

GCW data management as a bridge between scientific communities and WMO Information System

Øystein Godøy



WMO OMM



Purpose of GCW Data Management

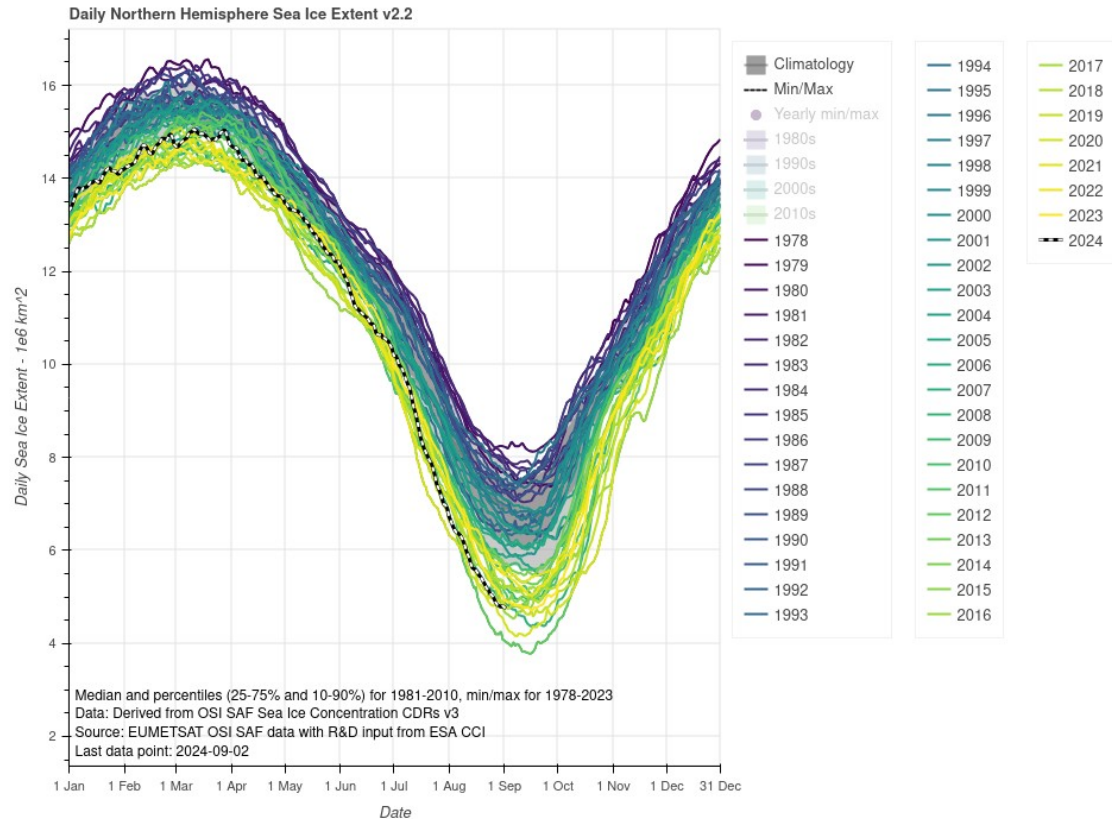
The screenshot shows the GCW Data Management web interface. At the top, there is a navigation bar with the WMO logo and 'World Meteorological Organization Global Cryosphere Watch' text, along with a 'Log in' link. Below this is a yellow navigation bar with links for 'GCW home page', 'Data Access', 'Support', and 'About the Data Portal'. The main content area is divided into several sections:

- Search:** A search bar with a dropdown menu for 'Contains all of these words' and a text input field for 'Enter your search here'. Below it are 'Start Date' and 'End Date' fields with date pickers.
- Filters:** A 'Has children' checkbox and 'Advanced options' section. A 'Dataset Level' section with a 'Parent (20604)' checkbox. An 'Iso Topic Category' list including 'climatology/Meteorology/Atmosphere (1213)', 'geoscientificInformation (203)', 'imagery/BaseMaps/EarthCover (193)', 'oceans (72)', 'elevation (50)', 'environment (21)', 'inlandWaters (15)', 'society (7)', 'biota (4)', and 'boundaries (3)'. A 'Keywords' section with a list of terms like 'EARTH SCIENCE > Cryosphere > Glaciers/Ice Sheets > Glacier Thickness Change (from geodetic method) (13221)'.
- Map:** A map of Greenland with a blue bounding box filter. A tooltip shows coordinates 'lon: 8.16, lat: 81.65'. Below the map, it says '20604 datasets found. Showing datasets 1 - 15 on page 1 of 1374 pages.' and 'Ice drift in Greenland seas'.
- Project/Collection/Personnel/Organisation:** Lists of related entities with checkboxes and counts. For example, 'Project' includes APPLICATE (450), SICOS (450), YOPP (450), OIB (109), and SMAP (102). 'Collection' includes ADC (20604), GCW (20604), SICOS (504), JARPL (452), and YOPP (452). 'Personnel' includes NSIDC User Services (1158), Øystein Godoy (457), FGDC User Services (146), NOAA User Services (102), and Thomas Jackson (90). 'Organisation' includes World Glacier Monitoring Service, University of Zurich (16752), Norwegian Meteorological Institute (459), Norwegian Meteorological Institute / Arctic Data Centre (32), Institute of Geophysics (24), and Norwegian Polar Institute (15).
- Data Center/Publisher:** Lists of data centers and publishers. 'Data Center' includes NASA NSIDC DAAC (832), WMO (450), FGDC (146), NOAA (102), and AGDC (72). 'Publisher' includes NASA NSIDC DAAC: National Snow and Ice Data Center Distributed Active Archive Center (792) and NSIDC: National Snow and Ice Data Center (446).
- Temporal Extent:** A form with 'Start date: 2005-05-10T10:33:00Z' and 'End date: 2006-02-10T23:31:00Z'.
- Data access:** A section with 'Data access:' and buttons for 'Show extended metadata', 'Child data [1,432]', and 'Export Metadata'.
- License:** A Creative Commons license icon and 'Access: Open'.

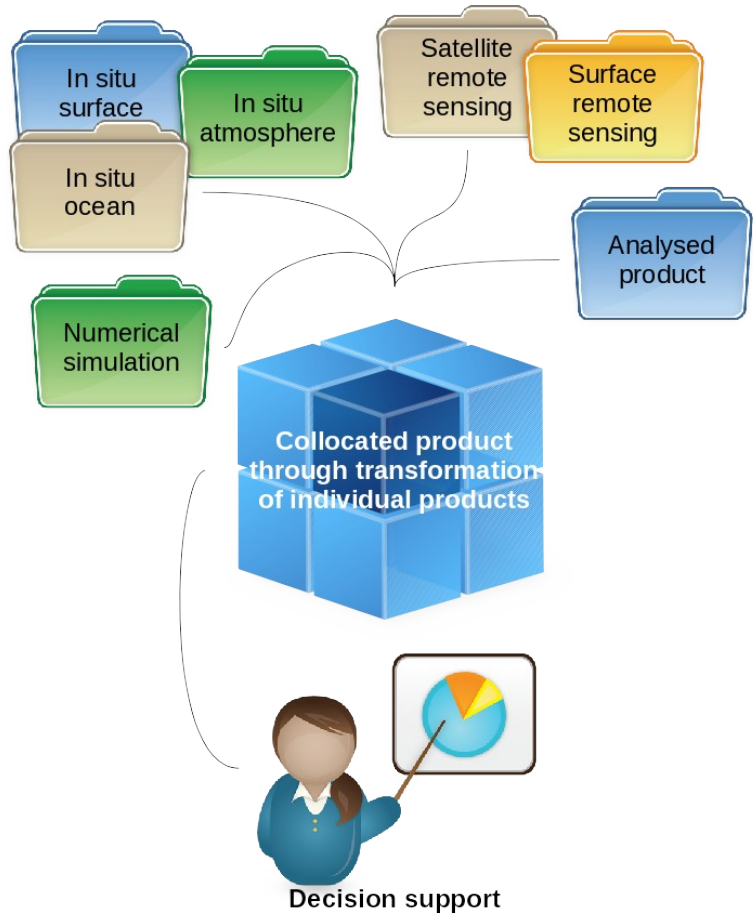
- To provide an overview of the datasets that are relevant for GCW
 - Authoritative, actionable, and accessible science-based information
 - For development of value added analysis and indicators
- To provide access to datasets
 - Real time data streams
 - Access to archived data
- To connect GCW stations with
 - WMO Information System
 - WIGOS

