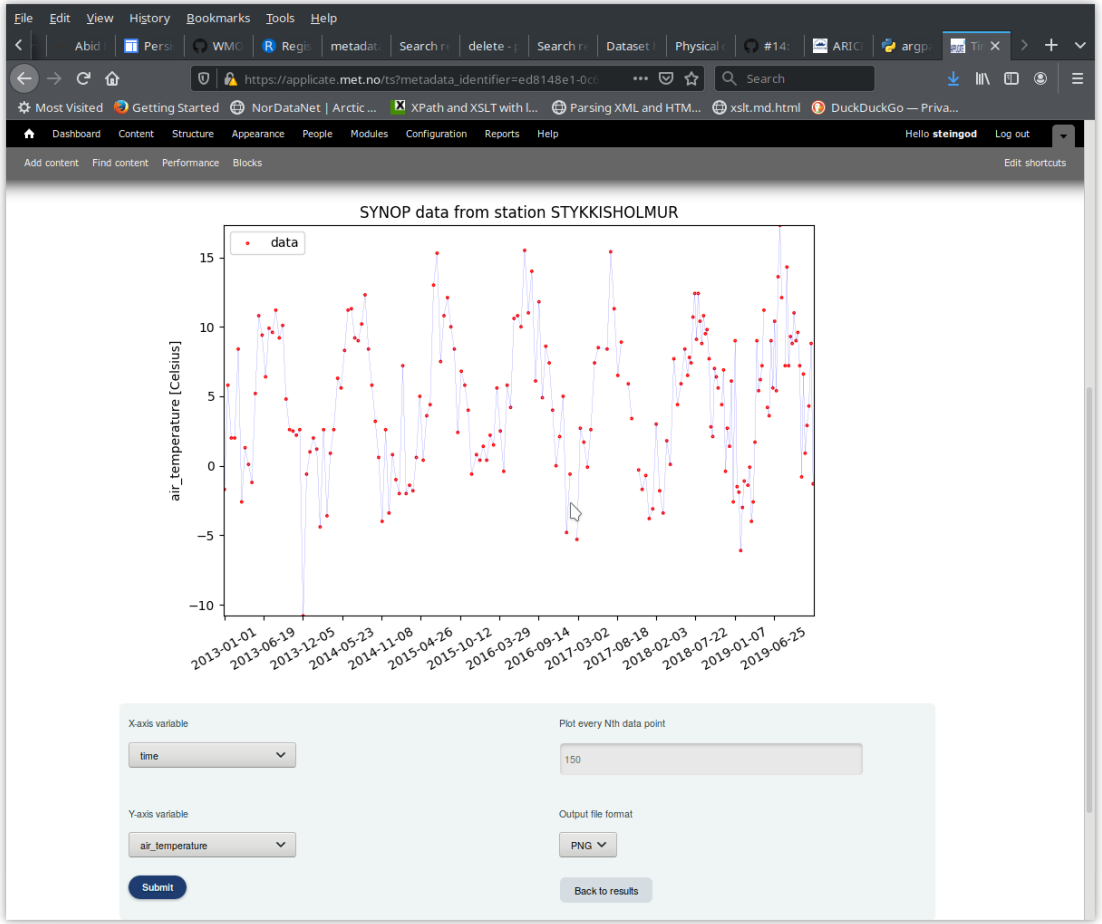


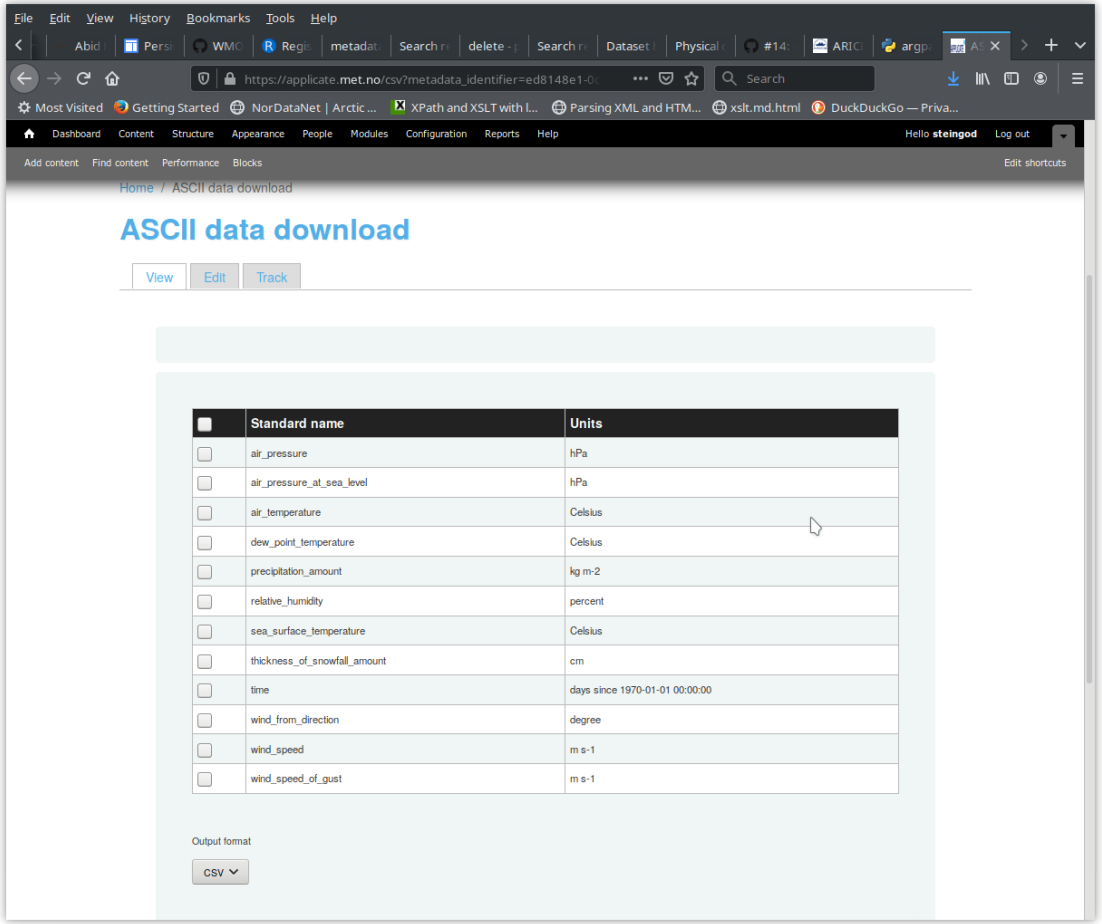
Services for gridded data





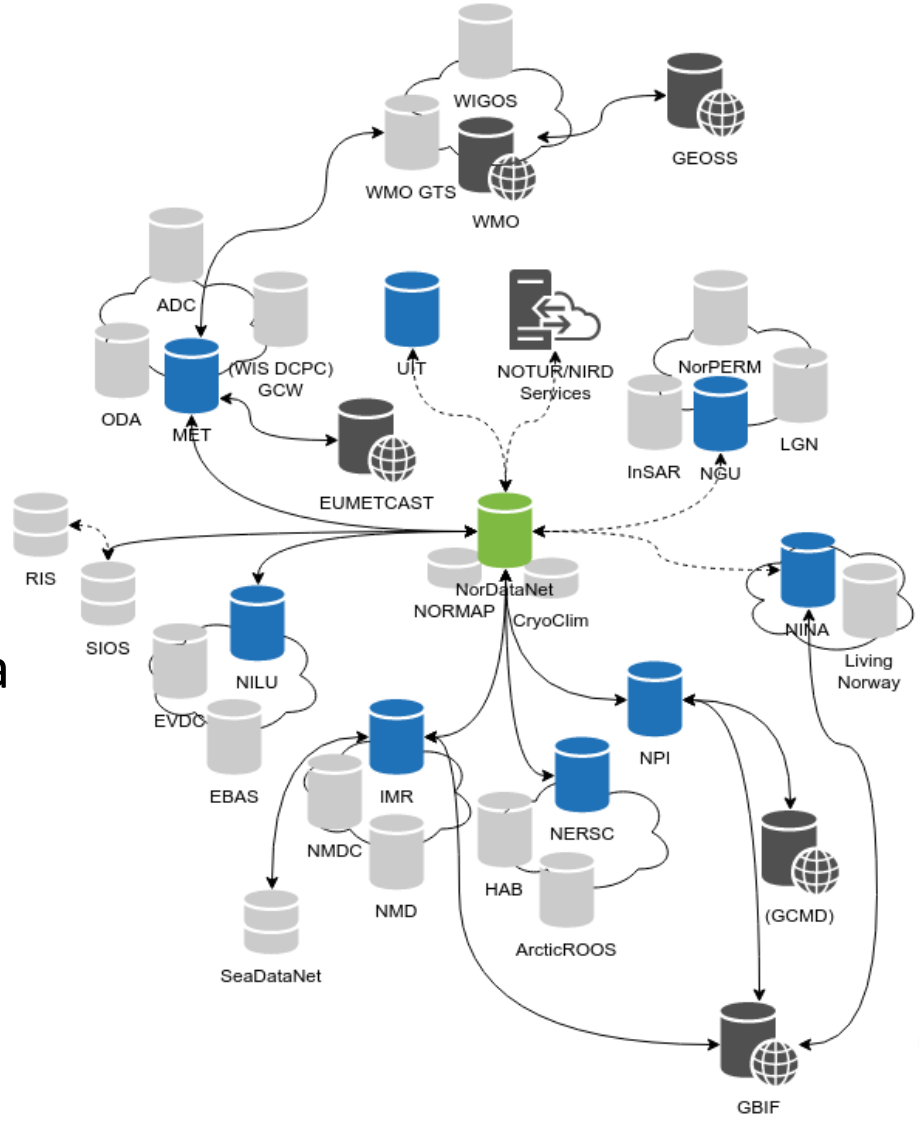
Services for observations at stations

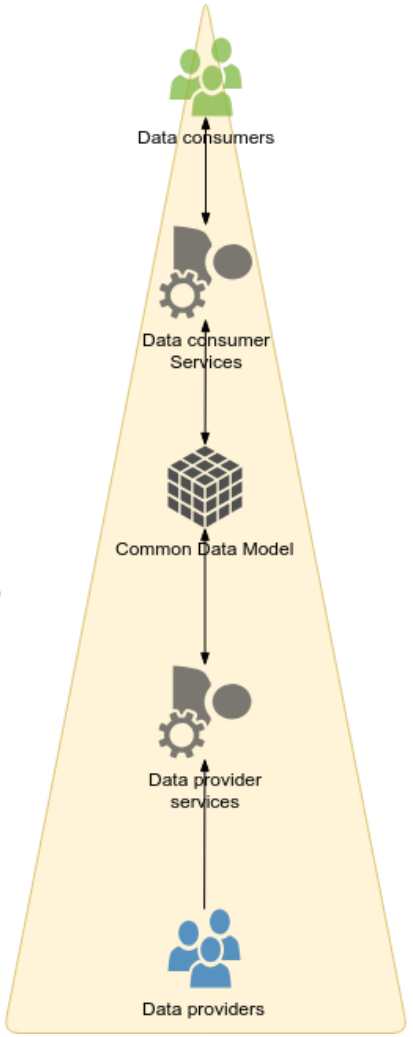




Norwegian Scientific Data Network - NorDataNet

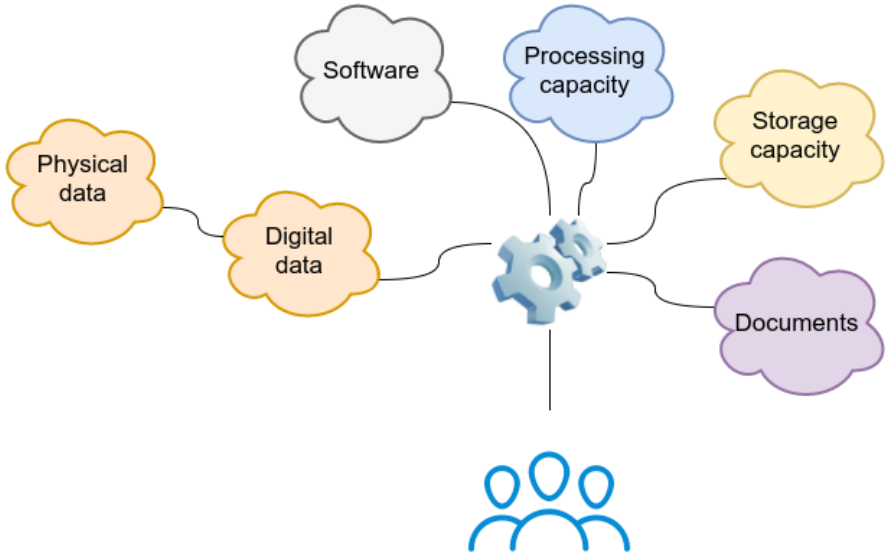
- Distributed data management network
- Development funded by the Research Council of Norway
- Activities are supporting other efforts like SIOS, GCW, NBS, YOPP, ...
- Building on existing data centres nationally and the legacy of IPY
- Adding efforts like NORMAP and CryoClim as collections
- Discovery metadata are harvested into a unified catalogue
- Moving towards standardised data hosted by core partners
  - Actionable data as basis for user oriented services
  - Externally harvested information may be of any kind

