

Distributed data management Challenges and potential solutions

Øystein Godøy



WMO OMM



Lessons learned

- International Polar Year
- WMO Global Cryosphere Watch
- Norwegian Satellite Earth Observation Database for Marine and Polar Research
- Norwegian Marine Data Centre
- Copernicus Marine Environmental Monitoring Service
- Earth System Grid Federation
- Norwegian Scientific Data Network



WMO OMM



GCW Generic Purpose

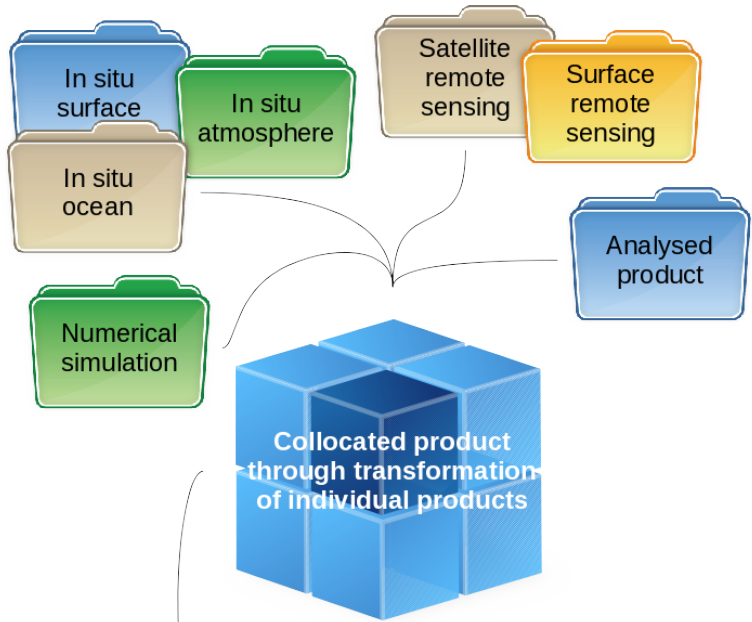
- To provide an overview of the datasets that are relevant for GCW
- To provide access to datasets
 - Real time data streams
 - Archive access
- Distributed Data Management
 - Metadata driven
 - Not hosting data
- To connect GCW with
 - WMO Information System
 - WIGOS

The screenshot shows the Global Cryosphere Watch (GCW) search results page. The page is titled "Search for GCW datasets" and features the WMO logo and "Global Cryosphere Watch" branding. The search results table is as follows:

Dataset name	Topics and variables	Contact (E-mail)	Abstract
cc-osisaf-nh	Cryosphere > Sea Ice > HIDDEN Cryosphere > Sea Ice > Sea Ice Concentration Cryosphere > Sea Ice > Sea Ice Edge Cryosphere > Sea Ice > Sea Ice Extent Geographical Region > Northern Hemisphere NORMET > Norwegian Meteorological Institute, Norway Not Available > Not Available > HIDDEN Oceans > Sea Ice > HIDDEN Oceans > Sea Ice > Sea Ice Concentration Oceans > Sea Ice > Sea Ice Edge Oceans > Sea Ice > Sea Ice	Not Available	Monthly sea ice concentration estimated from satellite data within the framework of EUMETSAT Ocean and Sea Ice SAF.

The detailed view of the dataset "cc-osisaf-nh" shows a map of the Arctic region and a list of topics and variables. The abstract states: "Monthly sea ice concentration estimated from satellite data within the framework of EUMETSAT Ocean and Sea Ice SAF."