Heterogeneous data

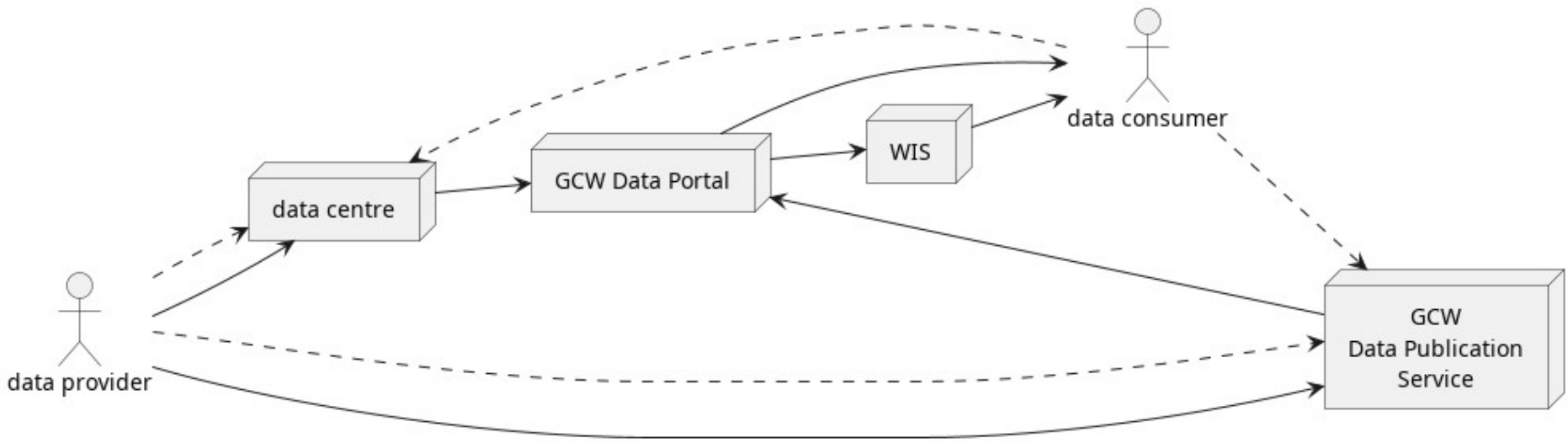
- Types of data
 - In situ observations
 - Weather data
 - Mass balance data
 - Surface irradiance data
 - ...
 - Remote sensing products
 - Numerical simulations
- Generic types of data
 - Gridded
 - Time-series
 - Profiles
 - ...



GCW Data Portal Approach

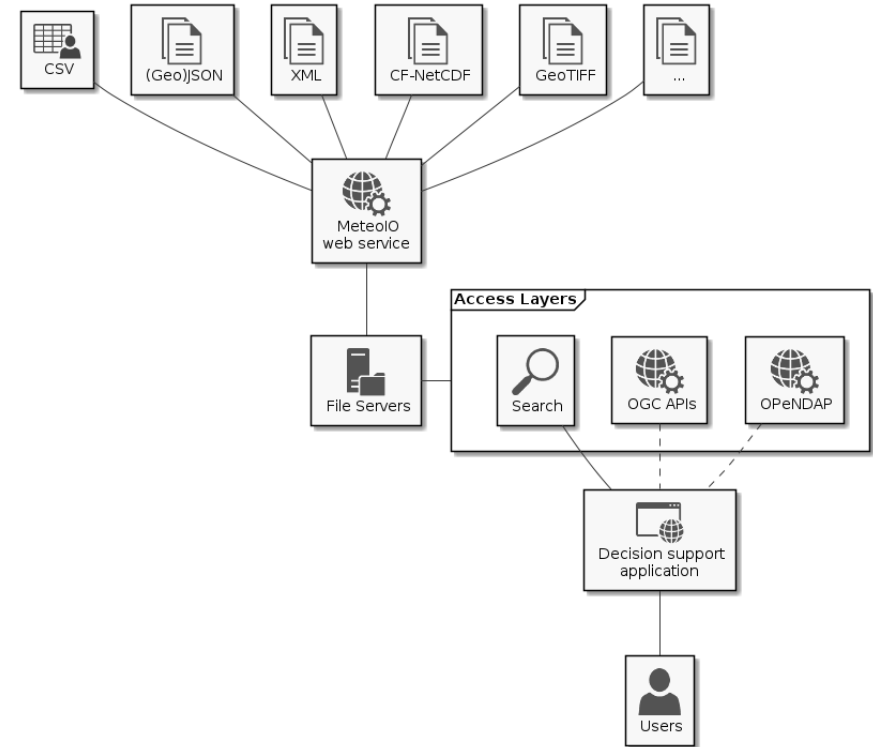


- Partnership based
 - Based on the WMO Unified Data Policy
- Dataset oriented
 - Driven by discovery metadata
- Open data space
 - Higher order services offered when the data space can be constrained
- Net centric
 - Linkages with data centres is vital
 - Relying on interoperability standards and semantic annotations
 - CF-NetCDF and standardisation efforts within WMO
 - Implies brokering of discovery metadata and data (including use metadata)
 - But need structured data to be served
- Interdisciplinary
 - Dataset agnostic in the open data space



Heterogeneous community

- Types of data centres
 - National Meteorological and Hydrological Services
 - Universities
 - Independent research institutions
 - (Industry)
- Varying degree of structured data management
- Not necessarily sharing the same objective
- Varying degree of interoperability for
 - metadata
 - data
- Mutual benefit of standardisation