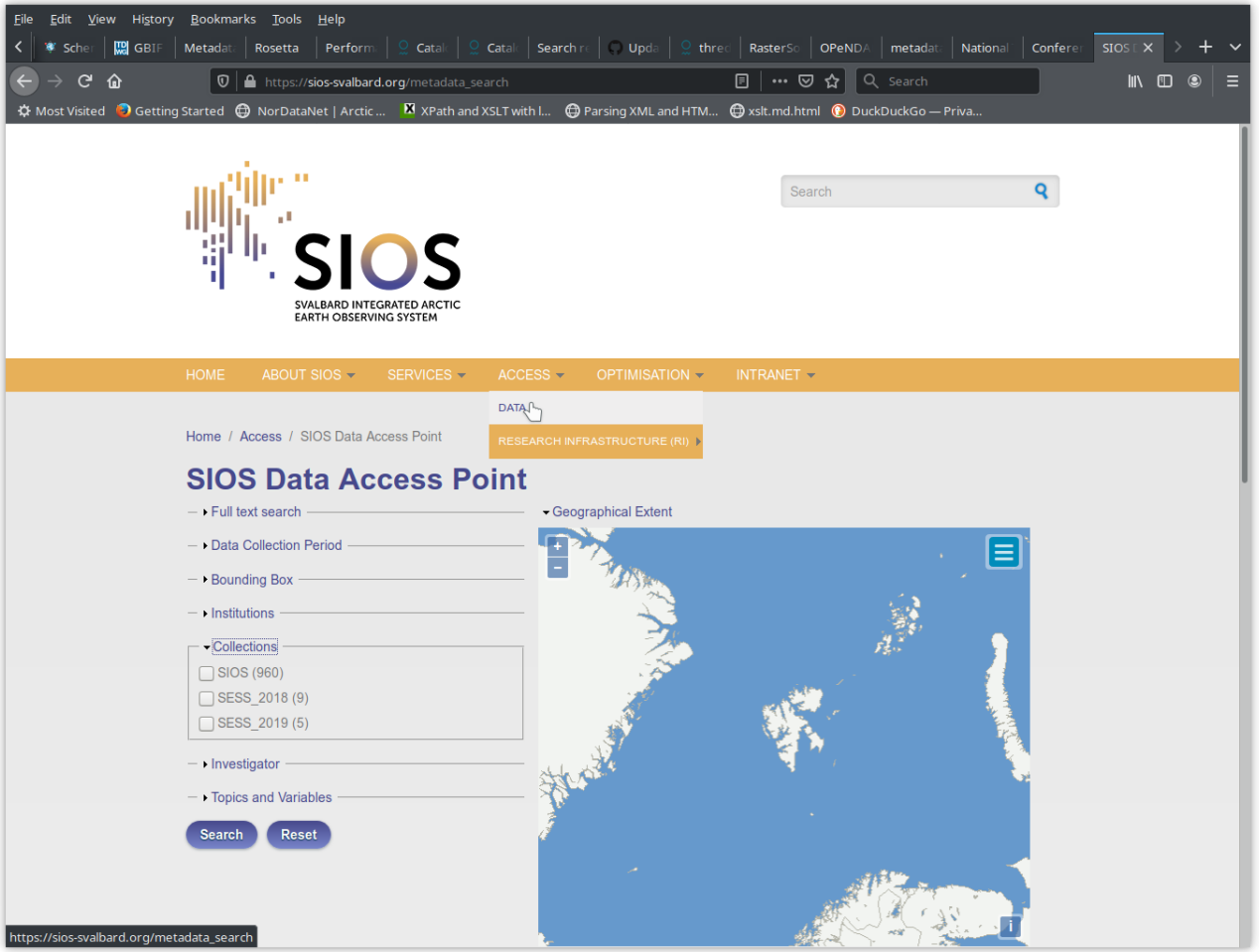




SIOS Data Management System

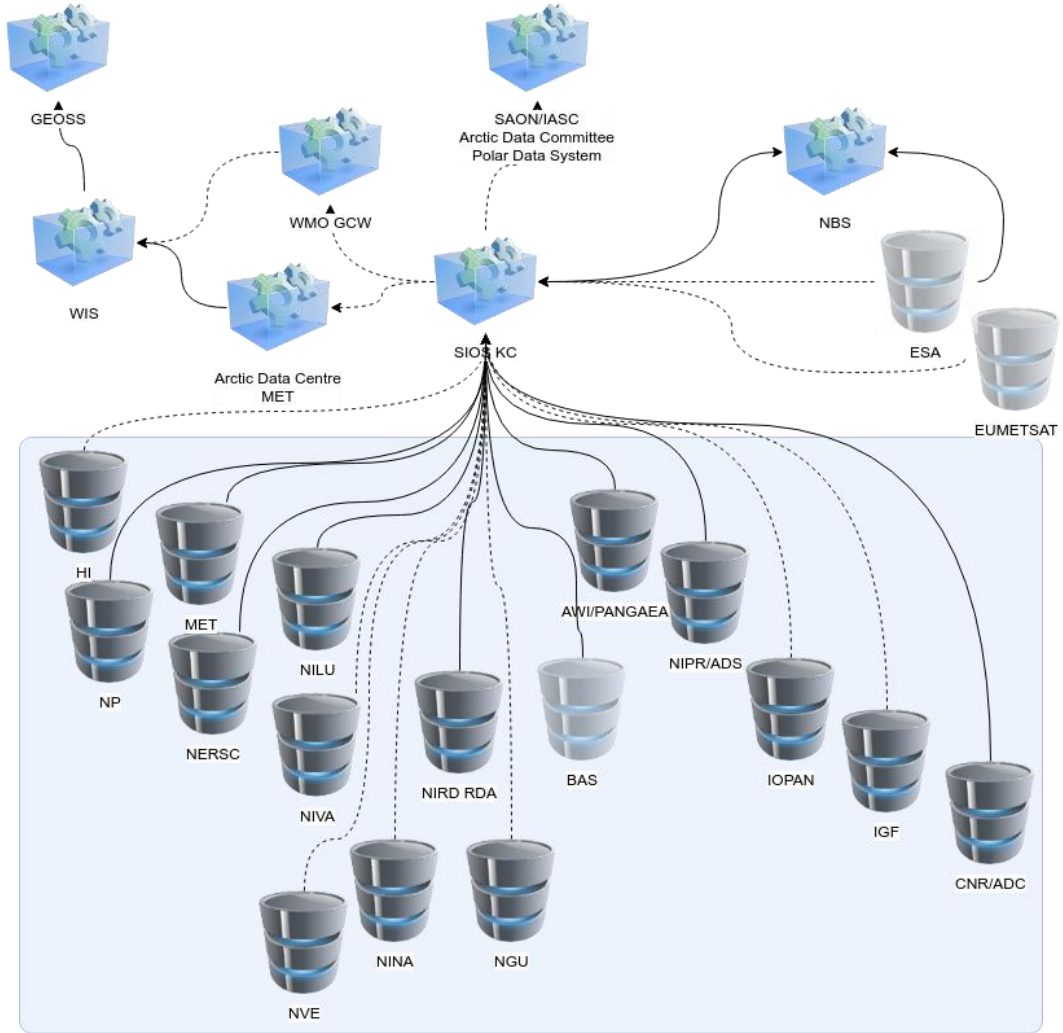
- Main principles
  - Open and free data sharing
  - Distributed data centres
    - Harvesting discovery metadata
  - International standards
  - Harmonisation of SIOS Core data





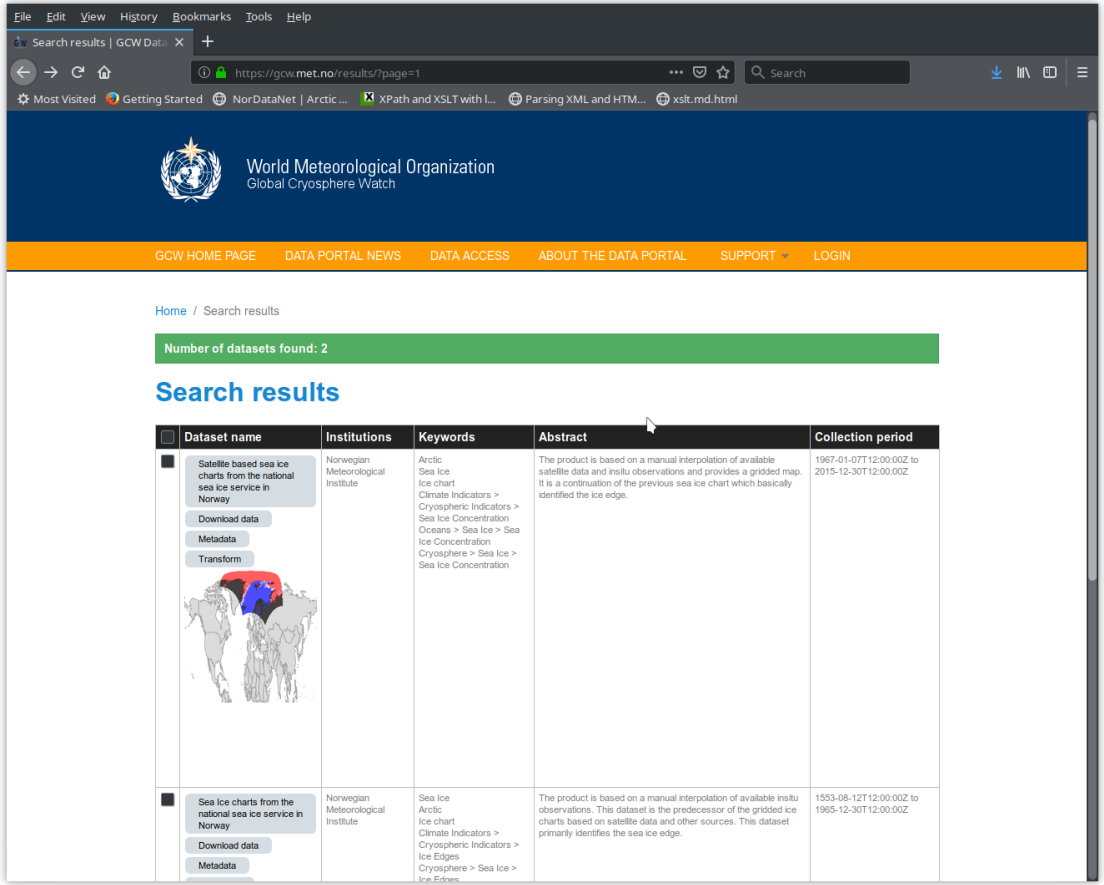
The SIOS Data Management System

- Integration of existing data centres into a unified system.
- Each data centre has its own procedures and technical solutions tailored to the needs of that data centre.
- SIOS will not change this, but bridge
  - Using internationally accepted interoperability standards and technologies
  - Which can be added as a layer between the data and the SIOS Data Management System