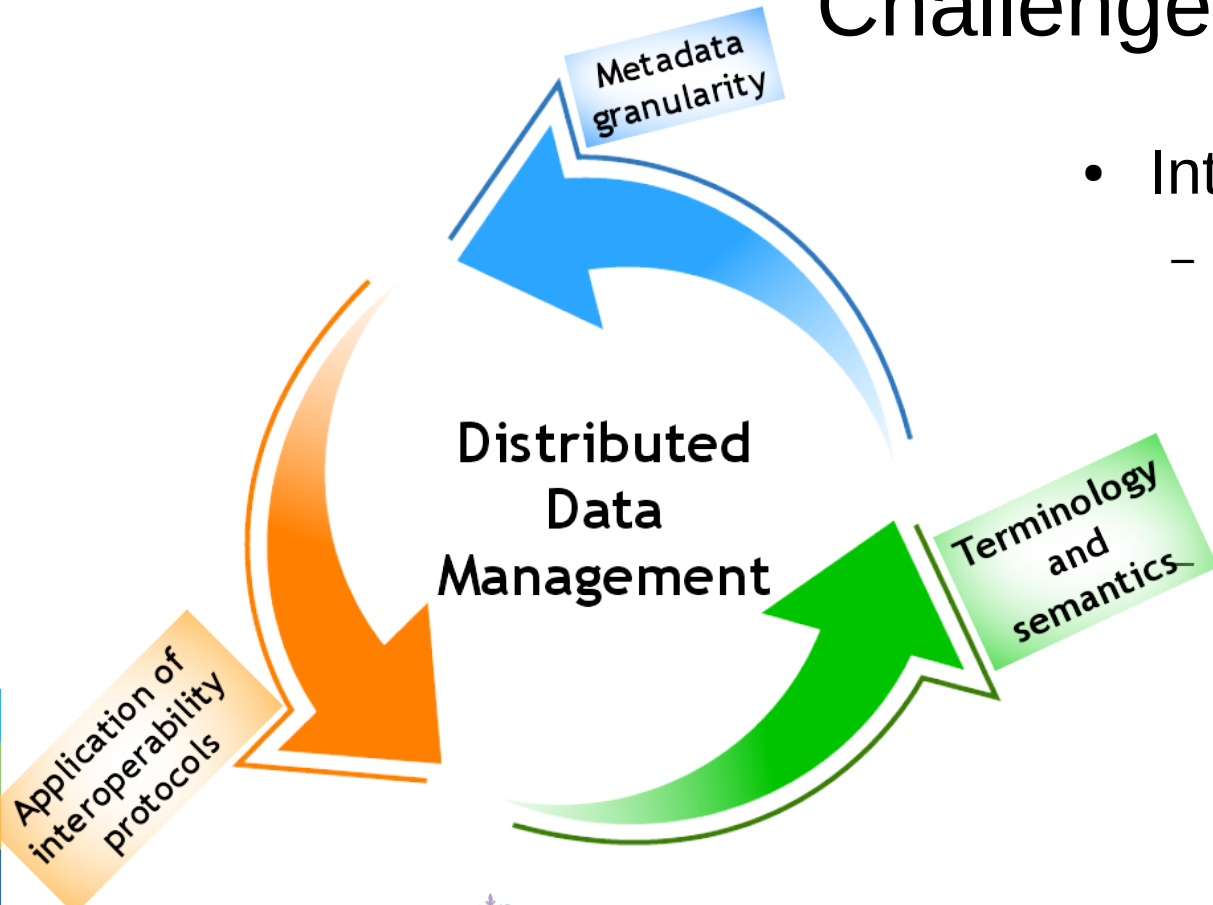
Approach



- Dataset oriented
 - Metadata driven
 - The WIS way at large
- Open data space
 - Higher order services offered when the data space can be constrained
- Net centric
 - Linkages with other data centres is vital
 - Implies brokering of metadata and data
 - Not too many
- Interdisciplinary
 - Dataset agnostic in the open data space



Challenges



- Interoperability

- Discovery Metadata

- Protocols (✓)
 - Structures (✓)
 - Semantics/terminology (-)

- Data documentation/exchange

- Protocols (✓)
 - Formats (-)
 - Semantics/terminology (-)
 - Common data model (-)

