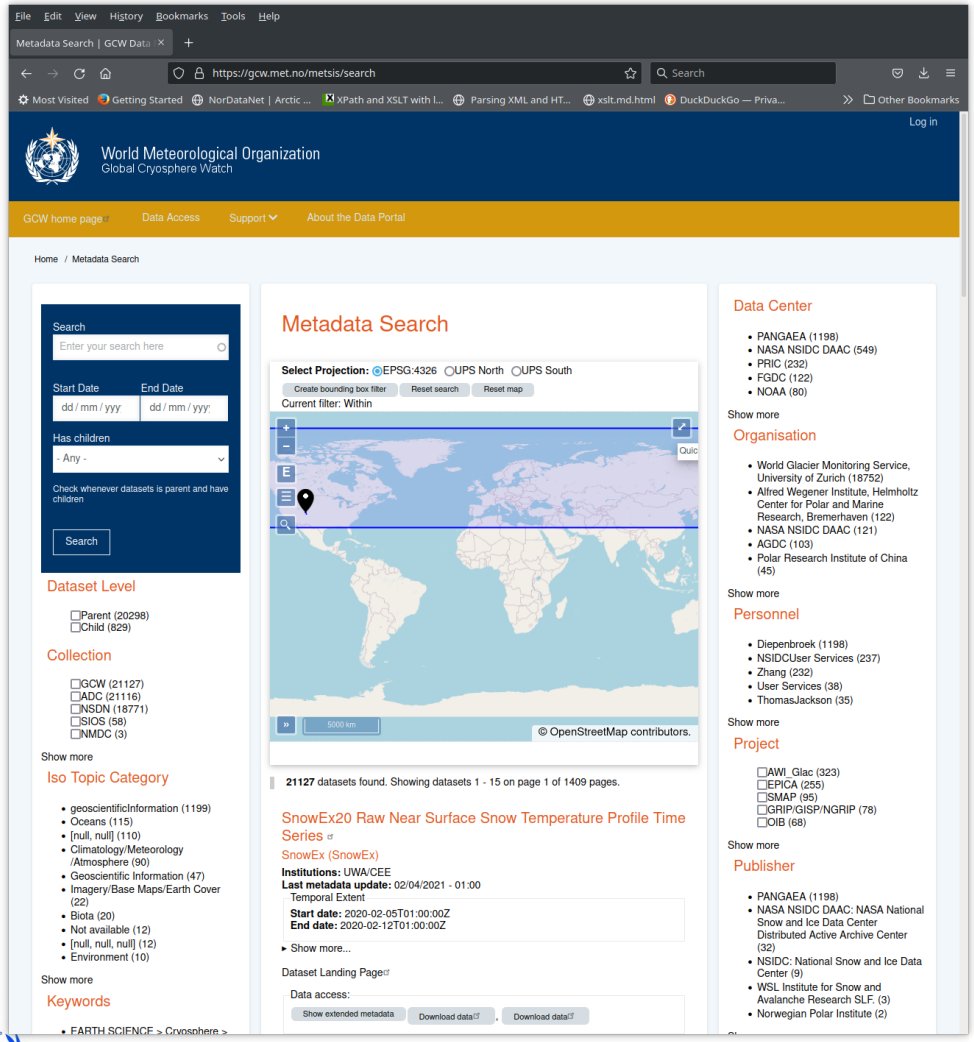
WEATHER CLIMATE WATER
TEMPS CLIMAT EAU

Open Access to GCW data

Øystein Godøy (Norwegian Meteorological Institute)
Mathias Bavay (WSL/SLF)
…

WMO OMM
World Meteorological Organization
Organisation météorologique mondiale

Purpose of GCW Data Management

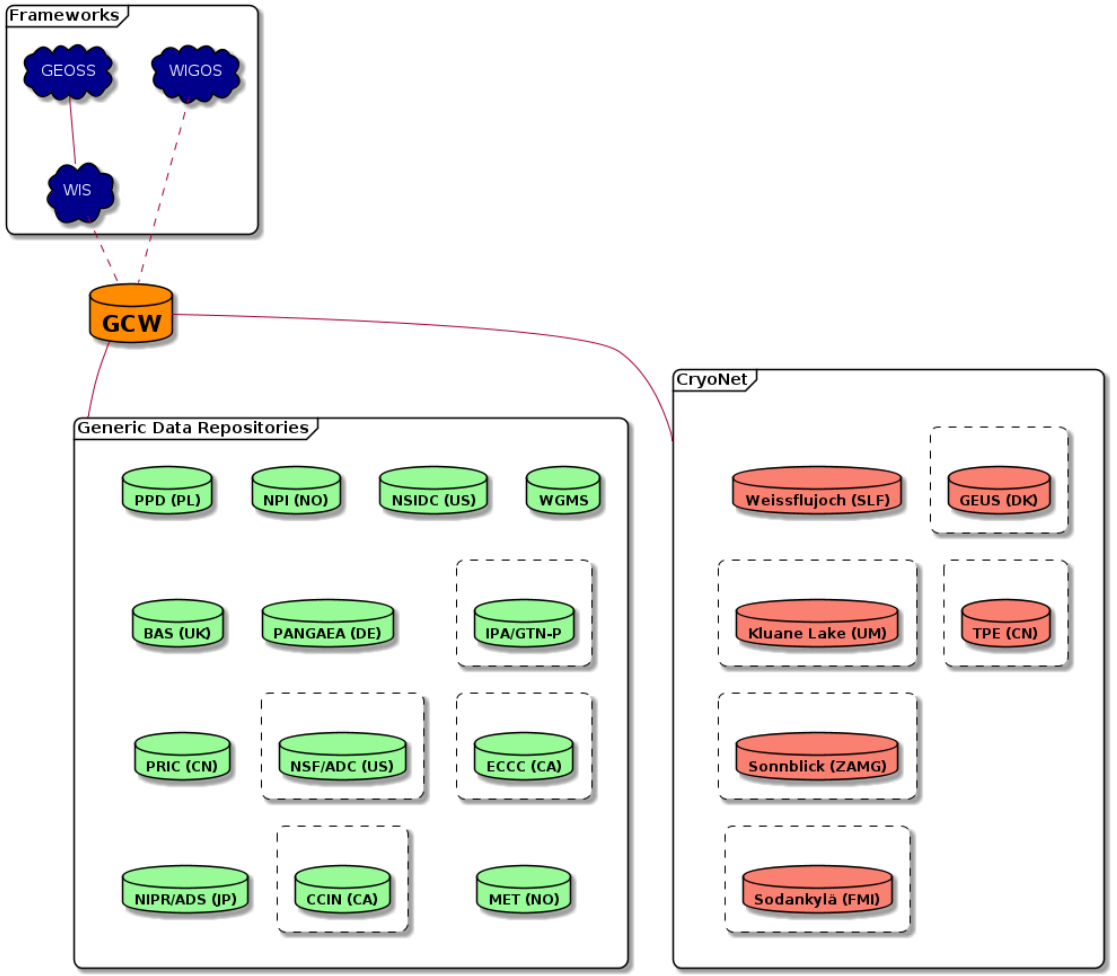


- To provide an overview of the datasets that are relevant for GCW
- To provide access to datasets
  - Real time data streams
  - Access to archived data
- Distributed Data Management
  - Metadata driven
  - Not hosting data
- To connect GCW stations with
  - WMO Information System
  - WIGOS

WMO OMM

Heterogeneous community

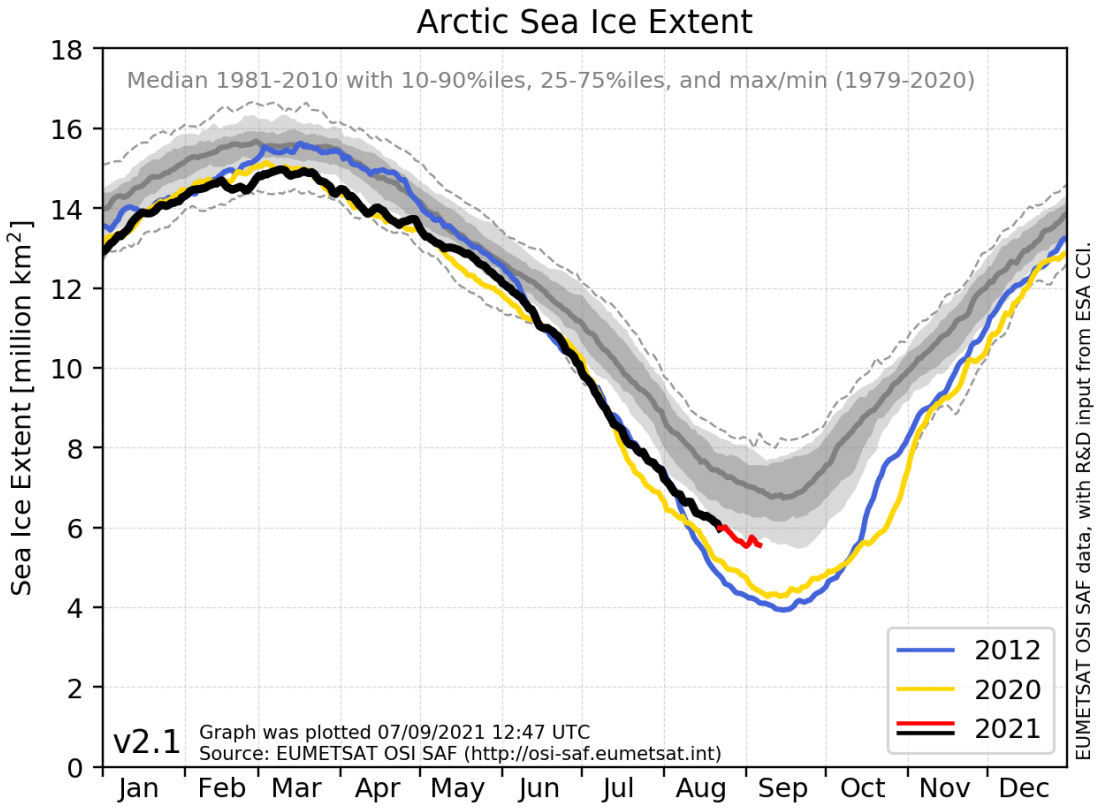
- Types of data centres
  - National Meteorological and Hydrological Services
  - Universities
  - Independent research institutions
- Varying degree of structured data management
- Not necessarily sharing the same objective
- Varying degree of interoperability for
  - metadata
  - data
- Mutual benefit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