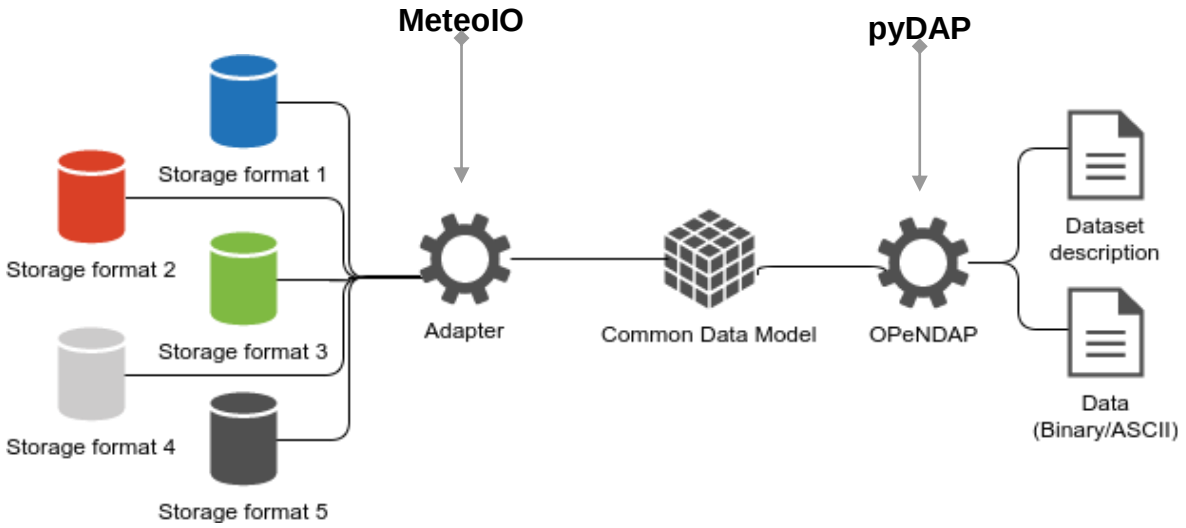
Purpose of GCW Data Management

- To provide an overview of the datasets that are relevant for WMO Global Cryosphere Watch
- To provide access to datasets
  - Real time data streams
  - Archive access
- Distributed Data Management
  - Metadata driven
  - *Not hosting data*
- To connect GCW with
  - WMO Information System
  - WMO Integrated Global Observing System



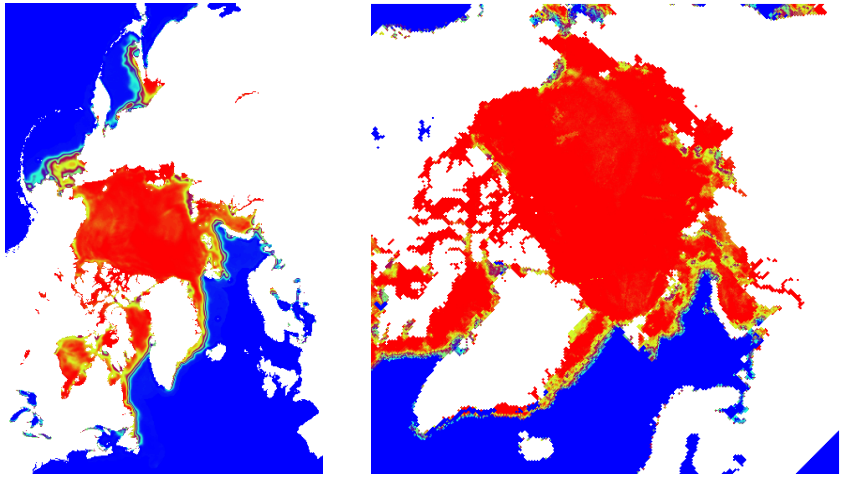
Interoperability with CryoNet stations

- WSL/SLF Software stack in support for GCW
  - Software for discovery and data interoperability
  - Solution capable of integration a wide range of input streams, including RDBMS
  - Includes QC and transformation from many formats to NetCDF/CF
- Takes care of data from measurement to published data where it can be picked up by services
- Low cost software stack for small data centres

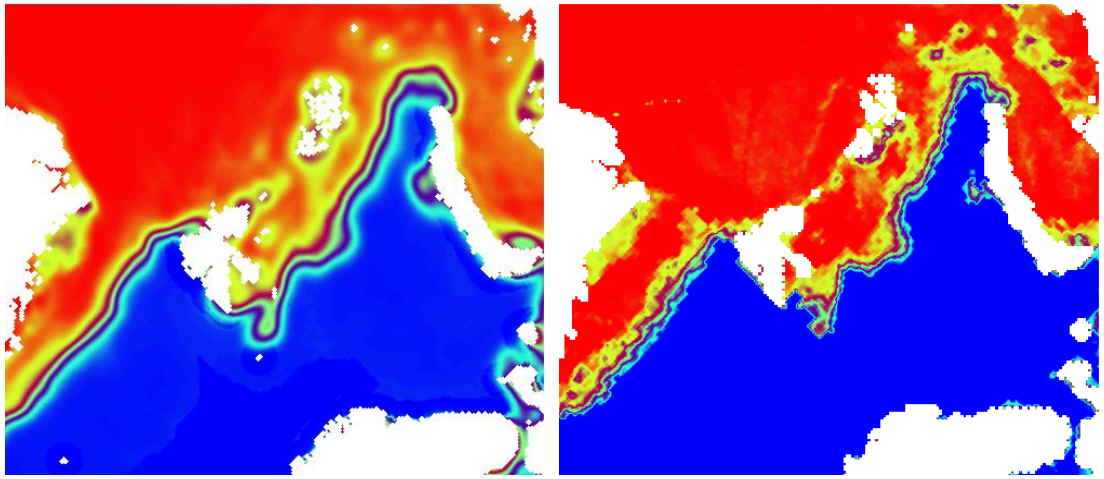


Transformations allow users to do comparisons of products and to extract tailored products for their specific need

