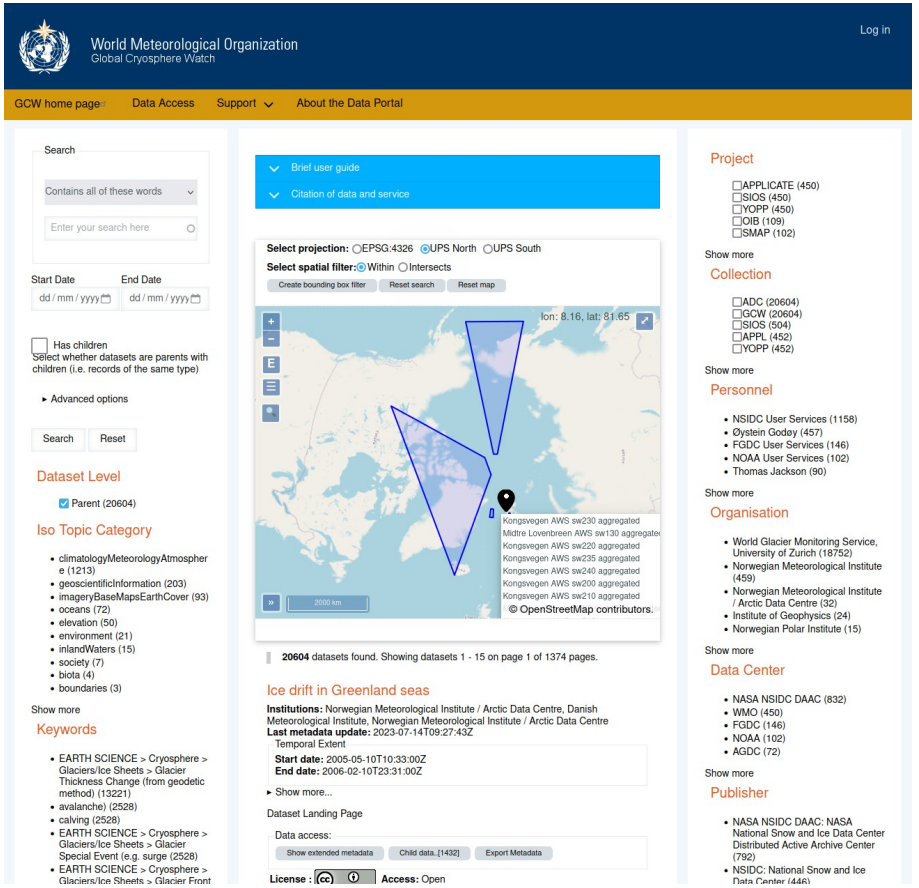
GCW data management as a bridge between scientific communities and WMO Information System