WMO OMM

Heterogeneous data

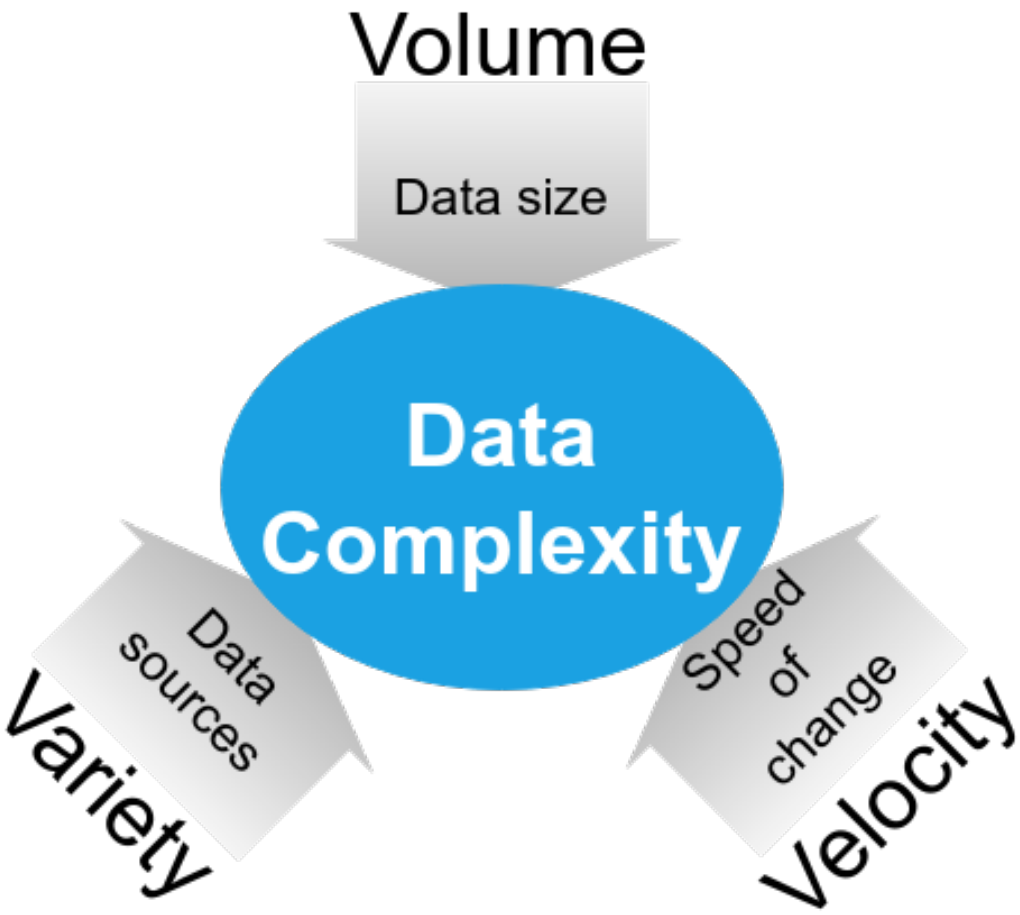
- Types of data
  - In situ observations
    - Weather data
    - Mass balance data
    - Surface irradiance data
    - ...
  - Remote sensing products
  - Numerical simulations
- Generic types of data
  - Gridded
  - Time-series
  - Profiles
  - ...



WMO OMM

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
  - …



WMO OMM

Project objectives

- To facilitate the access to available datasets from different institutions and projects, by bridging between scientific communities and WMO systems in support of the WMO activities (e.g. WMO operating plan).
- Improving interoperability of WMO GCW relevant datasets.
- Increasing the amount of data available to support cryosphere related goals of WMO, as delivered by GCW.
- Wherever possible efficiently link between WIGOS and WIS metadata.

WMO OMM

Project team

- Norway - MET
  - Øystein Godøy, Lara Ferrighi, Magnar Martinsen, Massimo Di Stefano, Elodie Fernandez
- Switzerland - WSL/SLF
  - Mathias Bavay, Charles Fierz, Joel Fiddes

Project plan - Deliverables

No.	Deliverable name	Lead	Del. date	Status
D1	Updated information model enabling linkages on dataset to WIGOS metadata	MET	2020Q2	Complete
D2	Dynamic visualisation of time series from NetCDF-CF and OPeNDAP	MET	2020Q2	Complete
D3	Updated harvesting of discovery metadata supporting OAI-PMH, OGC CSW and OpenSearch	MET	2021Q2	In progress
D4	NetCDF-CF guidelines for timeseries and profiles (e.g. permafrost)	MET	2021Q3	In progress
D5	Mapping harvested discovery metadata to WMO Core Profile	MET	2021Q3	In progress
D6	Extension of metadata harvesting to support schema.org provided current ESIP activities are approved by schema.org	MET	2022Q3	Pending funding
D7	Conversion of NetCDF-CF to WMO BUFR for permafrost profiles	MET	2023Q2	Funding through EU
D8	Web service converting non standardised data to NetCDF-CF using MeteoIO	WSL/SLF	2023Q4	Funding through EU

WMO OMM

Project plan - Milestones

No.	Milestone name	Lead	Due	Status
M1	New information model implemented	MET	2021Q1	Done
M2	Selected permafrost datasets available online and in real time	MET	2021Q4	In progress
M3	Harvested discovery metadata exposed through WIS	MET	2022Q1	In progress
M4	Transformation of NetCDF-CF to WMO BUFR for selected datasets	MET	2023	In planning

WMO OMM

WIS 2 Principles in the project

All applied principles – *Highlights in green*

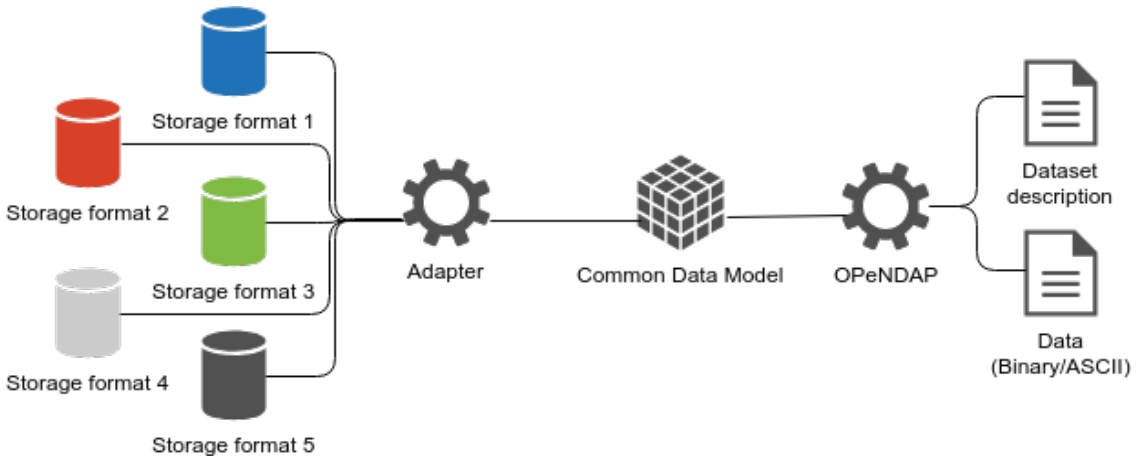
- Principle 1
- Principle 2
- Principle 3
- Principle 4
- Principle 5
- Principle 6
- Principle 7
- Principle 8
- Principle 9
- Principle 10
- Principle 11

WMO OMM

WIS 2 Principles in the project

- Principle 1: WIS 2.0: adopts Web technologies and leverages industry best practices and open standards.
  - GCW data management is based on harvesting discovery metadata through standardised web services for such exchange.
    - The information exchanged is standardised according to ISO19115 or GCMD DIF (currently) and data are encouraged to be served as NetCDF according to the Climate and Forecast convention. This links directly to a Service Oriented Approach relying on Semantic Web and Linked Data.
    - Specific standards used include OAI-PMH, OGC CSW, OpenSearch, OGC WMS, OPeNDAP, OGC WPS, SKOS/OWL, ISO19115, (SPARQL)

WIS 2 Principles in the project



- Principle 5: WIS 2.0: encourages NCs and DCPCs to provide 'data reduction' services via WIS that process 'big data' to create results or products that are small enough to be conveniently downloaded and used by those with minimal technical infrastructure.
  - GCW offers transformation services on top of data that are served according to the CF convention through OPeNDAP. These transformation services allows users (or applications to subset data in time, space or parameter space).
  - Data services rely heavily on OPeNDAP allowing easy subsetting of information on server side.
  - No need to download data
  - Services are running in a private cloud environment, but is configurable. Choice of solution is a cost / benefit analysis.

WMO OMM

WIS 2 Principles in the project

- Principle 8: WIS 2.0: will adopt direct data-exchange between provider and consumer.
  - GCW is currently considering the data provider and the host data centre of the data as the authoritative source for data. The direct access to a dataset is done by forwarding the data consumer to the web services offered by the host data centre.
  - The only exception to this in the current implementation is when higher order services offered in the GCW Data Portal are used to modify or combine data prior to data delivery.