Search results

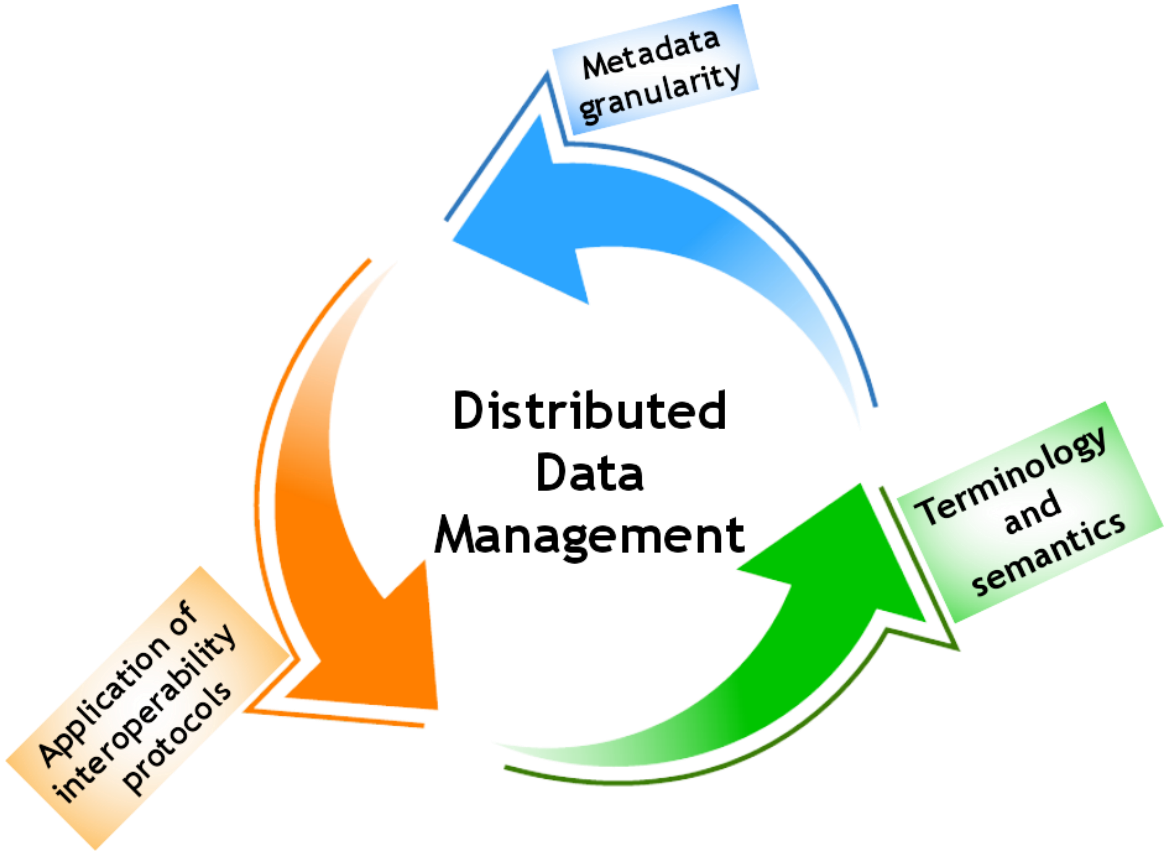


Transformation request



Bottlenecks

- Interoperability at the discovery level
  - Need to use standards in a “standardised” manner
- Interoperability at the data level
  - Need to engage data providers
  - Need tools helping them
- Application of controlled vocabularies and proper identification of which controlled vocabularies that are used
- Interaction with data providers in academia
  - Understanding emerging requirements from funding agencies on free and open data



Introduction on Data Management:

Policy, Best practices & Engagement with stakeholders / PART2

Halldór Jóhannsson (ARICE & Arctic Portal)






www.arice.eu

Arctic Research Icebreaker Cons.
ARICE

Grant agreement No 730965

ARICE

ARICE will improve the research icebreakers' services by working closely together with maritime industry on a so called "ships and platforms of opportunity" programme.

Through this programme, **commercial vessels operating in the Arctic Ocean will collect oceanic and atmospheric data** on their cruises.

Science and industry will work together to **explore new technologies, which can improve ship-based and autonomous measurements** in the Arctic Ocean.

ARICE will implement virtual and remote access of data via an innovative 3D Virtual Icebreaker, which will provide anyone with real-time information from the Arctic.

ARICE Objective

- **The objective is to establish the project data management system and develop and adapt strategies and tools for efficient data access and data dissemination.**

FAIR data

ARICE aims at its research data to be findable, accessible, interoperable and reusable (FAIR).

- Making data findable, including provisions for metadata
- Making data openly accessible
- Making data interoperable
- Increase data re-use (through clarifying licenses)

Tasks

Task 7.1 Data management requirements for curation, preservation and access to ARICE data

Task 7.2 Data Management System Design, Implementation and Certification

Task 7.3 Virtual access to data and dissemination via a **3D Virtual Icebreaker**

Scheme of the data management system

The Scheme of the data management system and its specifications has been developed and the report has been drafted.

We have created specifications for the server architecture, defined data products to be included in the system, and described the data flow between the data providers and final users.