Øystein Godøy

WMO OMM

GCW Global Cryosphere Watch

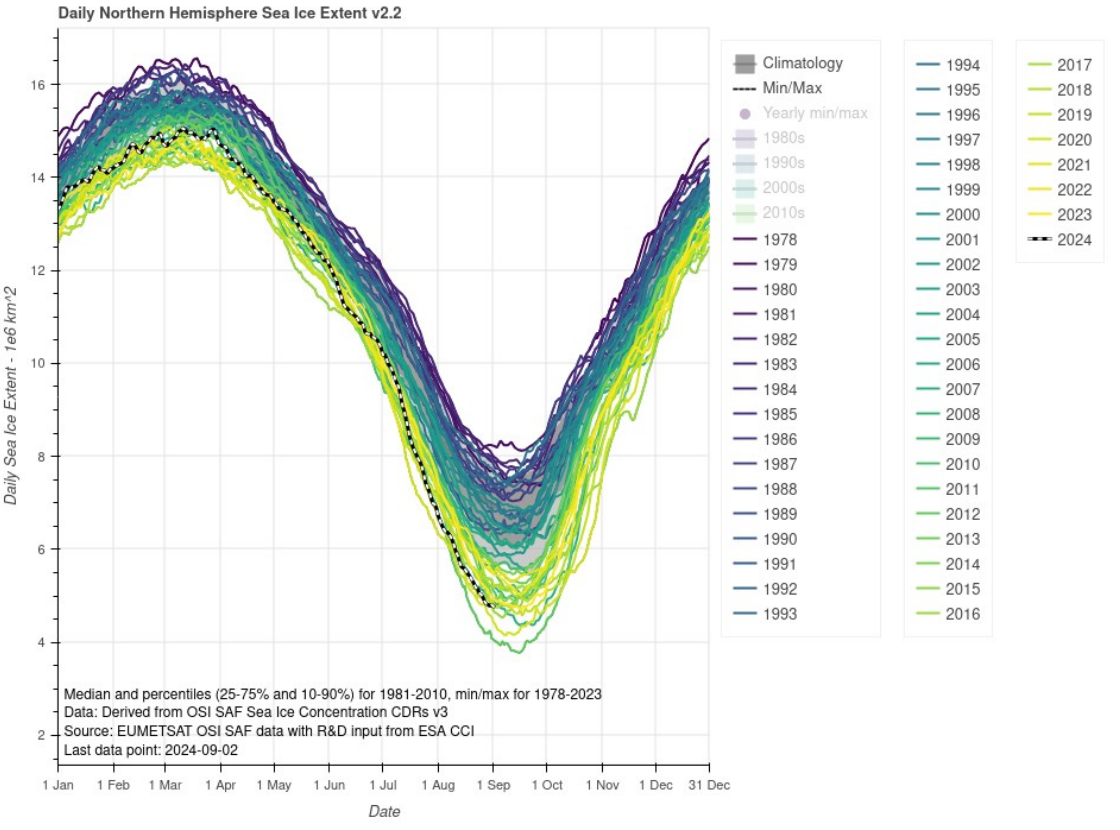
Purpose of GCW Data Management



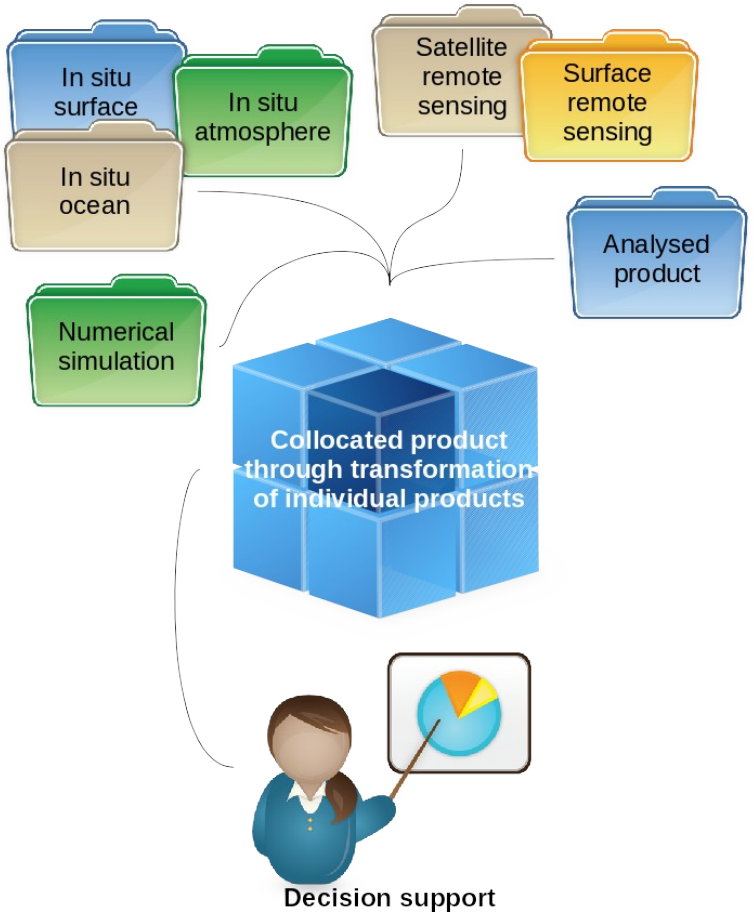
- To provide an overview of the datasets that are relevant for GCW
  - Authoritative, actionable, and accessible science-based information
  - For development of value added analysis and indicators
- To provide access to datasets
  - Real time data streams
  - Access to archived data
- To connect GCW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

Heterogeneous data

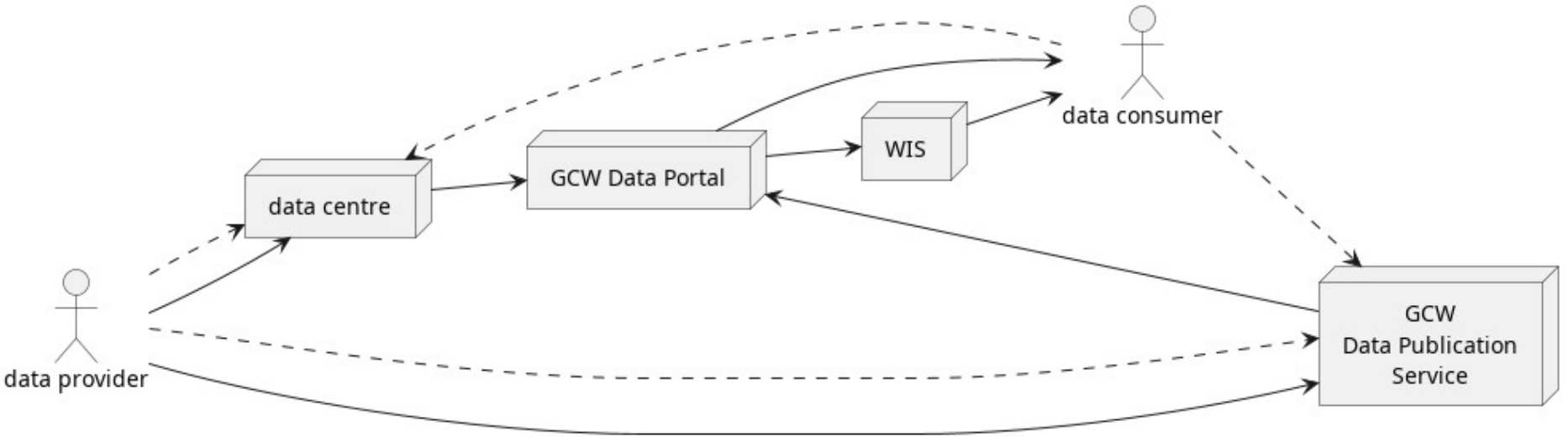
- Types of data
  - In situ observations
    - Weather data
    - Mass balance data
    - Surface irradiance data
    - ...
  - Remote sensing products
  - Numerical simulations
- Generic types of data
  - Gridded
  - Time-series
  - Profiles
  - …



GCW Data Portal Approach

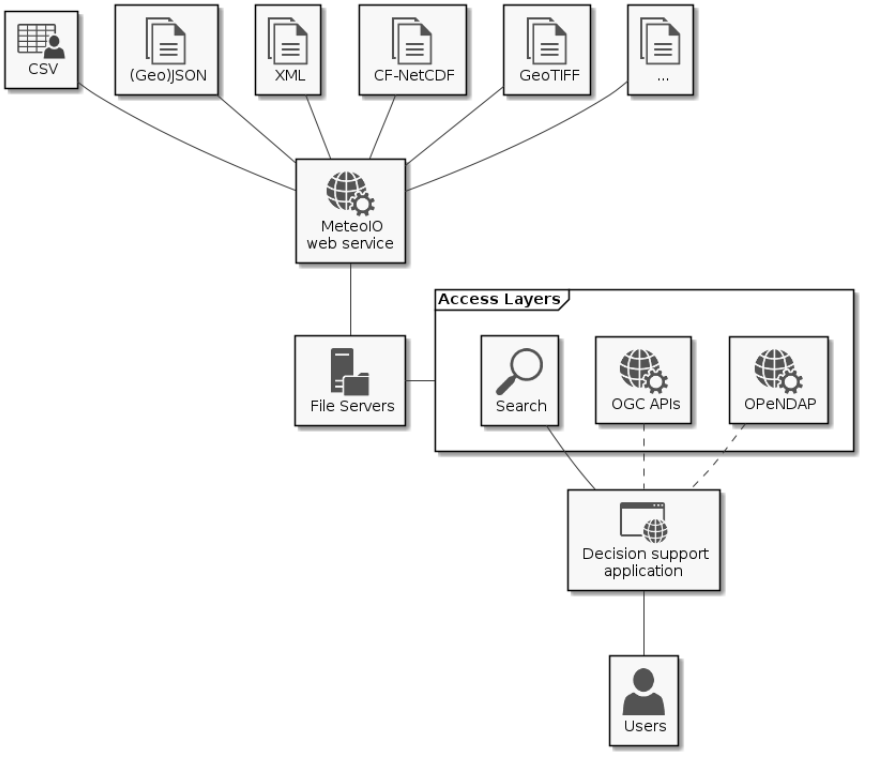


- Partnership based
  - Based on the WMO Unified Data Policy
- Dataset oriented
  - Driven by discovery metadata
- Open data space
  - Higher order services offered when the data space can be constrained
- Net centric
  - Linkages with data centres is vital
    - Relying on interoperability standards 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


data consumer
WIS
GCW Data Portal
data centre
data provider
GCW
Data Publication
Service