WIS 2 Principles in the project

- Principle 10: WIS 2.0: will provide a catalogue containing metadata that describes both data and the service(s) provided to access that data.
  - GCW maintains its own catalogue with discovery metadata, but holds currently no catalogue for web services.
  - Web services are identified through the dataset metadata relying on the principle of data being the important piece of information, services are mechanisms conveying the information.
  - Cataloguing the data discovery web services will be done, but not planned in the short term.
  - Integrating the existing GCW services with WIS 2.0 catalogue will be preferable. Currently the main effort of GCW is to ensure good enough quality on the discovery and use metadata supplied by contributors and transform this into WIS compliant information.

WIS 2 Principles in the project

- Principle 11: WIS 2.0: encourages data providers to publish metadata describing their data and Web services in a way that can be indexed by commercial search engines.
  - GCW is reimplementing web services offering discovery metadata and will in this context support OAI-PMH, OGC CSW and OpenSearch.
  - Details are still under discussion as well as how to ensure integrity in the value chain between the originating data centre and the higher order catalogues like WIS (to avoid duplicated of records).
  - GCW is also working with the ESIP community on extensions that will make schema.org useful for dataset discovery. The current definition of schema.org is insufficient for proper dataset discovery and filtering of information, but promising extensions are being discussed and the community working on this has good momentum.

WIS 2 Principles in the project

- Further applied principles
  - Principle 2: Uniform Resource Locators (URL)
    - Discovery metadata are harvested from contributing data centres using URLs. The discovery metadata harvested contains URLs for data access, licence information as well as interpretation of semantic annotation (on scientific parameters or purpose of a URL).
  - Principle 3: Public telecommunications networks
    - The backbone for all communication within GCW data management is the Internet. For specific purposes GCW will connect with private networks (e.g. WMO GTS).
  - Principle 4: Web service(s)
    - GCW data management relies on web services for information and data exchange and for high order services. This information follows the datasets.

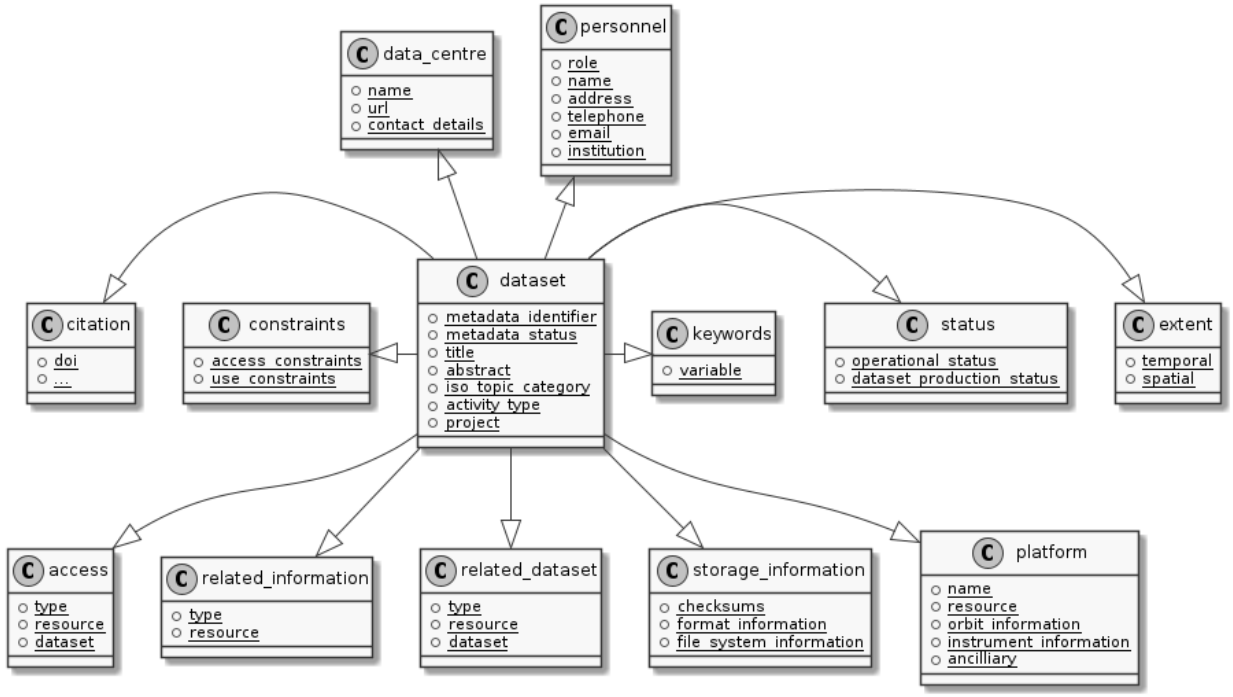
WIS 2 Principles in the project

- Further applied principles
  - Principle 6: Open standard messaging protocols
    - GCW does not currently have messaging protocol. Linkages towards EUMETCast/GEONETCast have been explored, implementation relies on user communities and demand.
  - Principle 7: Cache/store messages
    - GCW is currently not caching data nor distributing messages. Real time data are available as data streams using OPeNDAP. Specific message distribution will rely on the WMO message passing system.
  - Principle 9: Routing tables and bulletin headers
    - GCW core functionality relies on web services. GCW is currently not using WMO GTS for data transmission. The critical part is to connect efficiently to relevant WMO services.

Project data standards

- Data formats
  - CF-NetCDF with ACDD elements
    - For timeseries, profiles, trajectories
    - Under consideration for geometries
    - Relying on HDF5 as the storage layer
    - Has started to explore Zarr but not highest priority
  - WMO GRIB
  - WMO BUFR
- Application of the Attribute Convention for Dataset Discovery (ACDD) within NetCDF allows generation of discovery metadata for remote datasets if served through OPeNDAP and with THREDDS Catalogues
- GCW relies on UNIDATA's Common Data Model for all transformation services on top of data
  - Currently CDM, testing CDM2 for specific purposes