The data management system – 3D Icebreaker

The main DMS software will run on the ARICE webserver.
Its main functions will be:

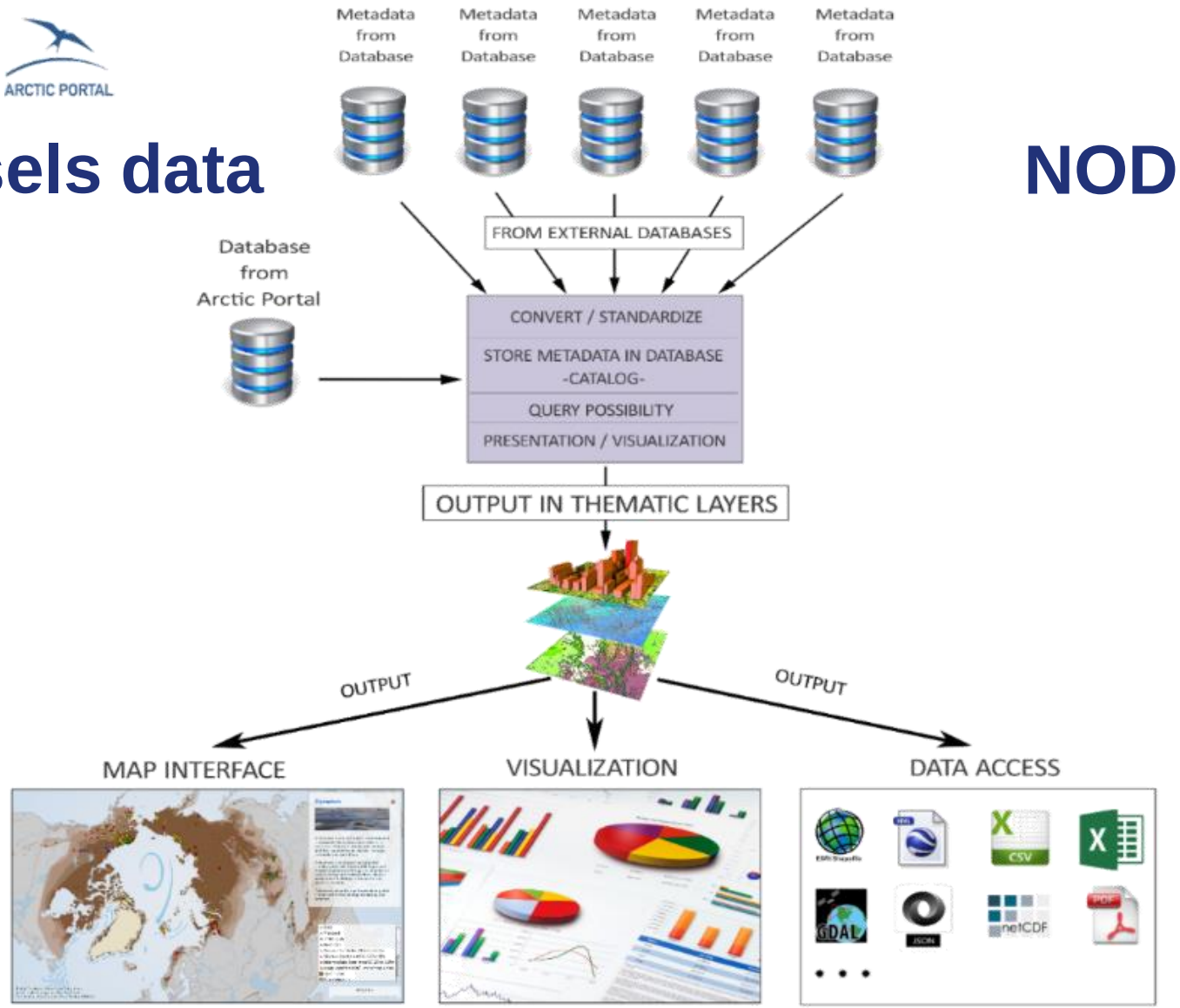
- to harvest and visualize data streamed / received from the icebreakers/vessels,
- to harvest and index NODCs metadata,
- to provide users with data search and conversion tools,
- to provide data through the 3D-Icebrakers interfacing tools,

For science, industry, other stakeholders and for education.