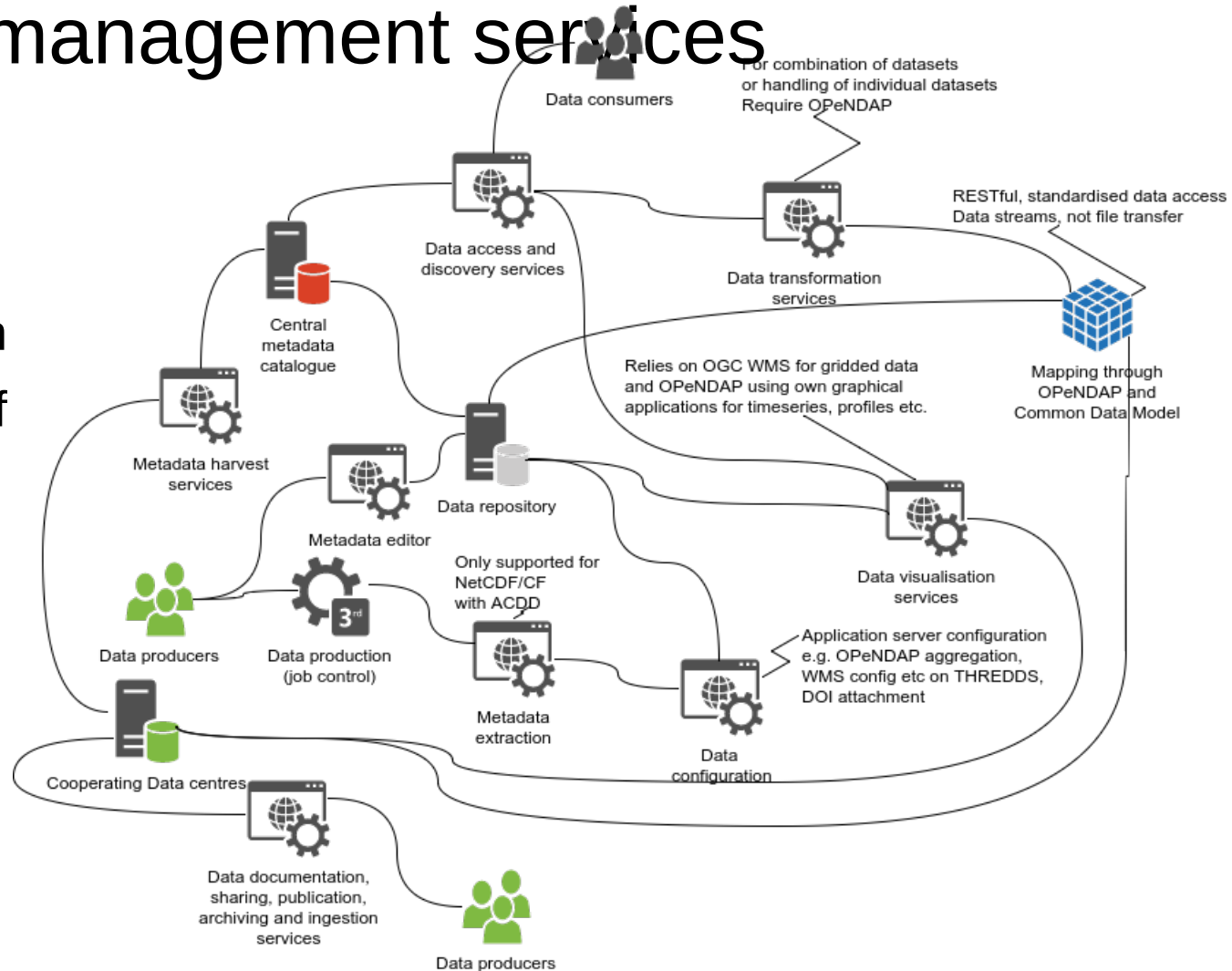


WMO OMM



Data management services

- Relies heavily on OPeNDAP
 - For data integration
 - Allows streaming of data
 - No difference between archive or real time data
 - Read what you need
 - No housekeeping

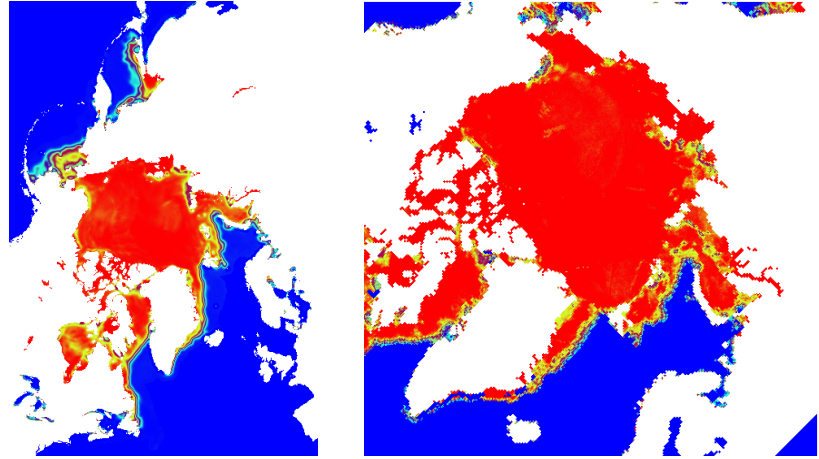


WMO OMM

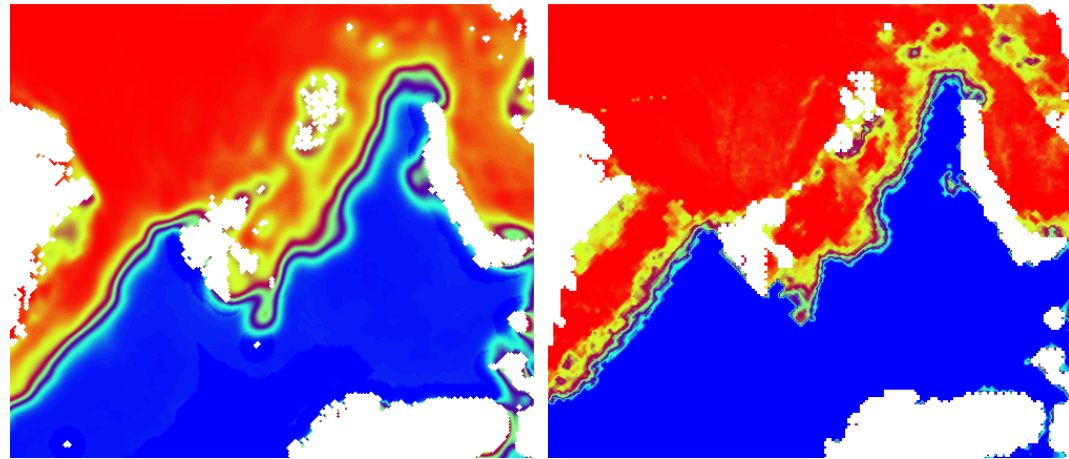


Transformations allow users to do comparisons of products and to extract tailored products for their specific need

Search results



Transformation request



WMO OMM



Summary

- While the acceptance of standardisation of discovery metadata is quite high
- the benefit of standardisation at the data level is far less understood
 - many communities lack an understanding for machine readable data exchange
- Agree on (discovery) metadata granularity
- Agree on controlled vocabularies and governance processes for these
 - Seems well established in GAW
- Agree on interoperability protocols for
 - Discovery metadata
 - Data
- Don't forget about interoperability at the data level
 - Archive data
 - Real time data
 - Why differentiate?



WMO OMM