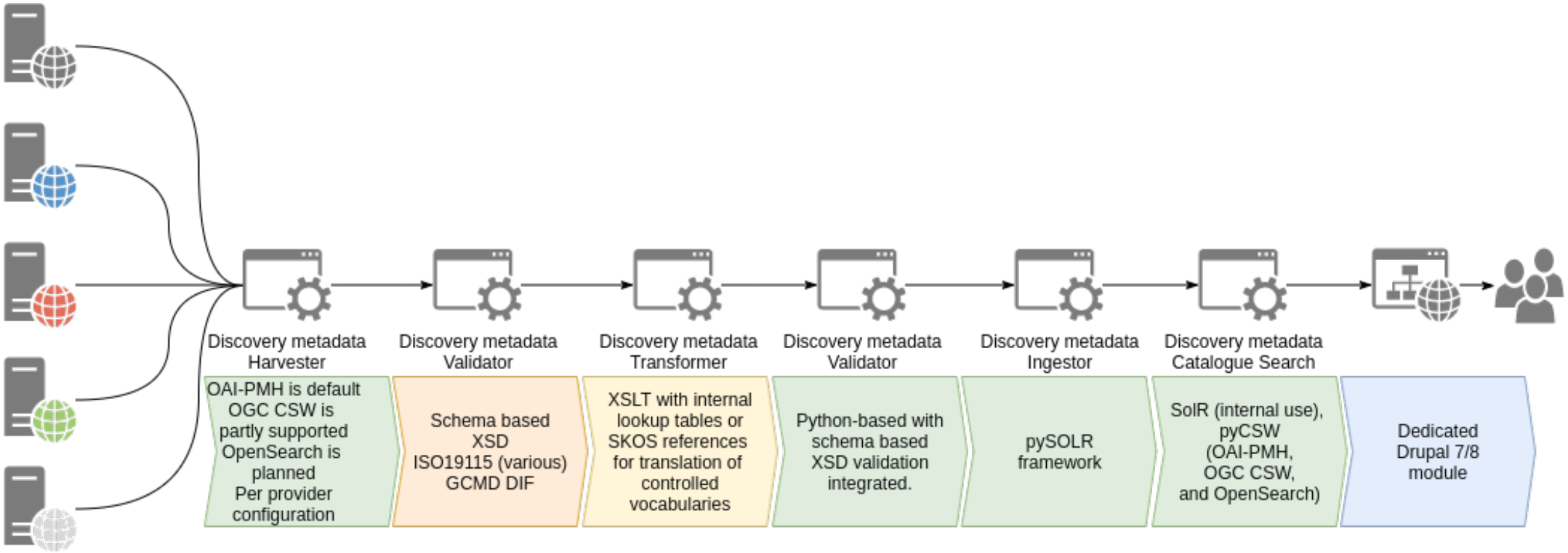
Project discovery metadata standards



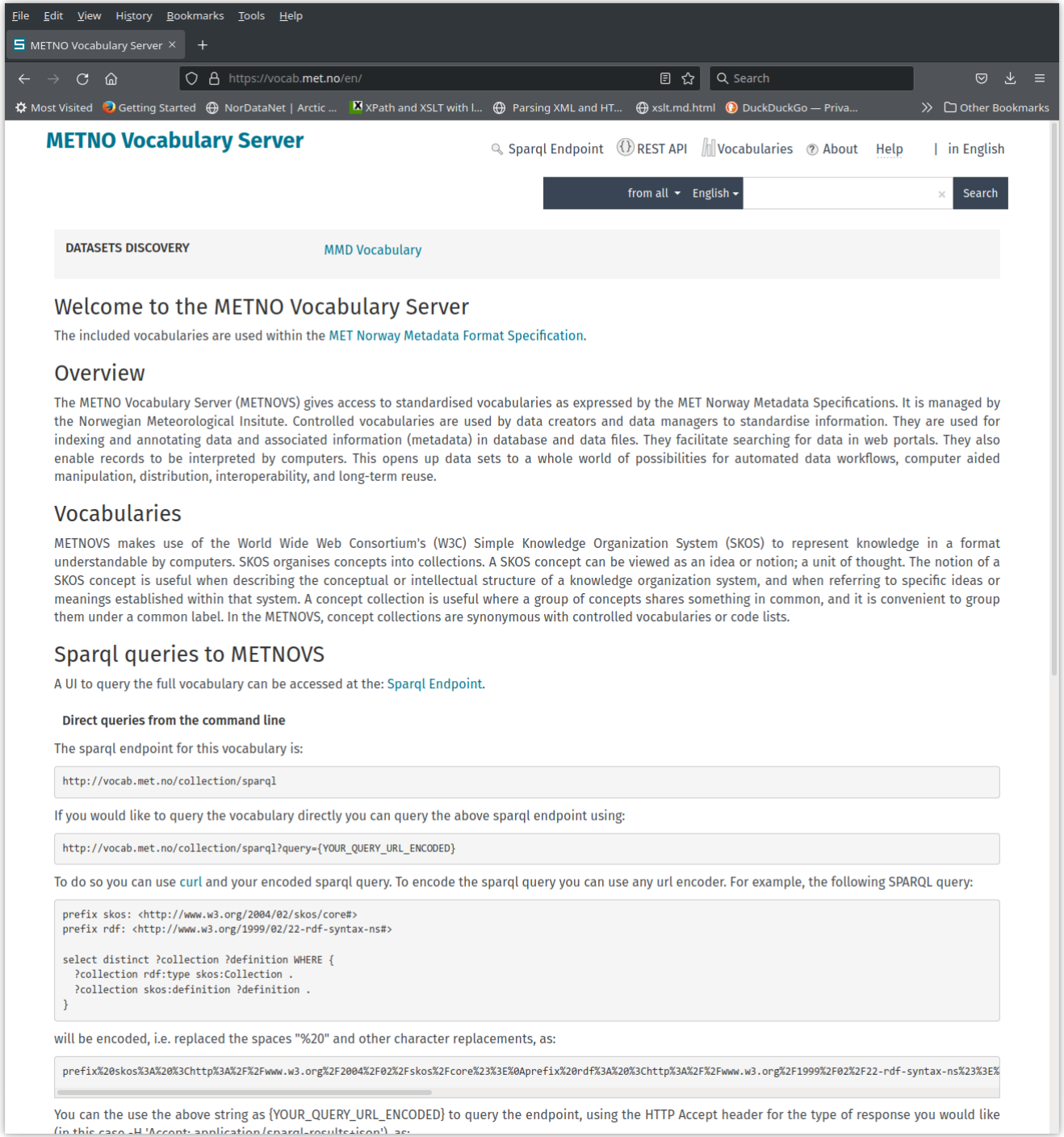
- Exchange protocols
  - OAI-PMH, OGC CSW, (OpenSearch)
- Information containers
  - ISO19115/ISO19139 (various profiles), GCMD DIF (multiple versions), ACDD, MMD
  - Looking into schema.org and DCAT
- Terminologies/ontologies
  - GCMD, OSGeo, (WIGOS)
- Using MMD as the metadata profile for the central services
  - This is built to be compatible with ISO19115 and GCMD DIF

WMO OMM

Discovery metadata workflow



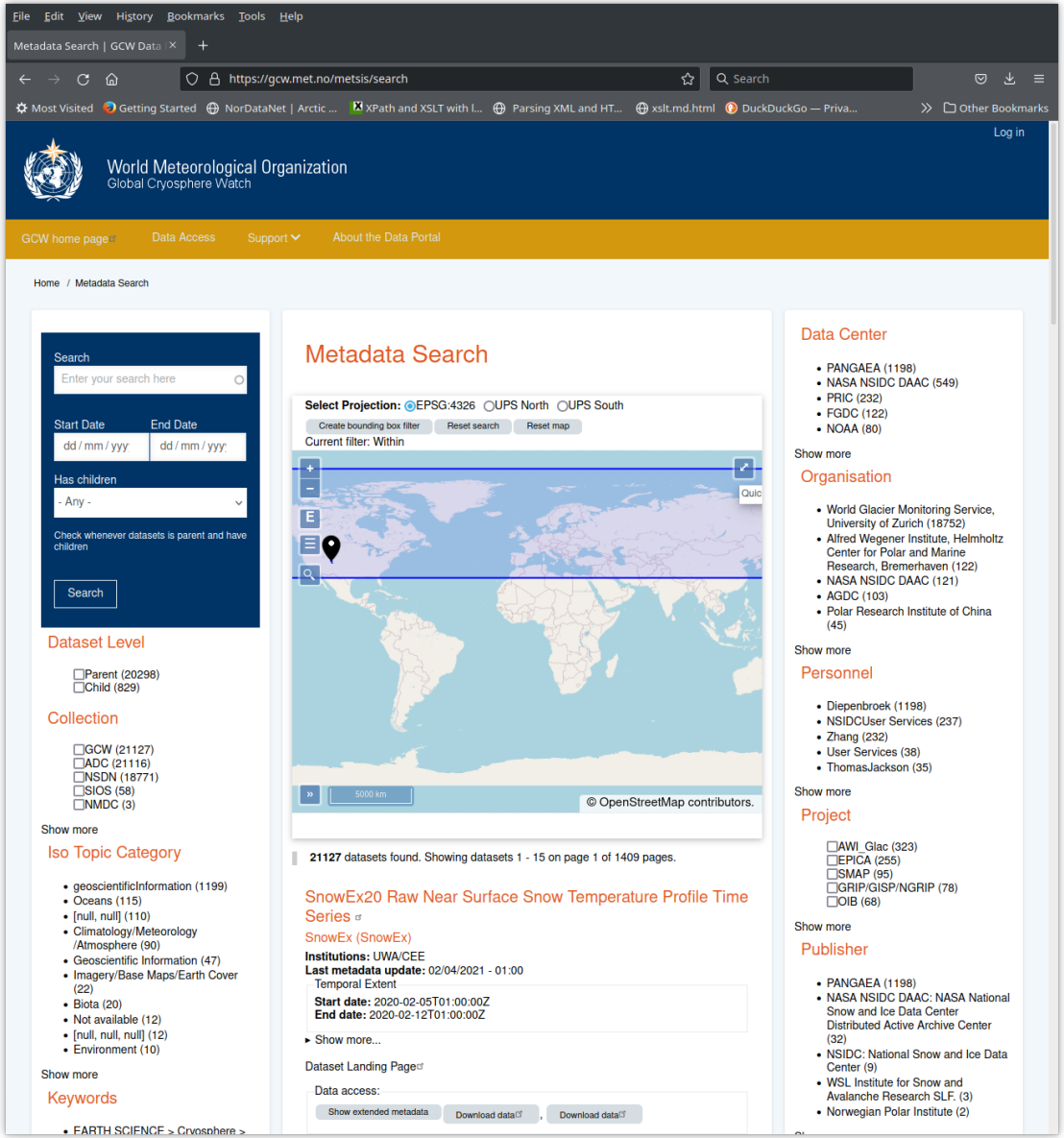
WMO OMM


File Edit View History Bookmarks Tools Help
METNO Vocabulary Server
https://vocab.met.no/en/
Search
Most Visited Getting Started NorDataNet | Arctic ... XPath and XSLT with l... Parsing XML and HT... xslt.md.html DuckDuckGo — Priva... Other Bookmarks
METNO Vocabulary Server
Sparql Endpoint REST API Vocabularies About Help | in English
from all English Search
DATASETS DISCOVERY
MMD Vocabulary
Welcome to the METNO Vocabulary Server
The included vocabularies are used within the MET Norway Metadata Format Specification.
Overview
The METNO Vocabulary Server (METNOVS) gives access to standardised vocabularies as expressed by the MET Norway Metadata Specifications. It is managed by the Norwegian Meteorological Insitute. Controlled vocabularies are used by data creators and data managers to standardise information. They are used for indexing and annotating data and associated information (metadata) in database and data files. They facilitate searching for data in web portals. They also enable records to be interpreted by computers. This opens up data sets to a whole world of possibilities for automated data workflows, computer aided manipulation, distribution, interoperability, and long-term reuse.
Vocabularies
METNOVS makes use of the World Wide Web Consortium's (W3C) Simple Knowledge Organization System (SKOS) to represent knowledge in a format understandable by computers. SKOS organises concepts into collections. A SKOS concept can be viewed as an idea or notion; a unit of thought. The notion of a SKOS concept is useful when describing the conceptual or intellectual structure of a knowledge organization system, and when referring to specific ideas or meanings established within that system. A concept collection is useful where a group of concepts shares something in common, and it is convenient to group them under a common label. In the METNOVS, concept collections are synonymous with controlled vocabularies or code lists.
Sparql queries to METNOVS
A UI to query the full vocabulary can be accessed at the: Sparql Endpoint.
Direct queries from the command line
The sparql endpoint for this vocabulary is:
http://vocab.met.no/collection/sparql
If you would like to query the vocabulary directly you can query the above sparql endpoint using:
http://vocab.met.no/collection/sparql?query={YOUR_QUERY_URL_ENCODED}
To do so you can use curl and your encoded sparql query. To encode the sparql query you can use any url encoder. For example, the following SPARQL query:
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
select distinct ?collection ?definition WHERE {
?collection rdf:type skos:Collection .
?collection skos:definition ?definition .
}
will be encoded, i.e. replaced the spaces "%20" and other character replacements, as:
prefix%20skos%3A%20%3Chttp%3A%2F%2Fwww.w3.org%2F2004%2F02%2Fskos%2Fcore%23%3E%0Aprefix%20rdf%3A%20%3Chttp%3A%2F%2Fwww.w3.org%2F1999%2F02%2F22-rdf-syntax-ns%23%3E%
You can the use the above string as {YOUR_QUERY_URL_ENCODED} to query the endpoint, using the HTTP Accept header for the type of response you would like

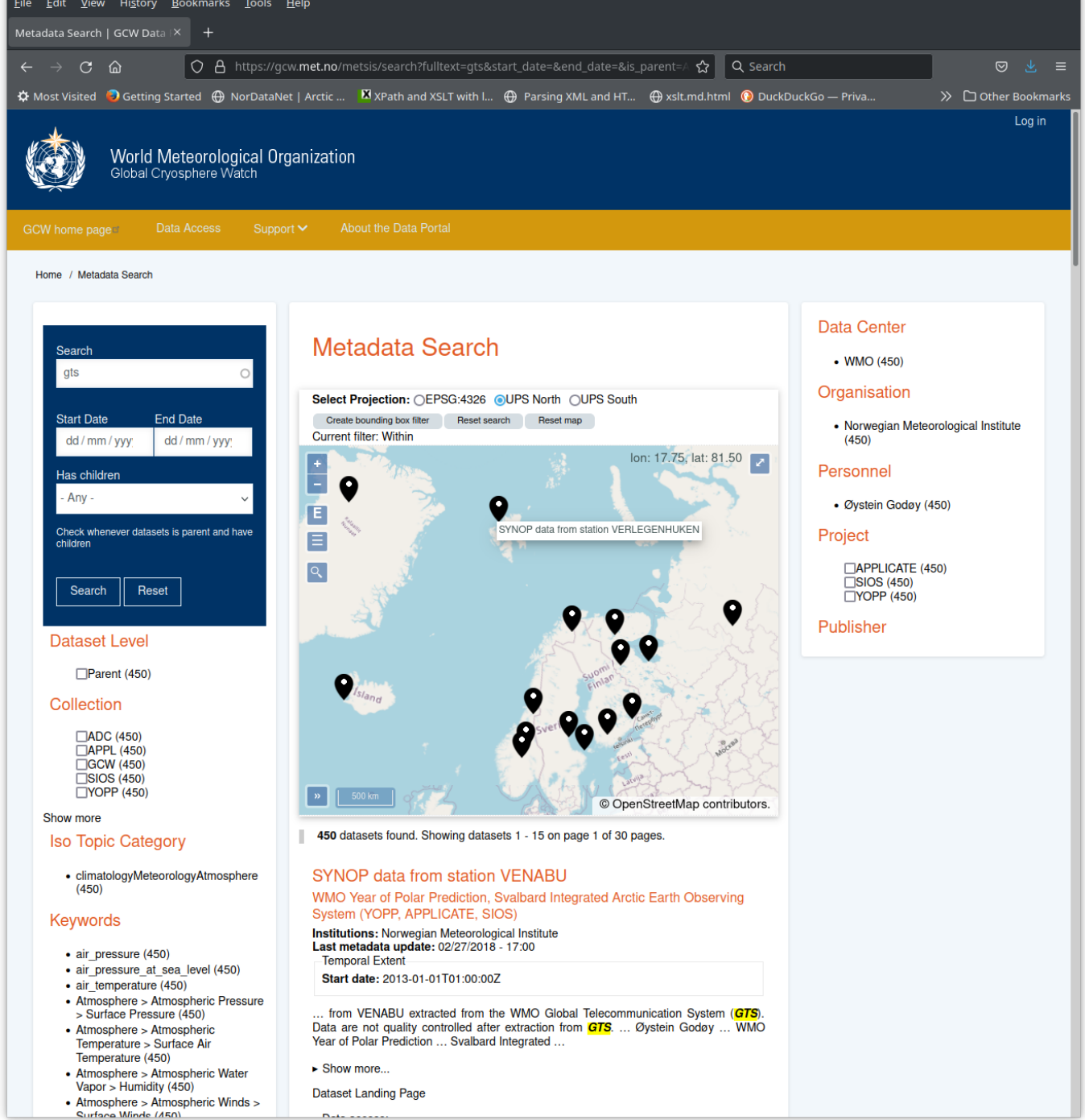
WMO OMM

Data discovery

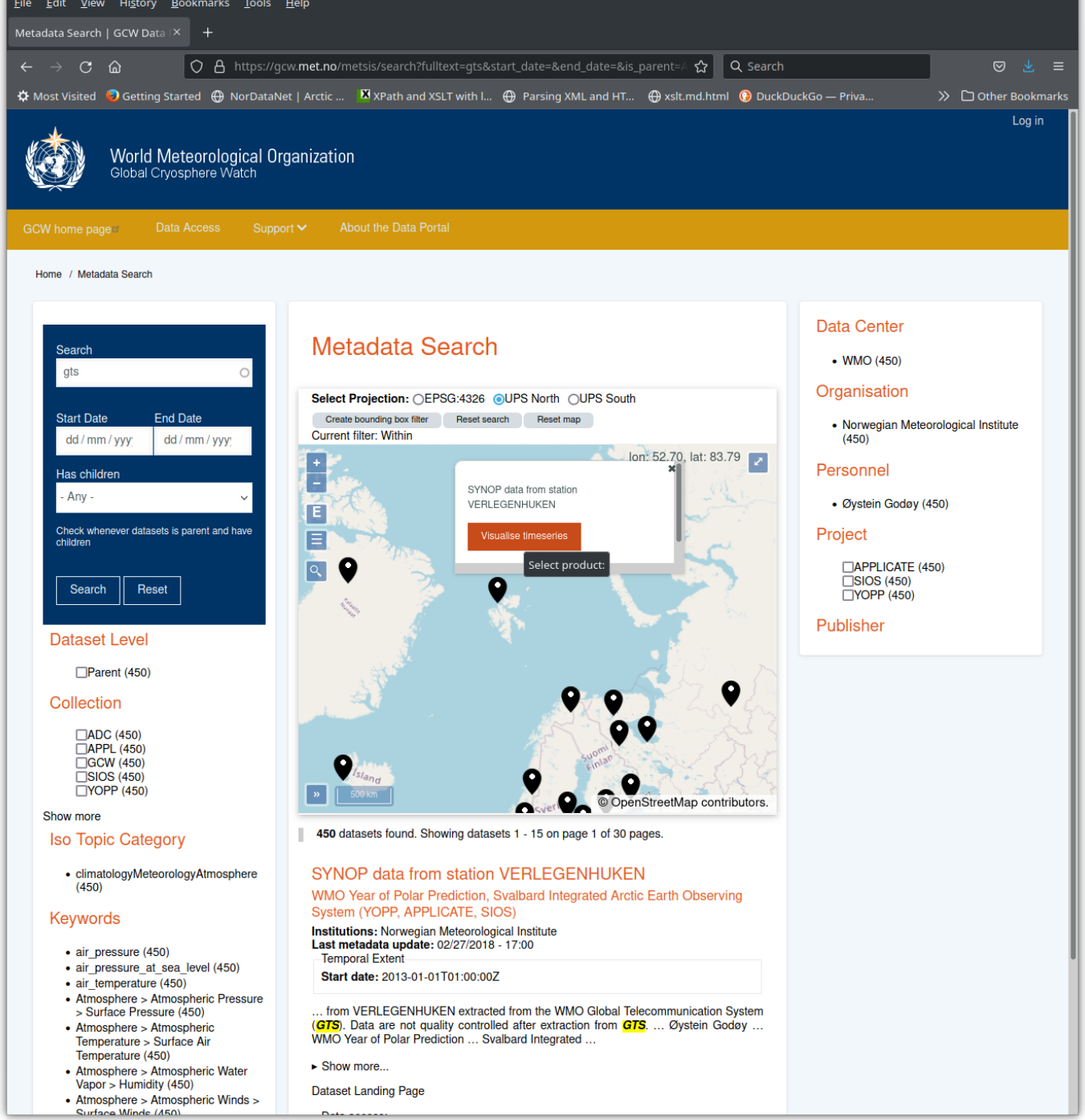
- The GCW data portal is available at https://gcw.met.no
- It serves a human interface and machine interfaces
- The human interface relies on SolR and extensive use of faceting
- Machine interfaces are implemented using pyCSW serving OAI-PMH, OGC CSW, (OpenSearch)
  - Challenges are related to backends for SolR and pyCSW and the information model for pyCSW (in particular handling of semantics)
  - Aiming to use SolR as storage backend for pyCSW, but implementing a direct feed from SolR to pyCSW first
  - All documents are stored as XML in a version control system (git) upon ingestion in the catalogue
- Harvesting information also by traversing THREDDS Catalogues and extracting information from CF-NetCDF with ACDD elements



WMO OMM


File Edit View History Bookmarks Tools Help
Metadata Search | GCW Data
https://gcw.met.no/metsis/search?fulltext=gts&start_date=&end_date=&is_parent=A
Search
Most Visited Getting Started NorDataNet | Arctic ... XPath and XSLT with l... Parsing XML and HT... xslt.md.html DuckDuckGo — Priva... Other Bookmarks
Log in
World Meteorological Organization
Global Cryosphere Watch
GCW home page Data Access Support About the Data Portal
Home / Metadata Search
Search
gts
Start Date
dd / mm / yyy
End Date
dd / mm / yyy
Has children
- Any -
Check whenever datasets is parent and have children
Search
Reset
Dataset Level
Parent (450)
Collection
ADC (450)
APPL (450)
GCW (450)
SIOS (450)
YOPP (450)
Show more
Iso Topic Category
climatologyMeteorologyAtmosphere (450)
Keywords
air_pressure (450)
air_pressure_at_sea_level (450)
air_temperature (450)
Atmosphere > Atmospheric Pressure > Surface Pressure (450)
Atmosphere > Atmospheric Temperature > Surface Air Temperature (450)
Atmosphere > Atmospheric Water Vapor > Humidity (450)
Atmosphere > Atmospheric Winds > Surface Winds (450)
Metadata Search
Select Projection: EPSG:4326 UPS North UPS South
Create bounding box filter
Reset search
Reset map
Current filter: Within
lon: 17.75, lat: 81.50
SYNOP data from station VERLEGENHUKEN
Island
Suomi Finland
Latvia
500 km
© OpenStreetMap contributors.
450 datasets found. Showing datasets 1 - 15 on page 1 of 30 pages.
SYNOP data from station VENABU
WMO Year of Polar Prediction, Svalbard Integrated Arctic Earth Observing System (YOPP, APPLICATE, SIOS)
Institutions: Norwegian Meteorological Institute
Last metadata update: 02/27/2018 - 17:00
Temporal Extent
Start date: 2013-01-01T01:00:00Z
... from VENABU extracted from the WMO Global Telecommunication System (GTS). Data are not quality controlled after extraction from GTS. ... Øystein Godøy ... WMO Year of Polar Prediction ... Svalbard Integrated ...
Show more...
Dataset Landing Page
Data Center
WMO (450)
Organisation
Norwegian Meteorological Institute (450)
Personnel
Øystein Godøy (450)
Project
APPLICATE (450)
SIOS (450)
YOPP (450)
Publisher

WMO OMM


Metadata Search | GCW Data
https://gcw.met.no/metsis/search?fulltext=gts&start_date=&end_date=&is_parent=A
Search
Most Visited
Getting Started
NorDataNet | Arctic ...
XPath and XSLT with l...
Parsing XML and HT...
xslt.md.html
DuckDuckGo — Priva...
Other Bookmarks
Log in
World Meteorological Organization
Global Cryosphere Watch
GCW home page
Data Access
Support
About the Data Portal
Home / Metadata Search
Search
gts
Start Date
End Date
dd / mm / yyyy
dd / mm / yyyy
Has children
- Any -
Check whenever datasets is parent and have children
Search
Reset
Dataset Level
Parent (450)
Collection
ADC (450)
APPL (450)
GCW (450)
SIOS (450)
YOPP (450)
Show more
Iso Topic Category
climatologyMeteorologyAtmosphere (450)
Keywords
air_pressure (450)
air_pressure_at_sea_level (450)
air_temperature (450)
Atmosphere > Atmospheric Pressure > Surface Pressure (450)
Atmosphere > Atmospheric Temperature > Surface Air Temperature (450)
Atmosphere > Atmospheric Water Vapor > Humidity (450)
Atmosphere > Atmospheric Winds > Surface Winds (450)
Metadata Search
Select Projection: EPSG:4326 UPS North UPS South
Create bounding box filter
Reset search
Reset map
Current filter: Within
lon: 52.70, lat: 83.79
SYNOP data from station VERLEGENHUKEN
Visualise timeseries
Select product:
Island
Suomi / Finlan
500 km
© OpenStreetMap contributors.
450 datasets found. Showing datasets 1 - 15 on page 1 of 30 pages.
SYNOP data from station VERLEGENHUKEN
WMO Year of Polar Prediction, Svalbard Integrated Arctic Earth Observing System (YOPP, APPLICATE, SIOS)
Institutions: Norwegian Meteorological Institute
Last metadata update: 02/27/2018 - 17:00
Temporal Extent
Start date: 2013-01-01T01:00:00Z
... from VERLEGENHUKEN extracted from the WMO Global Telecommunication System (GTS). Data are not quality controlled after extraction from GTS. ... Øystein Godøy ... WMO Year of Polar Prediction ... Svalbard Integrated ...
Show more...
Dataset Landing Page
Data Center
WMO (450)
Organisation
Norwegian Meteorological Institute (450)
Personnel
Øystein Godøy (450)
Project
APPLICATE (450)
SIOS (450)
YOPP (450)
Publisher

WMO OMM

File Edit View History Bookmarks Tools Help

File Edit View History Bookmarks Tools Help

Metadata Search | GCW Data

https://gcw.met.no/metsis/search?fulltext=bjørnøya&start_date=&end_date=&is_par

Search

Most Visited | Getting Started | NorDataNet | Arctic ... | XPath and XSLT with l... | Parsing XML and HT... | xslt.md.html | DuckDuckGo — Priva... | Other Bookmarks

GCW home page | Data Access | Support | About the Data Portal

- Atmosphere > Atmospheric Pressure > Surface Pressure (1)
- Atmosphere > Atmospheric Temperature > Surface Air Temperature (1)
- Atmosphere > Atmospheric Temperature > Surface Temperature > Air Temperature (1)
- Atmosphere > Atmospheric Water Vapor (1)
- Atmosphere > Atmospheric Water Vapor > Humidity (1)
- Atmosphere > Clouds > Cloud Amount/Frequency (1)

Show more

Dataset Landing Page

Data access:

Add to Basket | Show extended metadata | Visualise TimeSeries | Download as ASCII | OPeNDAP

License : Free **Access:** Free

Observations from *Bjørnøya*

Institutions: Norwegian Meteorological Institute
Last metadata update: 05/16/2019 - 23:40

Temporal Extent

Start date: 1910-08-29T01:00:00Z

... manual quality control routines. ... Vegard Kristiansen ... Observations from *Bjørnøya* ...

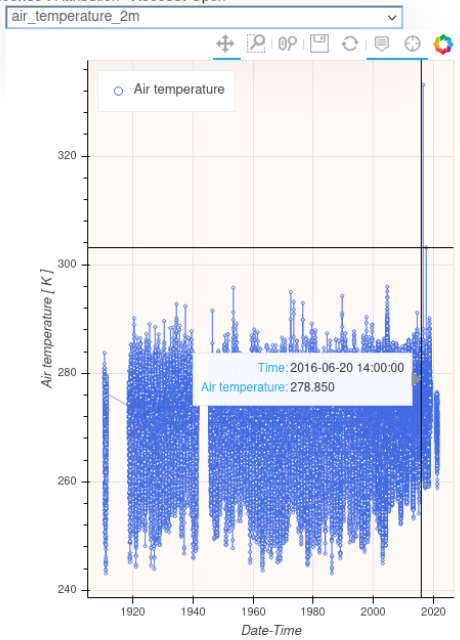
▸ Show more...

Dataset Landing Page

Data access:

Add to Basket | Show extended metadata | Visualise TimeSeries | Download data | Download as ASCII | OPeNDAP

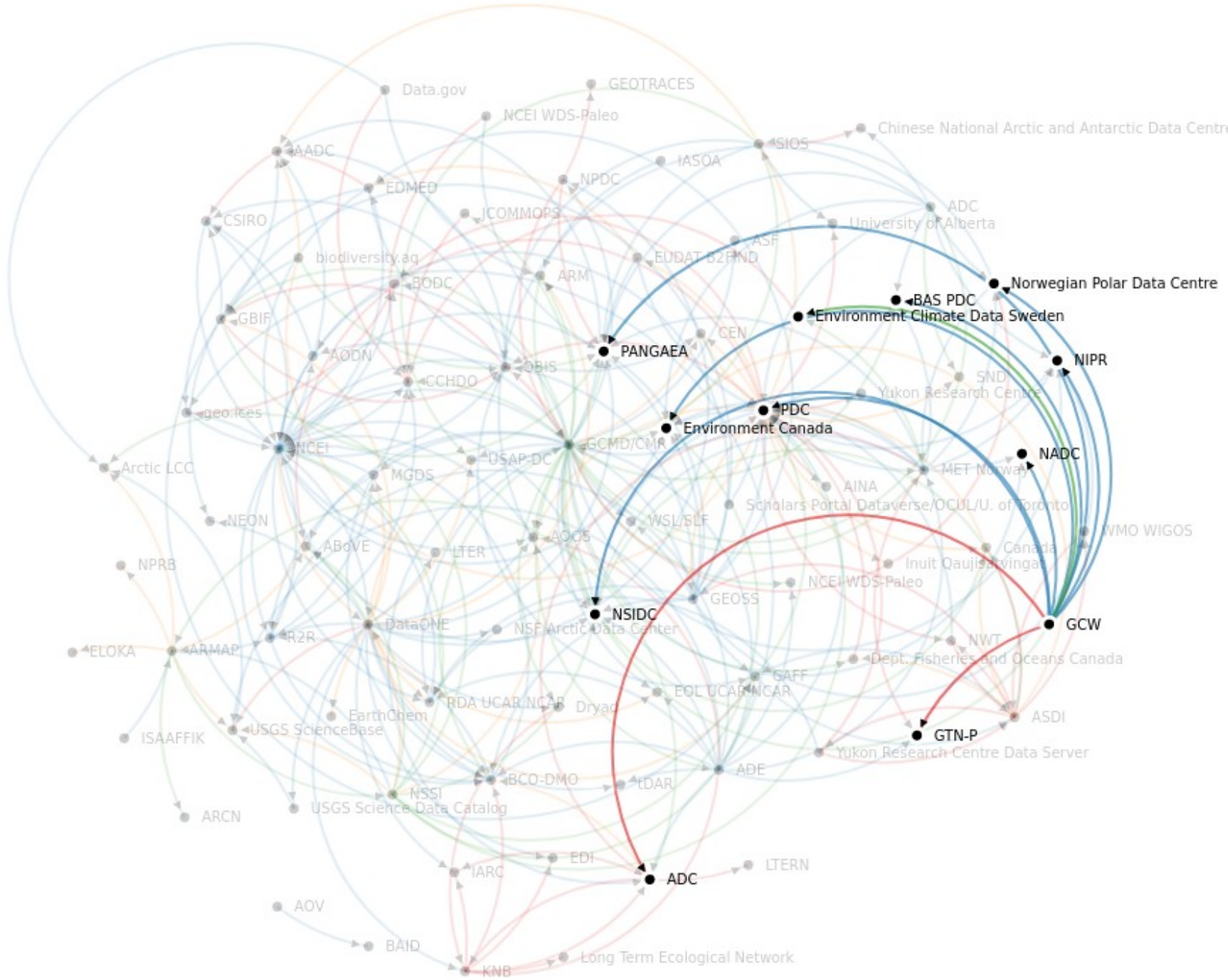
License : Attribution **Access:** Open



WMO OMM

Data exchange

- GCW is setting up services for data providers not able to serve their own data
  - Only serving in situ measurements for the time being relying on WSL/SLF or MET to serve the data according to GCW requirements
  - GCW recommends to serve data using OPeNDAP and WMS using THREDDS, Hyrax or pyDAP as servers and develops software components to help the community
- GCW will act as a mediator between non NMHS providers and the WMO systems
  - Entering and extracting information to serve the full community in mutual benefit


Data.gov
GEOTRACES
NCEI WDS-Paleo
Chinese National Arctic and Antarctic Data Centre
AADC
SIOS
IASOA
EDMED
NPDC
ADC
CSIRO
JCOMMOPS
University of Alberta
ASF
biodiversity.aq
EUDAT B2FIND
ARM
BODC
Norwegian Polar Data Centre
BAS PDC
GBIF
Environment Climate Data Sweden
CEN
PANGAEA
NIPR
AODN
OBIS
CCHDO
SND
Yukon Research Centre
PDC
geo.ices
Environment Canada
NCEI
GCMD/CMR
NADC
USAP-DC
Arctic LCC
MGDS
MET Norway
AINA
Scholars Portal Dataverse/OCUL/U. of Toronto
NEON
WSL/SLF
WMO WIGOS
AOOS
ABoVE
LTER
Canada
NPRB
Inuit Qaujisarvingat
NCEI WDS-Paleo
GEOSS
NSIDC
GCW
DataONE
NSF Arctic Data Center
R2R
NWT
ELOKA
ARMAP
Dept. Fisheries and Oceans Canada
SAFF
EOL UCAR/NCAR
RDA UCAR/NCAR
Dryad
EarthChem
ASDI
USGS ScienceBase
GTN-P
ISAAFFIK
Yukon Research Centre Data Server
ADE
BCO-DMO
tDAR
NSSI
USGS Science Data Catalog
ARCN
EDI
LTERN
IARC
ADC
AOV
BAID
Long Term Ecological Network
KNB

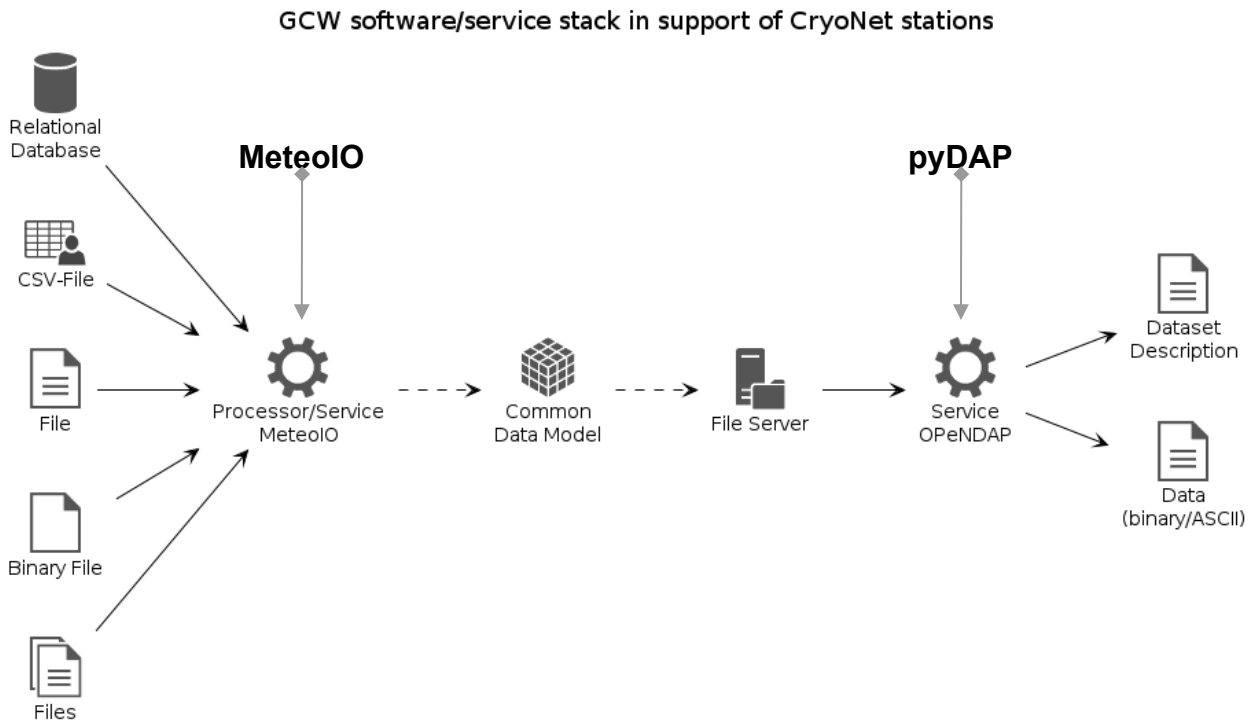
WMO OI

Contributions to Open Source development

- Everything developed is released using open software licenses
- Core functionality is included using open source projects like pyCSW
  - Contributions focus on information handling
  - API’s work well
- Specific focus is put on MeteoIO and making services for the community (web services, docker containers for easy implementation etc)
  - Through EU-funding in Arctic Passion

WMO OMM

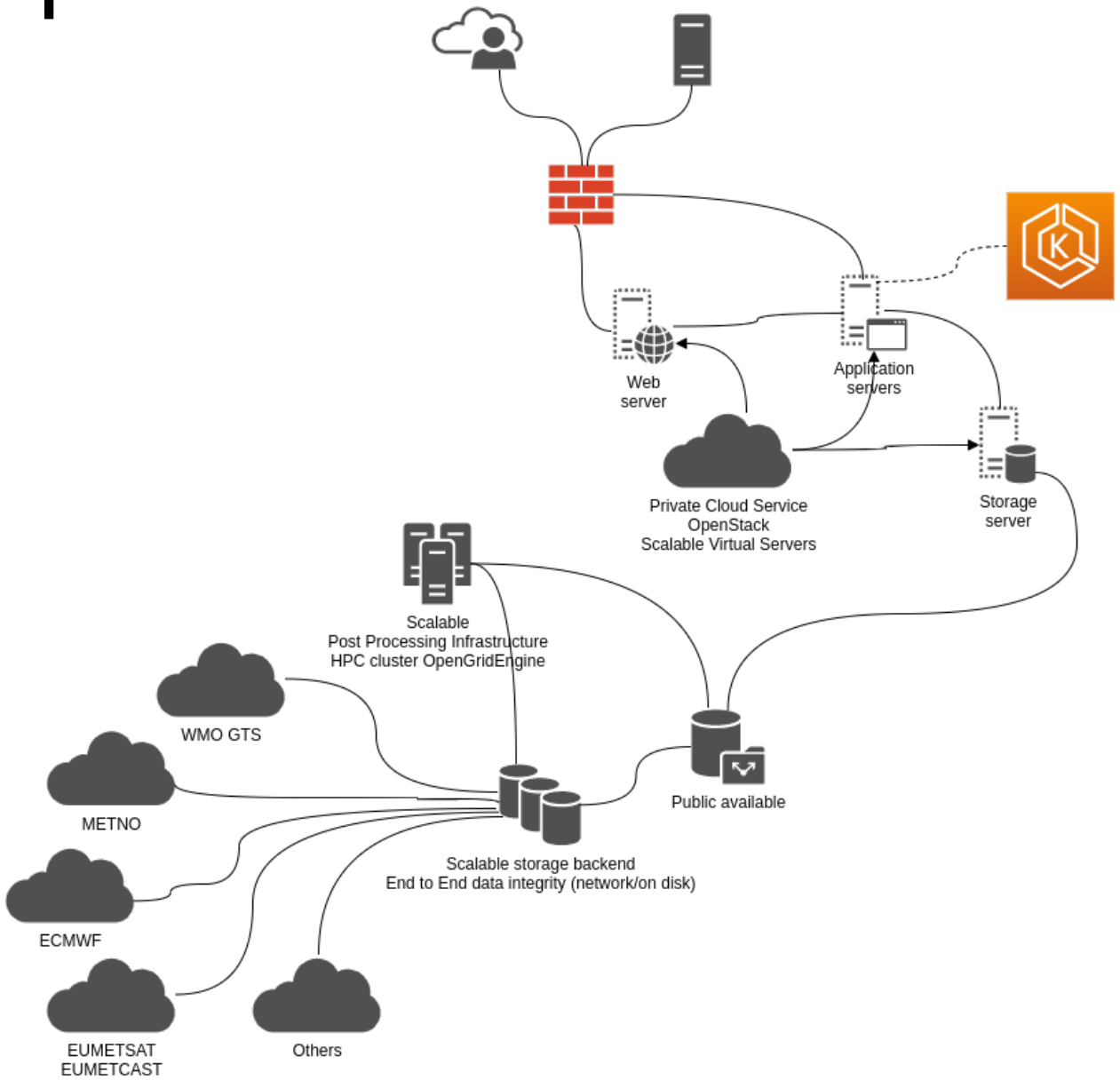
MeteoIO



- 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
- Development of web services and containers through Arctic Passion