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



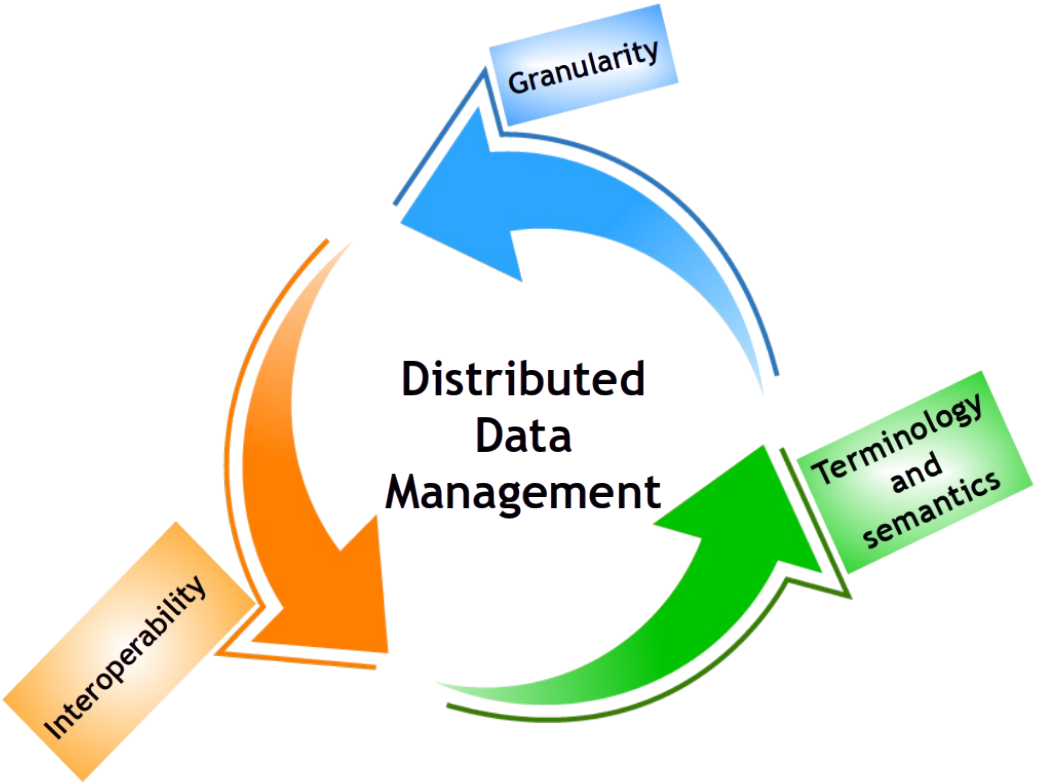
The FAIR guiding principles

- 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

Wilkinson et al. 2016

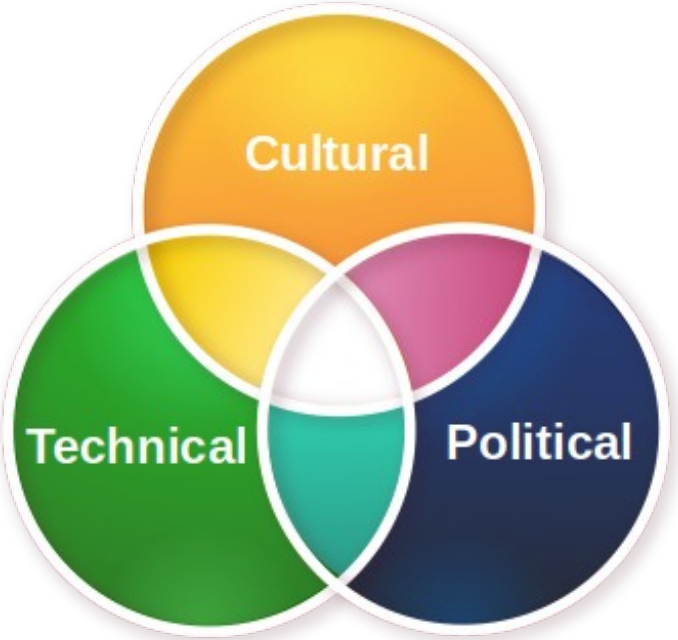
Challenges

- 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 GCW Global Cryosphere Watch