Metadata Templates and Gap Handling

M. Fiebig
NILU – Norsk Institutt for Luftforskning
L. Ferrighi
Norwegian Meteorological Institute

NorDataNet
Norwegian Scientific Data Network

Various Metadata Standards Exist: Examples

Dublin Core Element Set

- Emphasis on web resources, publications
- http://dublincore.org/documents/dces/

Darwin Core

- Emphasis on museum specimens & biodiversity
- http://rs.tdwg.org/dwc/index.htm

ISO 19115/19139 Geographic information – metadata

- Emphasis on geospatial data and services
- https://www.fgdc.gov/metadata/iso-standards

EU INSPIRE profile for ISO 19115/19139

- Infrastructure for Spatial Information in the European Community (INSPIRE)

…

NorDataNet
Norwegian Scientific Data Network

Choosing a Metadata Standard

- Choose the appropriate repository first
- Especially for topic repositories, these will tell you which metadata items (concepts) to report, and how (vocabulary).
- Important is which pieces of information are needed, not so much format.
- Topic repositories will often collect more metadata items than contained in the standards because of scientific need.
- Metadata standards are often chosen by use case. It is the repositories job to convert your metadata to the appropriate standard for a use case.
- Your organization might have policies on which repositories to use.
- Choose your metadata standard before measuring!

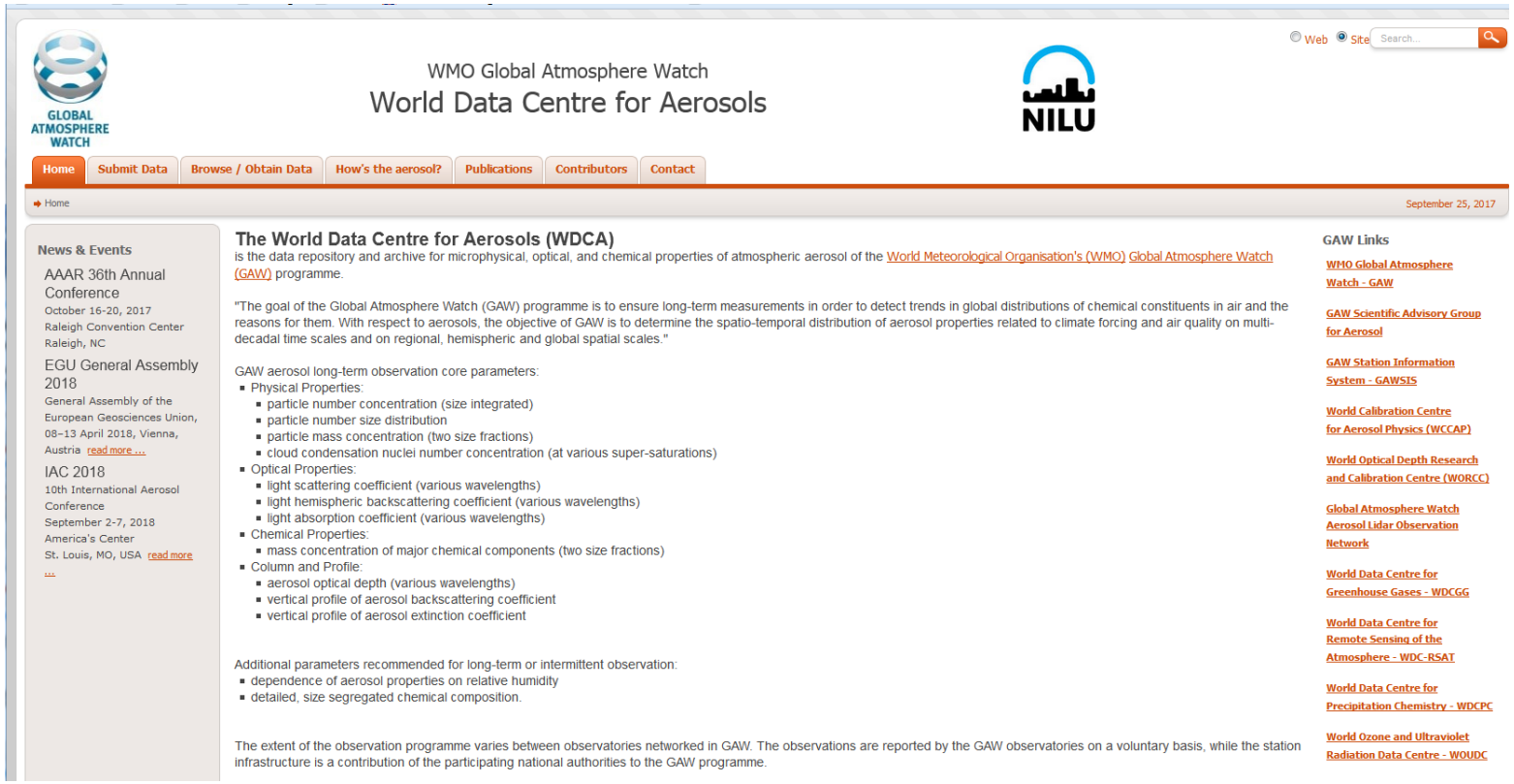
NorDataNet
Norwegian Scientific Data Network

Example 1: Nansen legacy Darwin Core Template Generator

- