INTER ⇄ ACT

Data documentation and dissemination supporting services, some standards and tools that may help

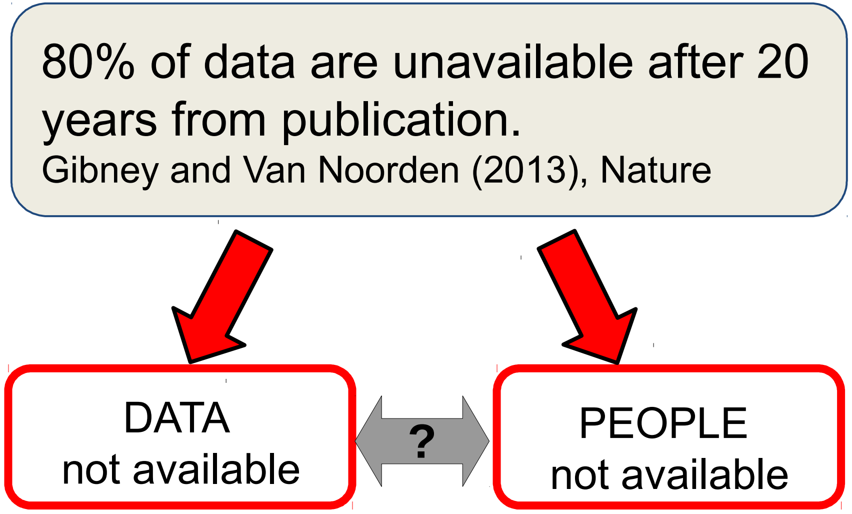
Øystein Godøy, Mathias Bavay and Massimo DiStefano

Norwegian Meteorological Institute

WSL SLF

INTER ⇄ ACT

Data Management



80% of data are unavailable after 20 years from publication.
Gibney and Van Noorden (2013), Nature

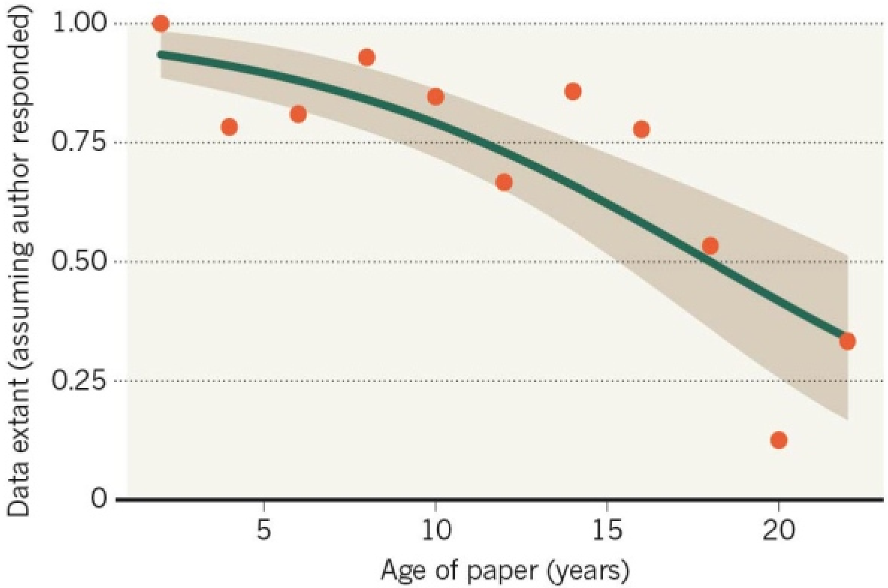
DATA
not available

?

PEOPLE
not available

MISSING DATA

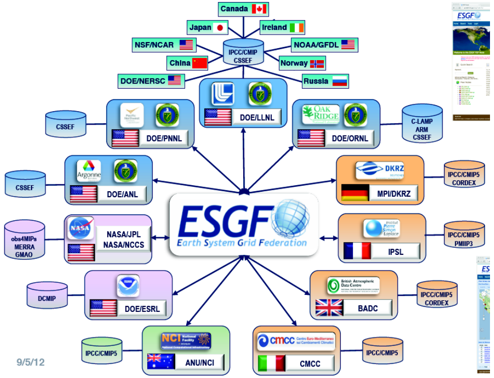
As research articles age, the odds of their raw data being extant drop dramatically.

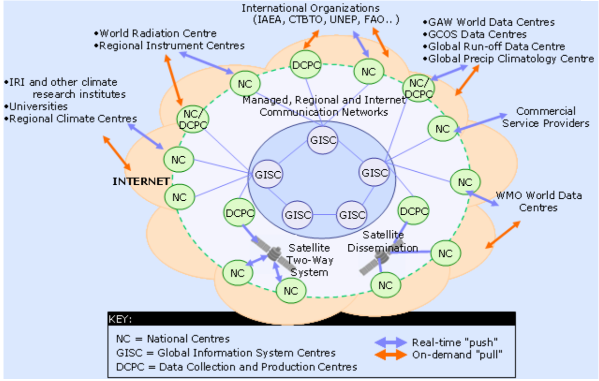


http://www.nature.com/news/scientists-losing-data-at-a-rapid-rate-1.14416

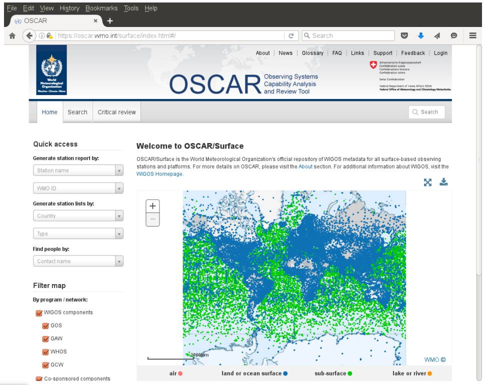
Visibility through integration

- Regional and global data management frameworks
  - GEO
  - INSPIRE
  - SAON/IASC Arctic Data Committee
  - WMO Information System
  - WMO Integrated Global Observing System
  - ICSU Word Data System
  - GBIF
  - ...



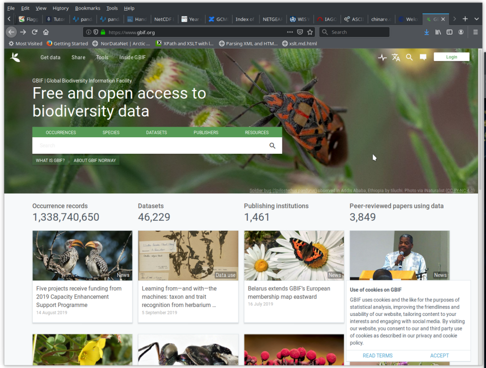


International Organizations
(IAEA, CTBTO, UNEP, FAO..)
•World Radiation Centre
•Regional Instrument Centres
•GAW World Data Centres
•GCOS Data Centres
•Global Run-off Data Centre
•Global Precip Climatology Centre
•IRI and other climate research institutes
•Universities
•Regional Climate Centres
Managed, Regional and Internet Communication Networks
Commercial Service Providers
INTERNET
WMO World Data Centres
Satellite Two-Way System
Satellite Dissemination
NC
DCPC
NC/DCPC
GISC
KEY:
NC = National Centres
GISC = Global Information System Centres
DCPC = Data Collection and Production Centres
Real-time "push"
On-demand "pull"



Welcome to OSCAR/Surface

OSCAR/Surface is the World Meteorological Organization's official repository of WIGOS metadata for all surface-based observing stations and platforms. For more details on OSCAR, please visit the About section. For additional information about WIGOS, visit the WIGOS Homepage.



Free and open access to biodiversity data

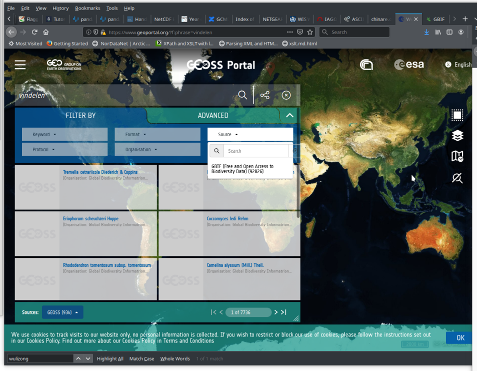
Occurrence records	Datasets	Publishing institutions	Peer-reviewed papers using data
1,338,740,650	46,229	1,461	3,849

Five projects receive funding from 2019 Capacity Enhancement Support Programme

Learning from—and with—the machines: taxon and trait recognition from herbarium ...

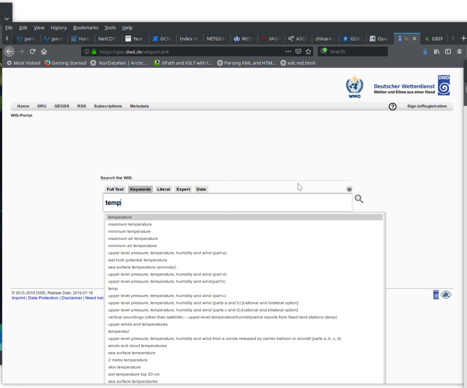
Belarus extends GBIF's European membership award

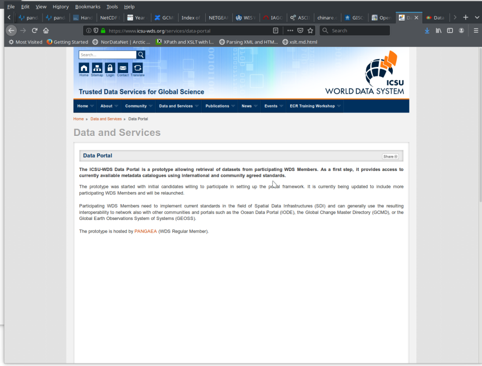
Use of cookies on GBIF



https://www.geoportal.org/?f:phrase=vindelen
GEOSS Portal
esa
vindelen
FILTER BY
ADVANCED
Keyword
Format
Source
Protocol
Organisation
GBIF (Free and Open Access to Biodiversity Data) (92826)
Sources:
OK

We use cookies to track visits to our website only, no personal information is collected. If you wish to restrict or block our use of cookies, please follow the instructions set out in our Cookies Policy. Find out more about our Cookies Policy in Terms and Conditions





Data and Services

Data Portal

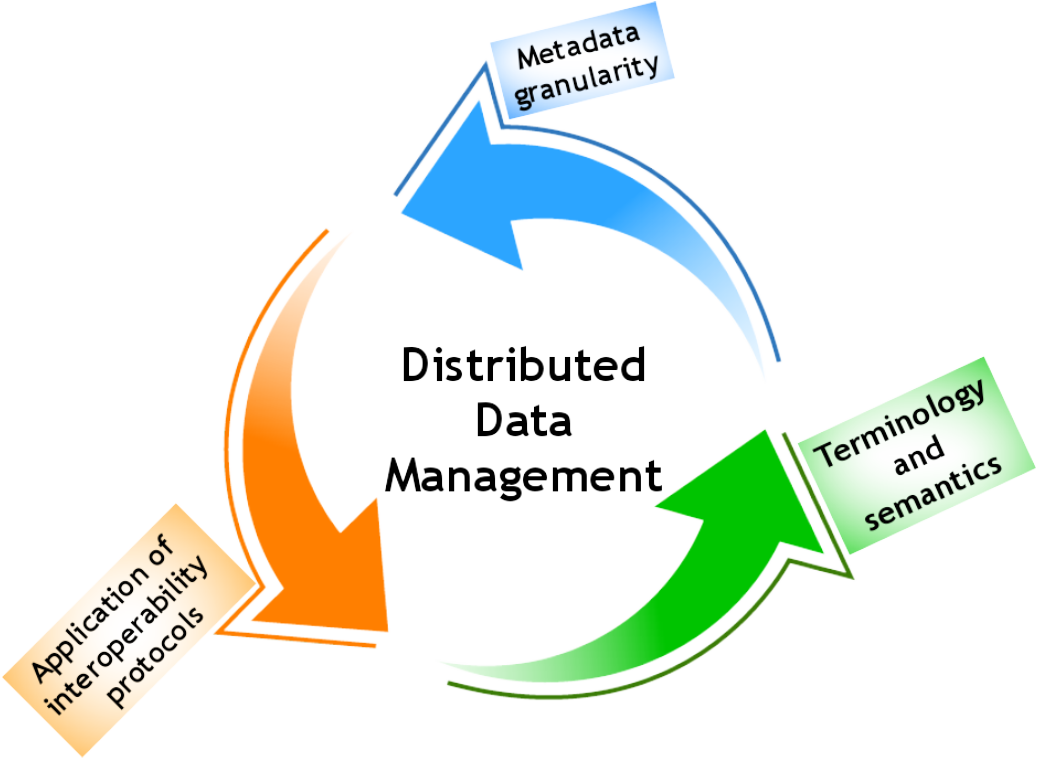
The ICSU-WDS Data Portal is a prototype allowing retrieval of datasets from participating WDS Members. As a first step, it provides access to currently available metadata catalogues using international and community agreed standards.

The prototype was started with initial candidates willing to participate in setting up the portal framework. It is currently being updated to include more participating WDS Members and will be relaunched.

Participating WDS Members need to implement current standards in the field of Spatial Data Infrastructures (SDI) and can generally use the resulting interoperability to network also with other communities and portals such as the Ocean Data Portal (ODP), the Global Change Master Directory (GCMD), or the Global Earth Observations System of Systems (GEOSS).

The prototype is hosted by PANGAEA (WDS Regular Member).

Challenges during integration


Metadata granularity
Distributed Data Management
Terminology and semantics
Application of interoperability protocols

- Interoperability
  - Discovery Metadata
    - Exchange Protocols (✓)
    - Structures (✓)
    - Semantics/terminology (-)
  - Data
    - Exchange Protocols (✓)
    - Formats (-)
    - Use metadata (✓)
    - Semantics/terminology (-)
    - Common data model (-)
- Cultural
  - Sharing data...

INTER ⇄ ACT

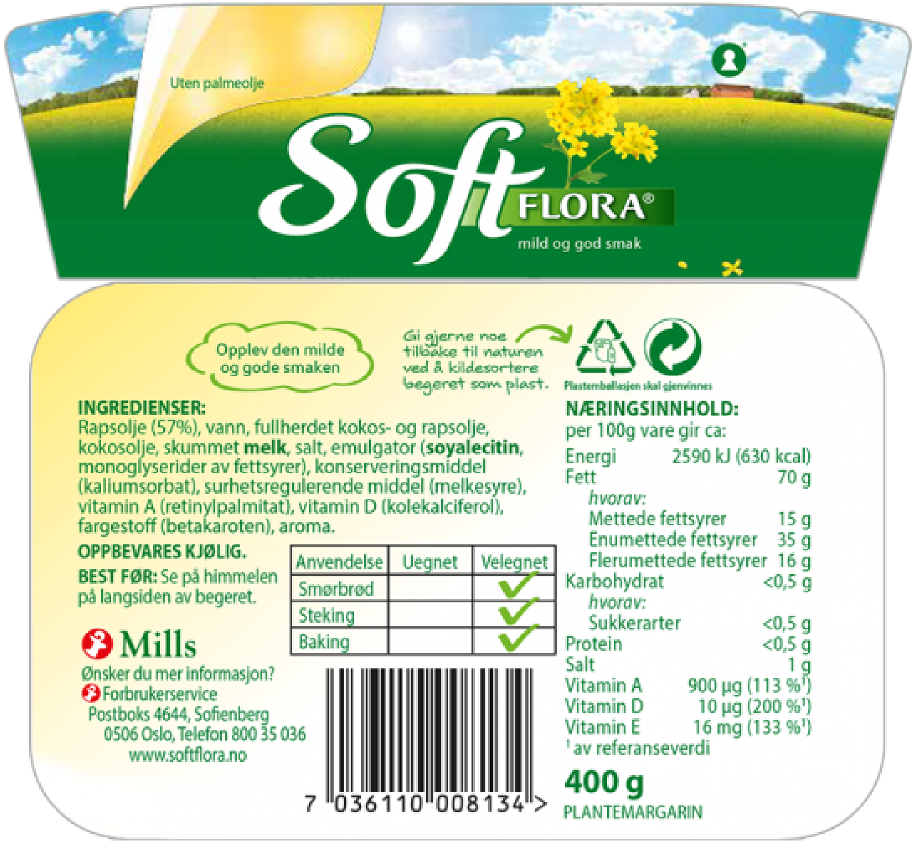
The FAIR guiding principles

- 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**

Standards and tools

- Standards
  - Discovery metadata
    - ISO19115
    - GCMD DIF
    - ACDD
    - OGC CSW
    - OAI-PMH
  - Use metadata
    - GBIF
    - CF
  - File formats (standardised)
    - NetCDF/CF
    - Excel/GBIF
    - ...

Must put data in context



Opplev den milde og gode smaken

Gi gjerne noe tilbake til naturen ved å kildesortere begeret som plast.

Plastemballasjen skal gjenvinnes

INGREDIENSER:
Rapsolje (57%), vann, fullherdet kokos- og rapsolje, kokosolje, skummet **melk**, salt, emulgator (**soyalecitin**, monoglyserider av fettsyrer), konserveringsmiddel (kaliumsorbat), surhetsregulerende middel (melkesyre), vitamin A (retinylpalmitat), vitamin D (kolekalciferol), fargestoff (betakaroten), aroma.

OPPBEVARES KJØLIG.

BEST FØR: Se på himmelen på langsiden av begeret.

Anvendelse	Uegnet	Velegnet
Smørbrød		✓
Steking		✓
Baking		✓

Mills

Ønsker du mer informasjon?
Forbrukerservice
Postboks 4644, Sofienberg
0506 Oslo, Telefon 800 35 036
www.softflora.no

7 036110 008134 >

NÆRINGSINNHOLD:
per 100g vare gir ca:

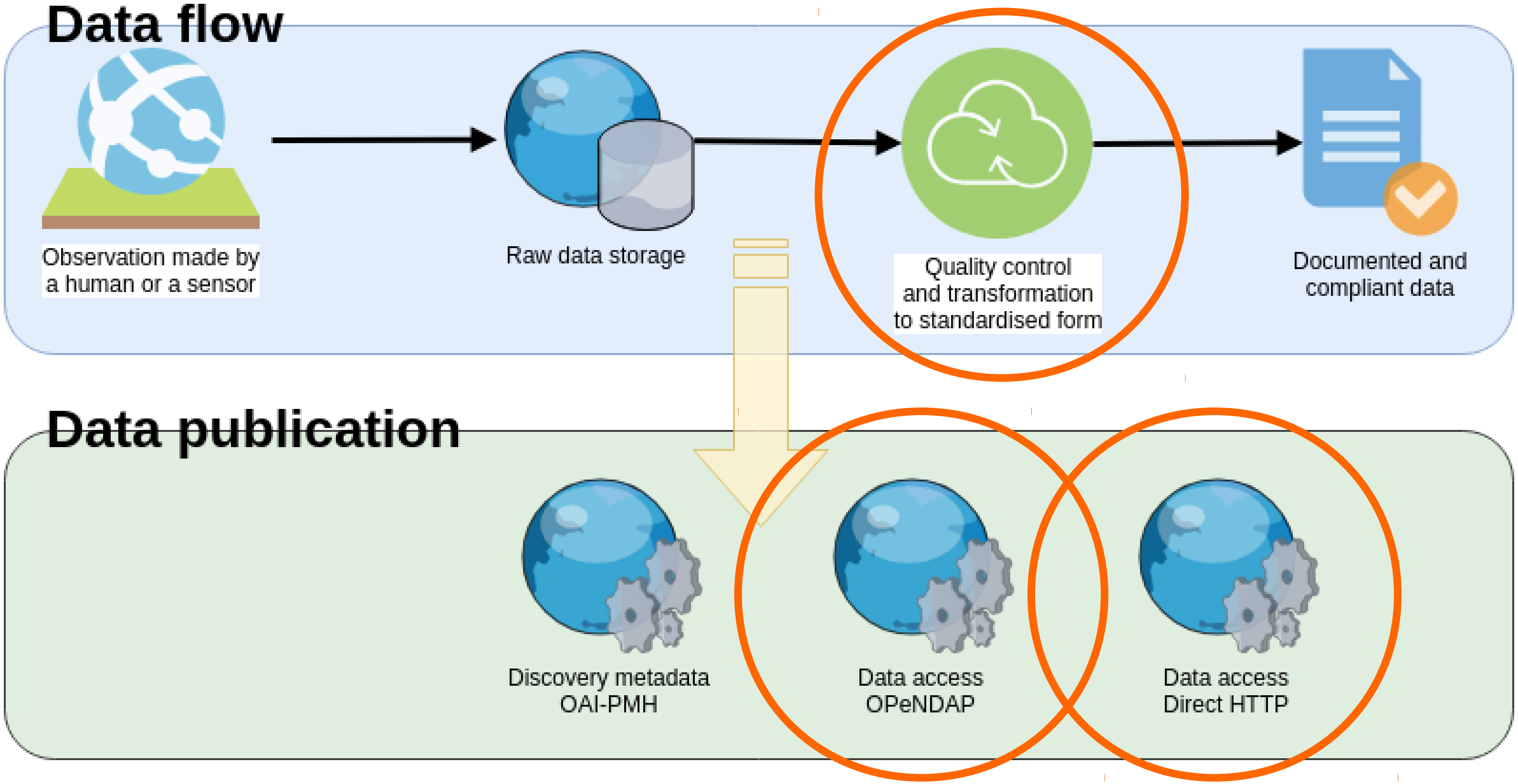
Energi	2590 kJ (630 kcal)
Fett	70 g
hvorav:	
Mettede fettsyrer	15 g
Enumettede fettsyrer	35 g
Flerumettede fettsyrer	16 g
Karbohydrat	<0,5 g
hvorav:	
Sukkerarter	<0,5 g
Protein	<0,5 g
Salt	1 g
Vitamin A	900 µg (113 %[1])
Vitamin D	10 µg (200 %[1])
Vitamin E	16 mg (133 %[1])

[1] av referanseverdi

400 g
PLANTEMARGARIN

- What’s the meaning of a number?
  - Basic metadata are needed for any use of data
- Data can be used in different ways
  - For adequate use of data, adequate information about the data is critical
- The whole is more than the sum of the pieces
  - Smart combination of information has a much larger potential than single observations
- Must
  - Make data talk together
  - Make data traceable
  - Make data count

The promised GCW/SLF software stack



Data flow

Observation made by a human or a sensor

Raw data storage

Quality control and transformation to standardised form

Documented and compliant data

Data publication

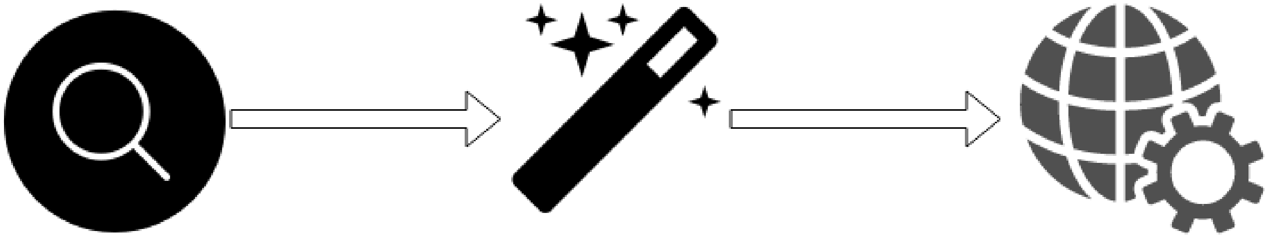
Discovery metadata OAI-PMH

Data access OPeNDAP

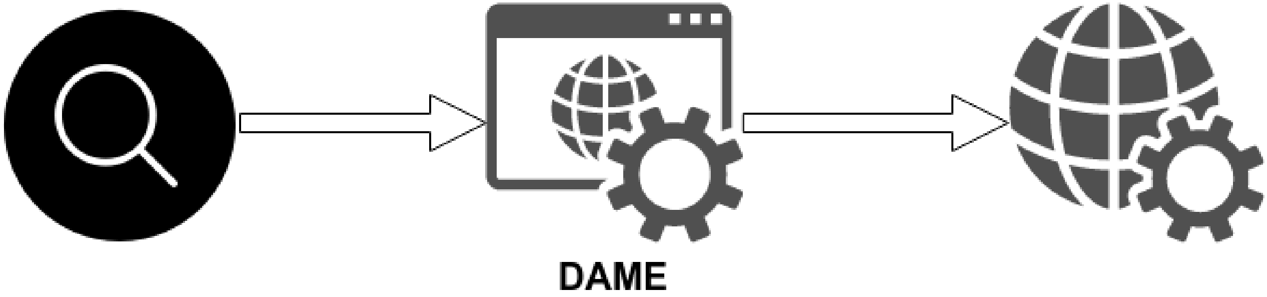
Data access Direct HTTP

The concept...

- Requirements
  - No-brainer data use (quick usage for any application)
  - Scalability (i.e. minimum effort to add more stations)
- Constraints
  - Diversity of formats & protocols
  - Diversity of variable names
  - Diversity of units & other metadata

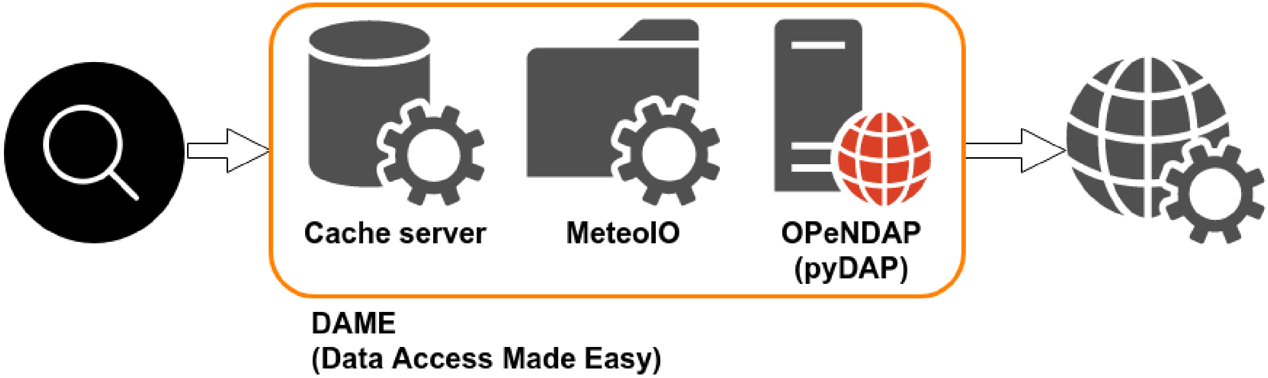


The magic...



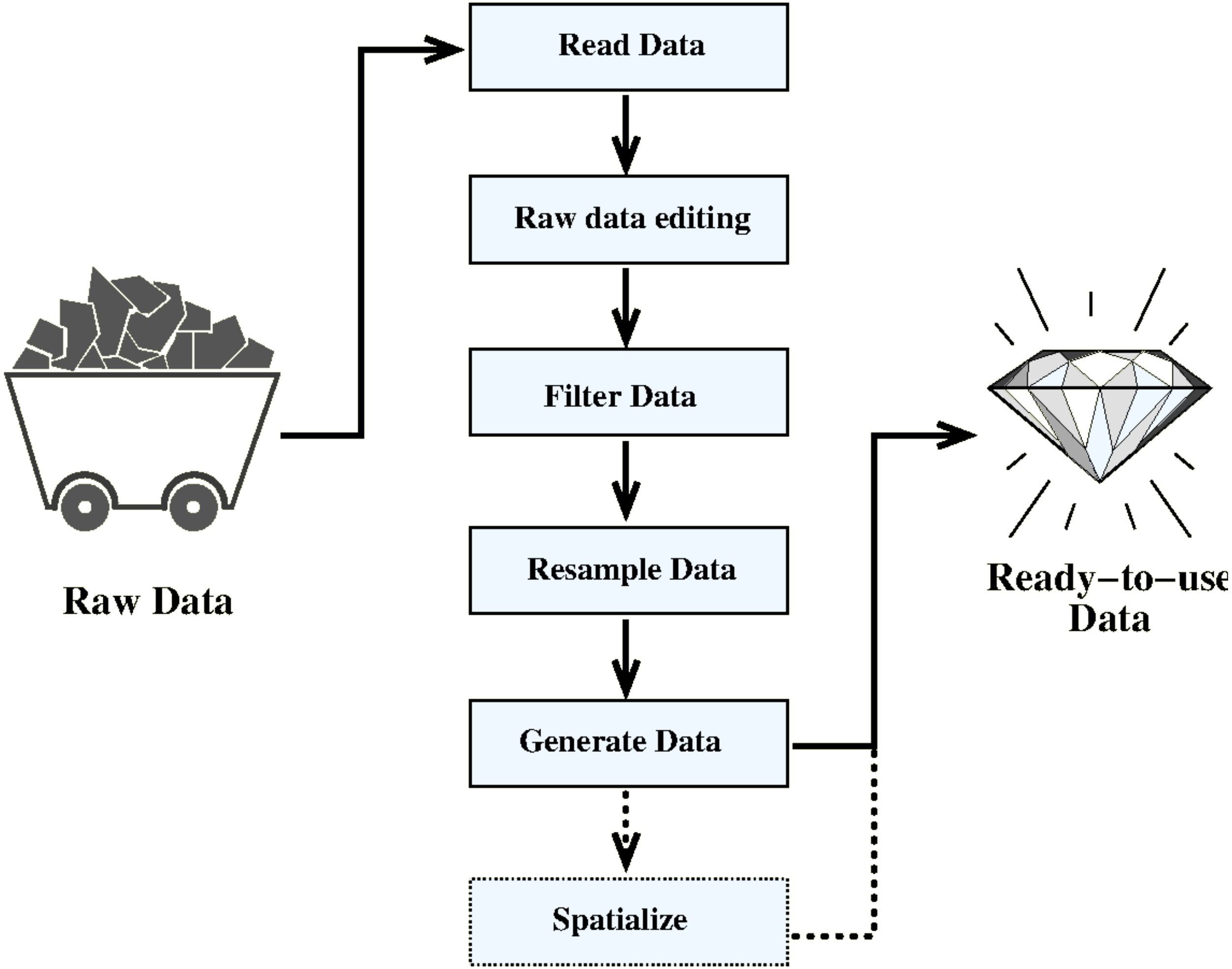
- Data read in native format
- Data converted to standard compliant NetCDF CF1.6 with ACDD metadata
- Standard field names
- Standard metadata
- Standard search metadata
- NetCDF/CF served through OPeNDAP
  - i.e. FAIR data and services

The magic...

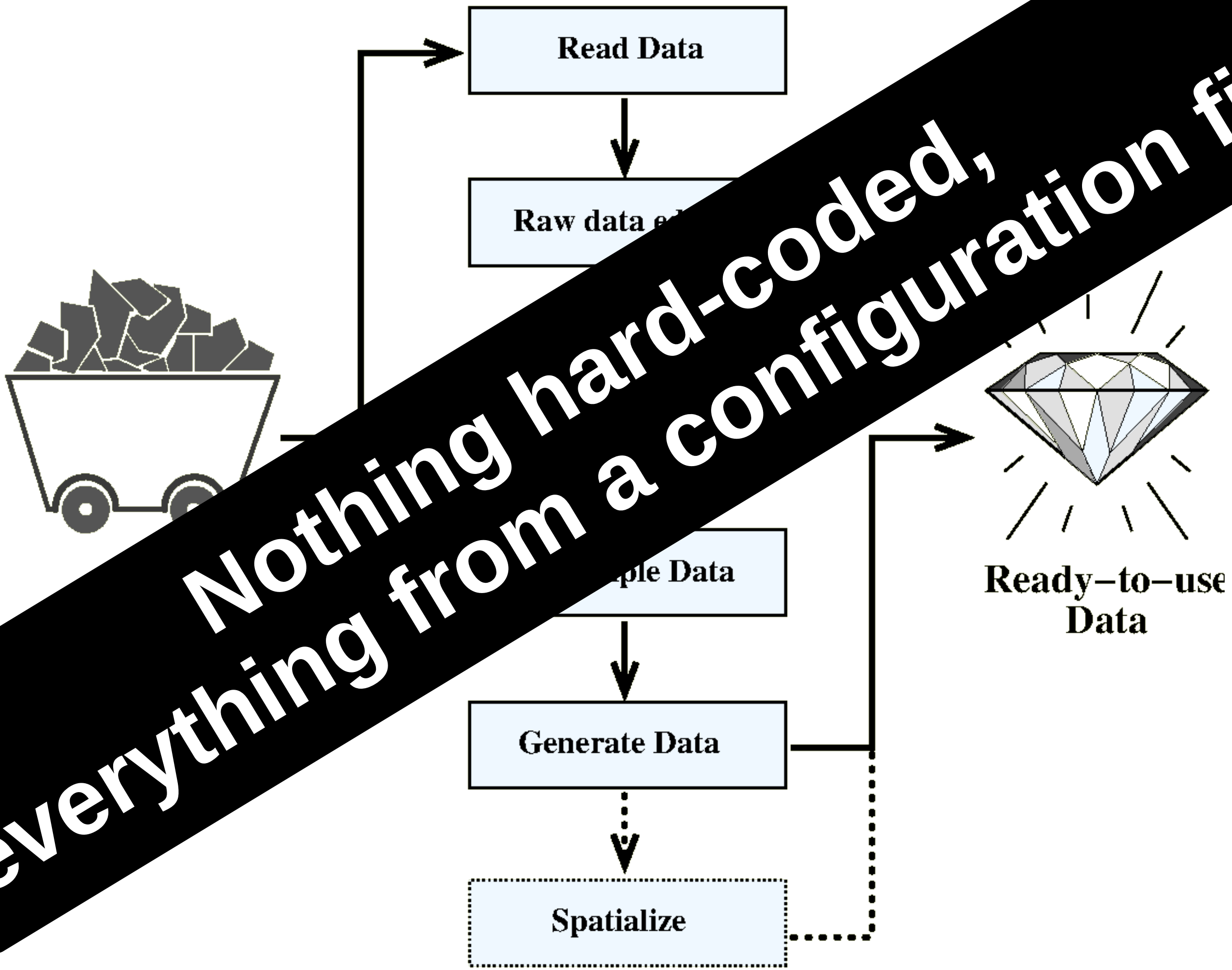


- The MeteoIO library is the engine that reads the native data
- MeteoOI does name mapping, units conversions, merging, time corrections, filtering...
- MeteoIO writes the data back to NetCDF/CF
- Data are served through pyDAP

MeteoIO’s Workflow



MeteoIO’s Workflow



Read Data

Raw data e

ple Data

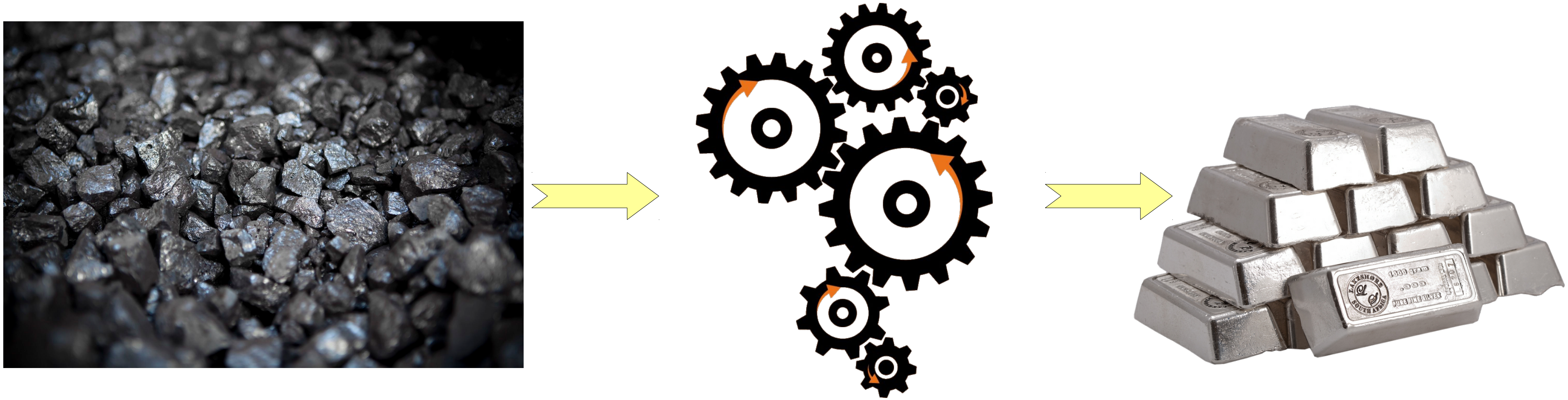
Generate Data

Spatialize

Ready–to–use Data

Nothing hard-coded, everything from a configuration file

DAME: Data standardization



Raw data,
vast variety
of formats

Read with different plugins
Process to standard levels
Add search metadata
Write with a common plugin

Standardised
product

DAME: Data standardization

Raw data
vast vari
of formats

erent plugins
to standard levels
dd search metadata
rite with a common plugin

Standardised
product

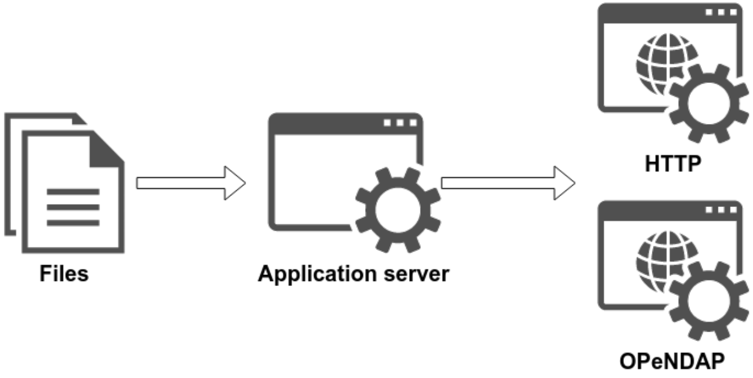
Backup configuration file
Implies traceability

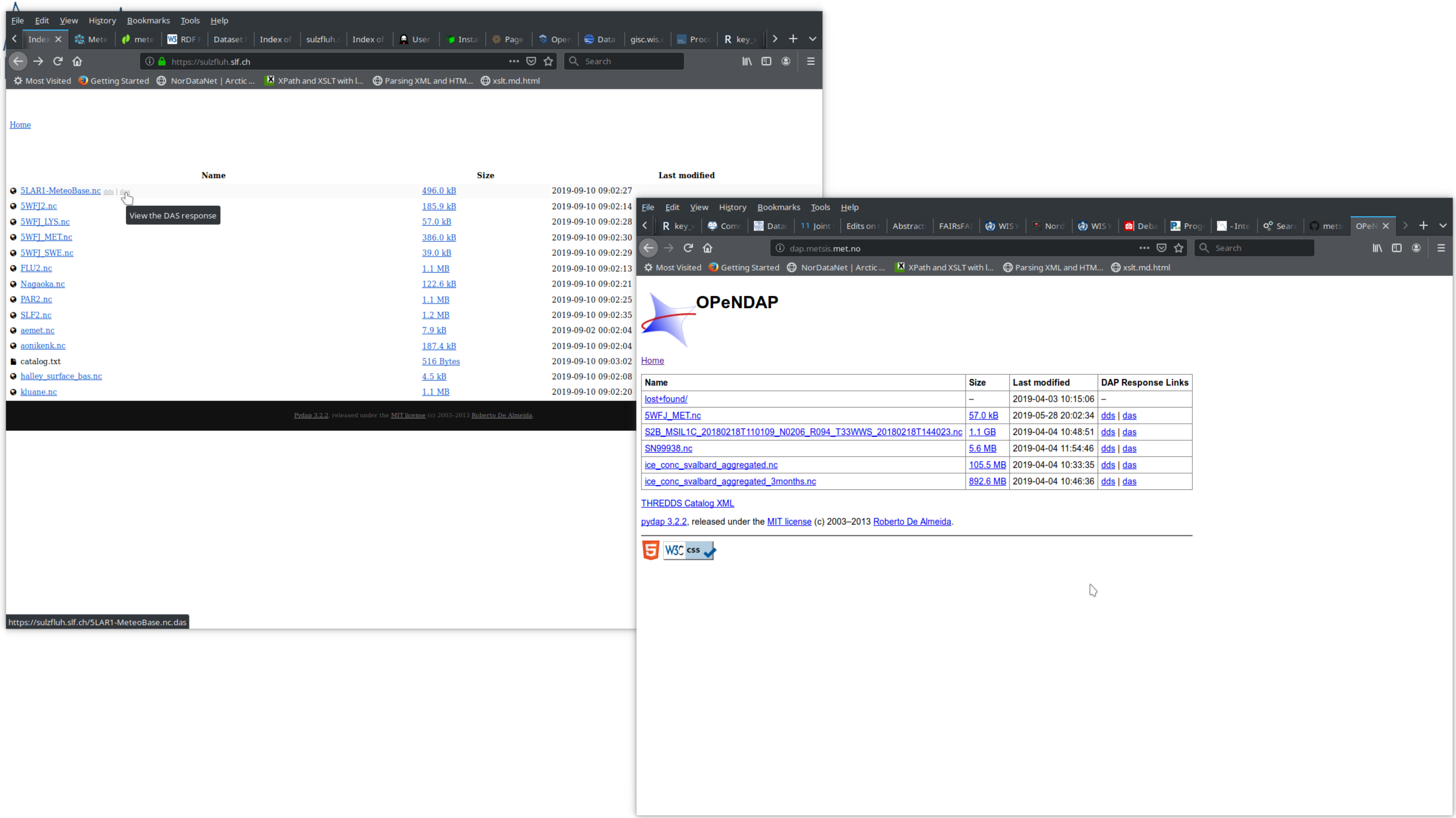
Summary on MeteoIO

- Not only meteorological data
- For each station, a configuration file (a few lines long)
- A small script calls MeteoIO on all configuration files at regular intervals and copies the resulting NetCDF/CF files to the OPeNDAP server
- Further reading
  - Bavay, M., and T. Egger. *"MeteoIO 2.4. 2: a preprocessing library for meteorological data."* Geoscientific Model Development (2014).
  - Get MeteoIO at http://models.slf.ch

Serving data through web services

- OPeNDAP allows data streaming and integration directly in analysis tools like R, Python and Matlab
- Using the lightweight pyDAP library/application
    - Unicode decoding issues are fixed when dealing with NetCDF files with Unicode characters in the metadata (this involved fix in both pyDAP and one of its dependencies webob)
    - Fixing Key-Value issues when dealing with NetCDF where the dimension are not listed also as variables [dims not in vars]
    - Adding docker support to serve data through externally mounted docker volume - pyDAP running as Apache WSGI
    - Adding script to generate Debian packages for both webob and pyDAP libraries.
        - The build of packages is automated when building the docker image.





Home

Name	Size	Last modified
5LAR1-MeteoBase.nc dds | das	496.0 kB	2019-09-10 09:02:27
5WFJ2.nc	185.9 kB	2019-09-10 09:02:14
5WFJ_LYS.nc	57.0 kB	2019-09-10 09:02:28
5WFJ_MET.nc	386.0 kB	2019-09-10 09:02:30
5WFJ_SWE.nc	39.0 kB	2019-09-10 09:02:29
FLU2.nc	1.1 MB	2019-09-10 09:02:13
Nagaoka.nc	122.6 kB	2019-09-10 09:02:21
PAR2.nc	1.1 MB	2019-09-10 09:02:25
SLF2.nc	1.2 MB	2019-09-10 09:02:35
aemet.nc	7.9 kB	2019-09-02 00:02:04
aonikenk.nc	187.4 kB	2019-09-10 09:02:04
catalog.txt	516 Bytes	2019-09-10 09:03:02
halley_surface_bas.nc	4.5 kB	2019-09-10 09:02:08
kluane.nc	1.1 MB	2019-09-10 09:02:20

Pydap 3.2.2, released under the MIT license (c) 2003-2013 Roberto De Almeida.

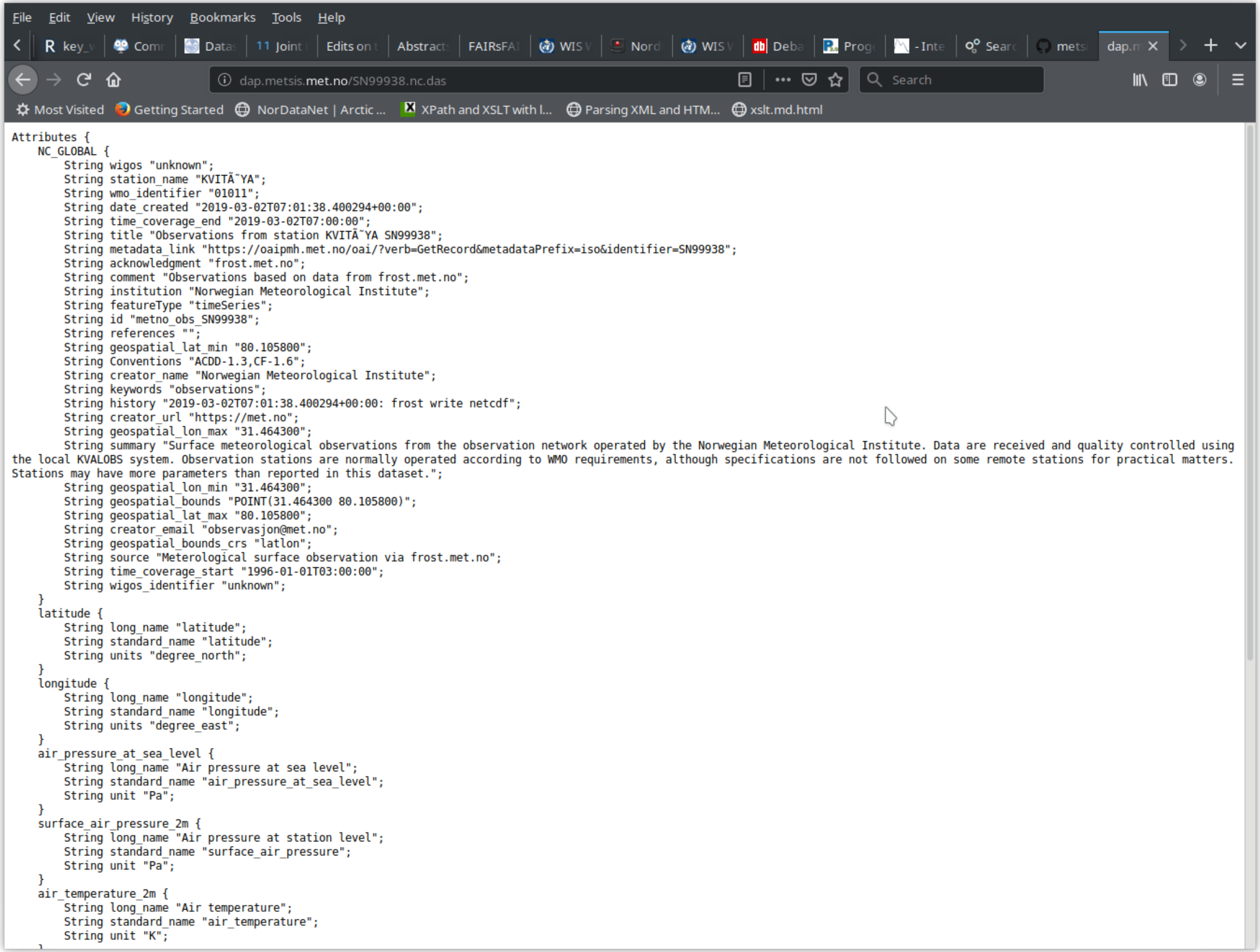
OPeNDAP

Home

Name	Size	Last modified	DAP Response Links
lost+found/	–	2019-04-03 10:15:06	–
5WFJ_MET.nc	57.0 kB	2019-05-28 20:02:34	dds | das
S2B_MSIL1C_20180218T110109_N0206_R094_T33WWS_20180218T144023.nc	1.1 GB	2019-04-04 10:48:51	dds | das
SN99938.nc	5.6 MB	2019-04-04 11:54:46	dds | das
ice_conc_svalbard_aggregated.nc	105.5 MB	2019-04-04 10:33:35	dds | das
ice_conc_svalbard_aggregated_3months.nc	892.6 MB	2019-04-04 10:46:36	dds | das

THREDDS Catalog XML

pydap 3.2.2, released under the MIT license (c) 2003–2013 Roberto De Almeida.



```
Attributes {
    NC_GLOBAL {
        String wigos "unknown";
        String station_name "KVITÃ˜YA";
        String wmo_identifier "01011";
        String date_created "2019-03-02T07:01:38.400294+00:00";
        String time_coverage_end "2019-03-02T07:00:00";
        String title "Observations from station KVITÃ˜YA SN99938";
        String metadata_link "https://oaipmh.met.no/oai/?verb=GetRecord&metadataPrefix=iso&identifier=SN99938";
        String acknowledgment "frost.met.no";
        String comment "Observations based on data from frost.met.no";
        String institution "Norwegian Meteorological Institute";
        String featureType "timeSeries";
        String id "metno_obs_SN99938";
        String references "";
        String geospatial_lat_min "80.105800";
        String Conventions "ACDD-1.3,CF-1.6";
        String creator_name "Norwegian Meteorological Institute";
        String keywords "observations";
        String history "2019-03-02T07:01:38.400294+00:00: frost write netcdf";
        String creator_url "https://met.no";
        String geospatial_lon_max "31.464300";
        String summary "Surface meteorological observations from the observation network operated by the Norwegian Meteorological Institute. Data are received and quality controlled using
the local KVALOBS system. Observation stations are normally operated according to WMO requirements, although specifications are not followed on some remote stations for practical matters.
Stations may have more parameters than reported in this dataset.";
        String geospatial_lon_min "31.464300";
        String geospatial_bounds "POINT(31.464300 80.105800)";
        String geospatial_lat_max "80.105800";
        String creator_email "observasjon@met.no";
        String geospatial_bounds_crs "latlon";
        String source "Meterological surface observation via frost.met.no";
        String time_coverage_start "1996-01-01T03:00:00";
        String wigos_identifier "unknown";
    }
    latitude {
        String long_name "latitude";
        String standard_name "latitude";
        String units "degree_north";
    }
    longitude {
        String long_name "longitude";
        String standard_name "longitude";
        String units "degree_east";
    }
    air_pressure_at_sea_level {
        String long_name "Air pressure at sea level";
        String standard_name "air_pressure_at_sea_level";
        String unit "Pa";
    }
    surface_air_pressure_2m {
        String long_name "Air pressure at station level";
        String standard_name "surface_air_pressure";
        String unit "Pa";
    }
    air_temperature_2m {
        String long_name "Air temperature";
        String standard_name "air_temperature";
        String unit "K";
    }
```

Summary

- MeteoIO for transformation to FAIR and quality control of sensor data
- PyDAP for public access to data through interoperability interfaces
- Need for online transformation services?