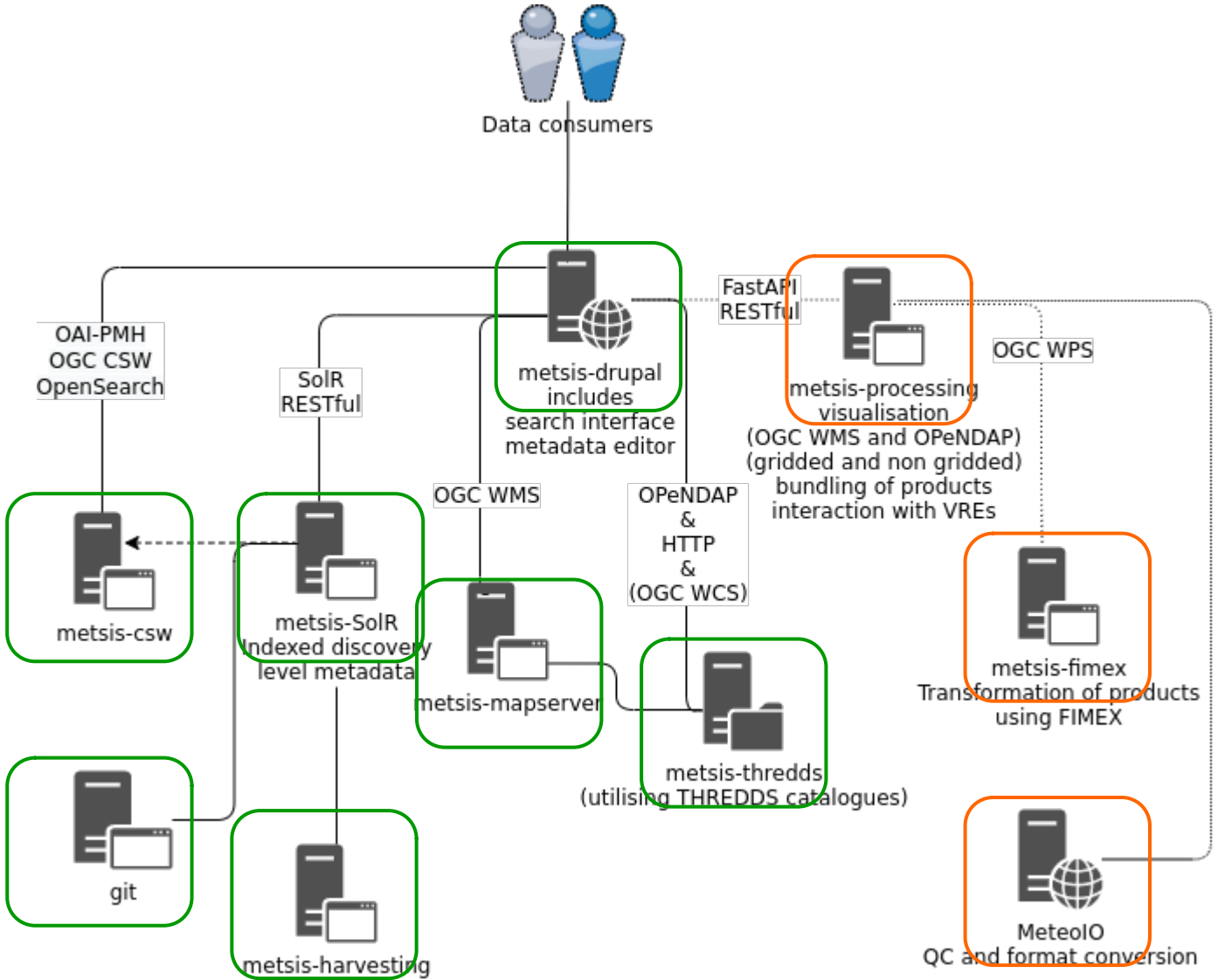
WMO OMM

Implementation environment



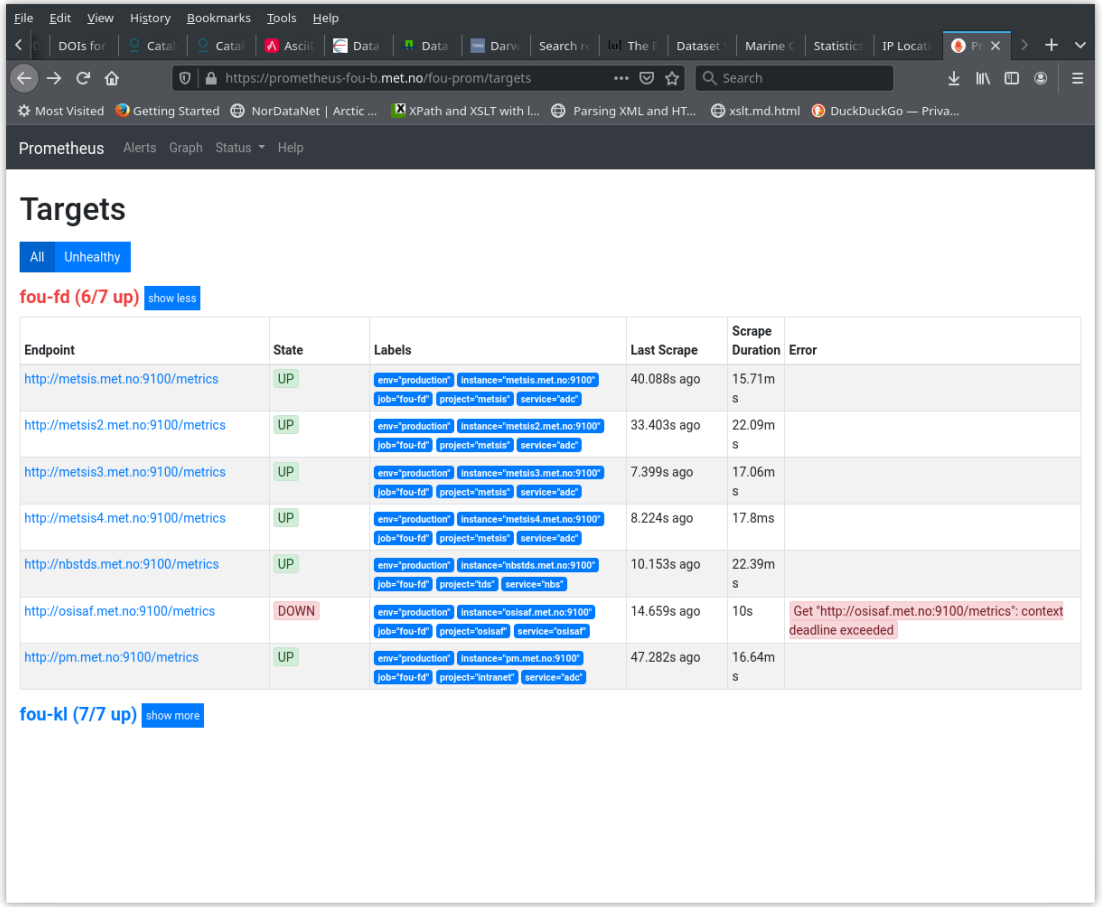
WMO OMM

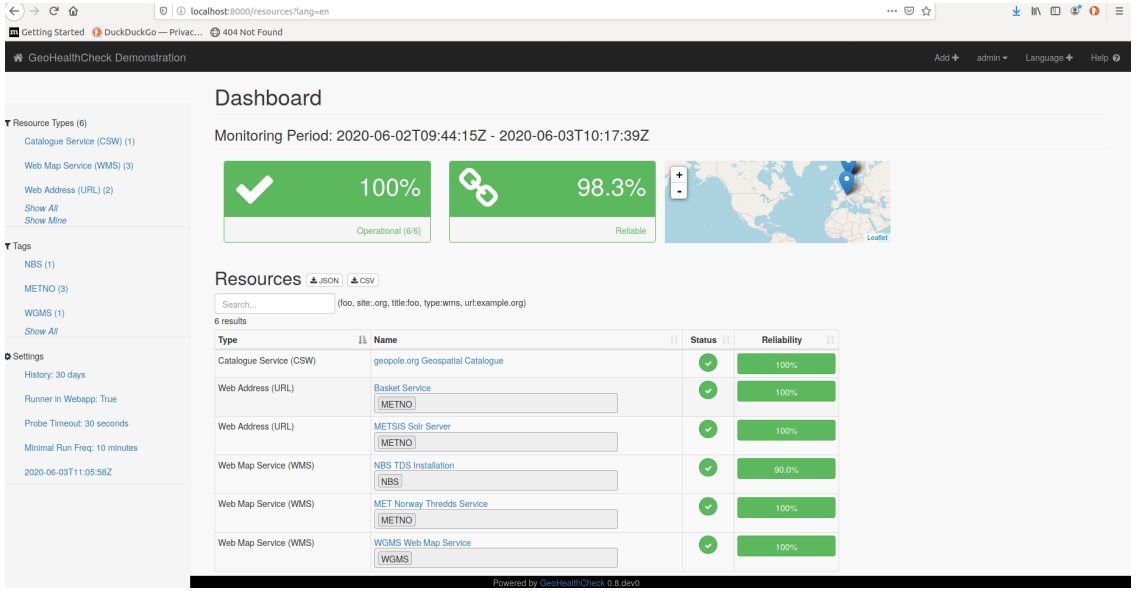
Architecture overview



WMO OMM

Monitoring





WMO OMM

Building the system

- Relies on external funding received through a number of projects, both national and European
- Project funding that has supported the development includes (non exhaustive list)
  - Arctic Passion (H2020)
  - SIOS KC (RCN)
  - Norwegian Scientific Data Network (RCN)
  - Norwegian Ground Segment for Satellite Data (NOSA)
  - ESA Cryosphere Virtual Laboratory (ESA)
  - ENVRI-FAIR (H2020)
  - APPLICATE (H2020)
  - …
- In order to establish distributed data management services, utilisation of synergies with research and other communities is vital
  - GCW is engaged in IASC/SAON activities, specifically on the Arctic Data Committee and in several working groups (e.g. on semantics and documentation of observing assets)
  - GCW interacts with RDA and ICS CODATA/WDS where relevant

Input to WIS2

- WMO formats has little traction outside the NMHS, need to be pragmatic and support other standards (each for its specific purpose)
- Improve the usage of CF-NetCDF and ACDD and actively engage in developments
  - It simplifies interaction with external communities
- Evaluate Zarr as backend for CF-NetCDF
- Include OPeNDAP as a mechanism for data exchange
  - Used internally in Copernicus services, ESGF etc
- Connect with external communities on data management
  - e.g. on data management RDA, CODATA, WDS, on refinement of schema.org (ESIP)
  - Interact actively with semantic communities/activities (e.g. ESIP, ENVO, ...)
- Primary challenge on data provider side is too complex standards (e.g. WMO Core Profile), understanding of semantics and how to use this in metadata
  - simplification is needed
- Identify open source software/approaches that could benefit a larger community and actively contribute to this
  - e.g. pyCSW, THREDDS/HYRAX/pyDAP, MeteoIO, ...

WEATHER CLIMATE WATER
TEMPS CLIMAT EAU

Thank you
Merci

WMO OMM
World Meteorological Organization
Organisation météorologique mondiale