The data management system

Vessels data

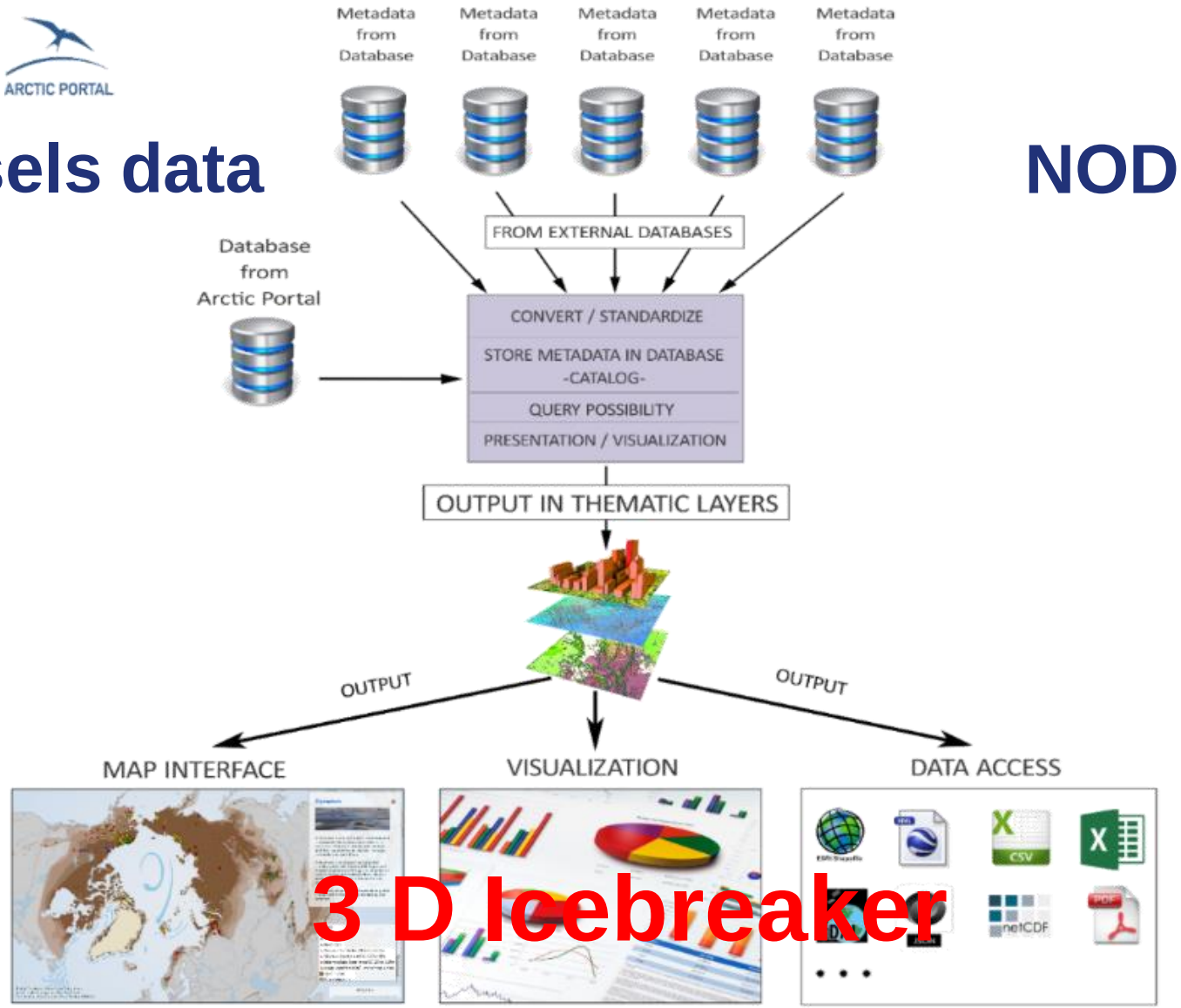
NODC data



The data management system

Vessels data

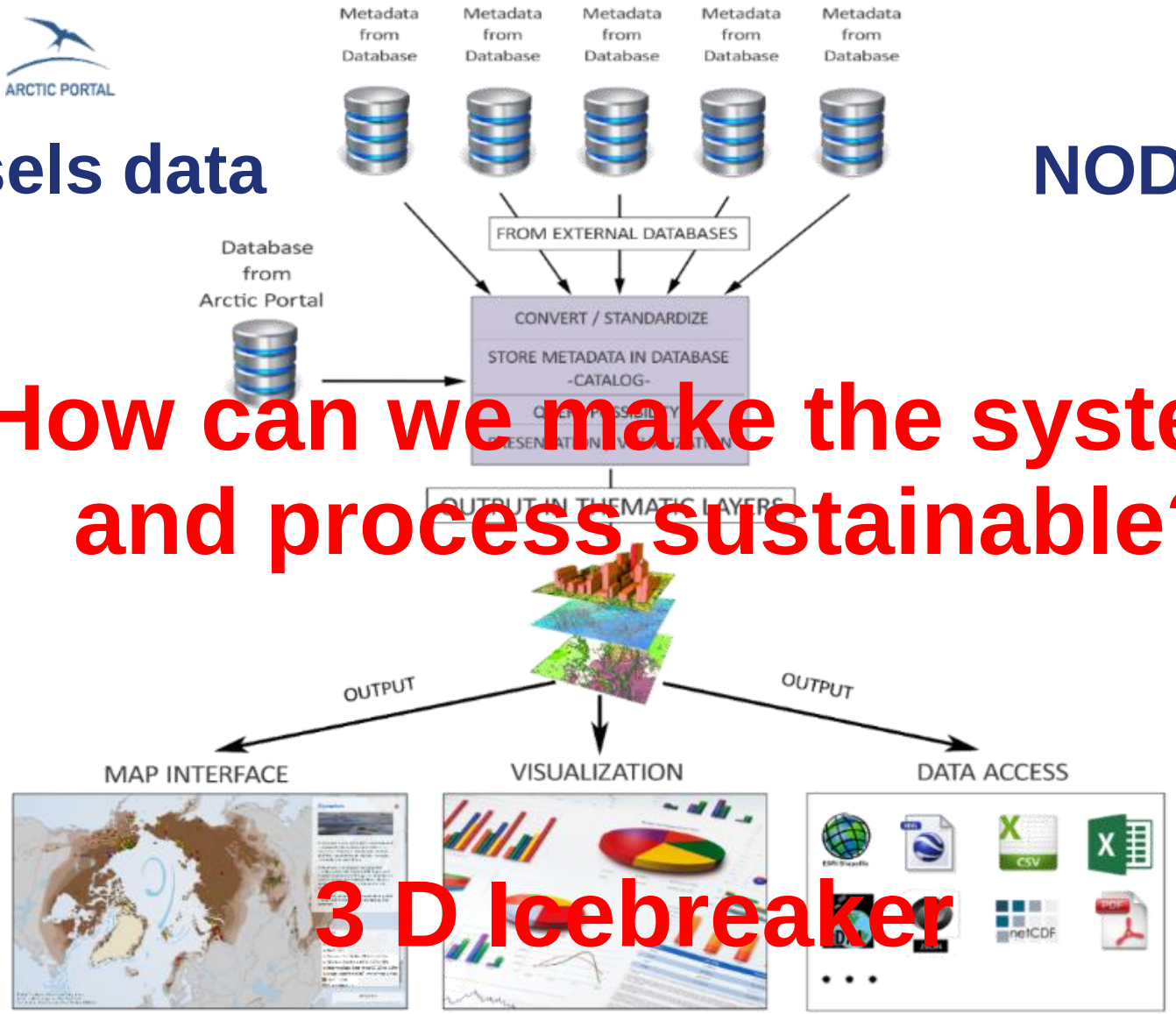
NODC data



The data management system

Vessels data

NODC data



How can we make the system and process sustainable?

3 D Icebreaker

ARICE Webinar
Data Management

Thank you very much!






Arctic Research Icebreaker Cons.
ARICE

An international collaboration strategy for meeting the needs of marine based research in the Arctic

Webinar recording will be available here

Home | About | News | Outreach | Training | Apply for Ship Time | Call Results | Intranet