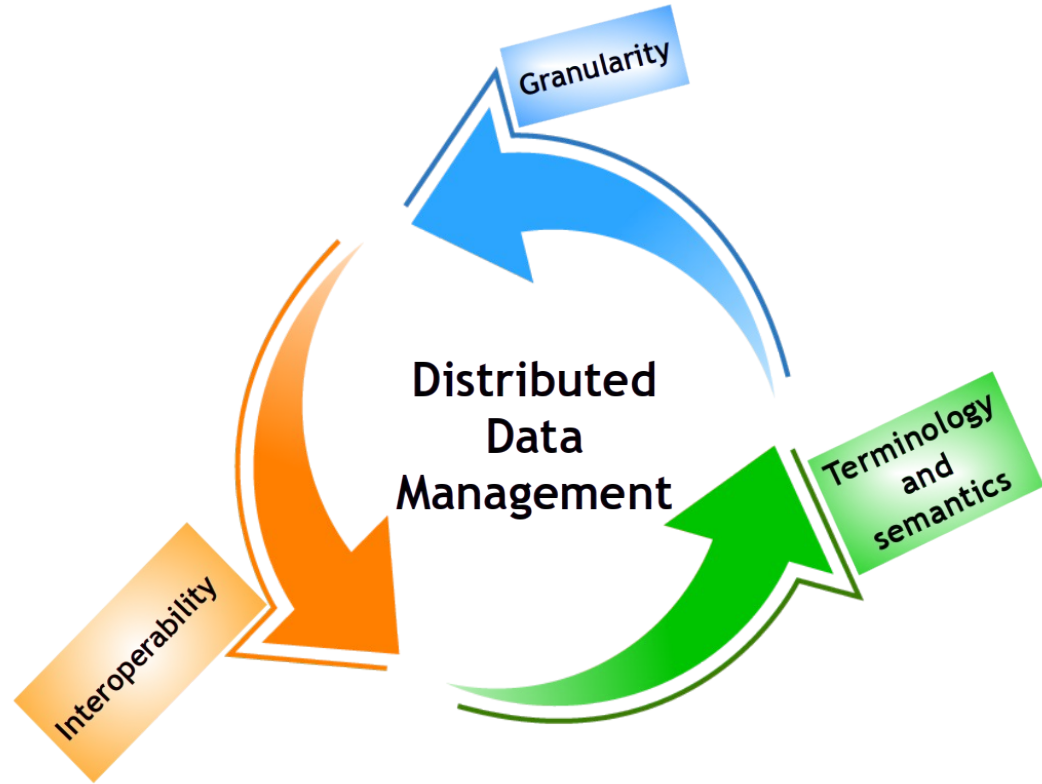


The FAIR guiding principles

- 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

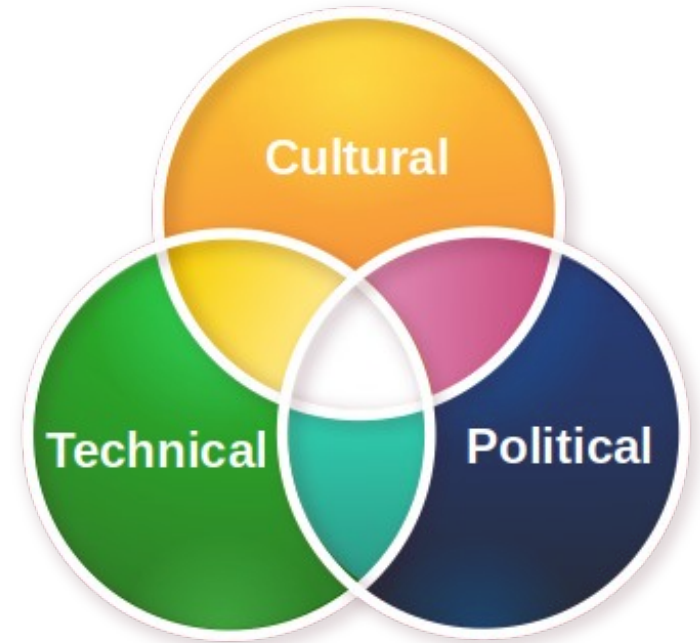
Challenges

- Machine Interoperability
 - Discovery Metadata
 - Protocols (✓)
 - Structures (✓)
 - Semantics/terminology (✓)
 - Data
 - Protocols (✓)
 - Formats (✓)
 - Semantics/terminology (✓)
 - Common data model (✓)



Summary

- Need to consolidate on data documentation and sharing approaches and not constantly explore new opportunities
 - Integration of new technologies comes at a substantial cost and cause technological debt
- Think beyond your own use case when publishing data
 - Granularity on datasets
 - Reduced latency
 - Cal/Val require consolidated datasets
- Data exchange is as good as we make it



WMO OMM

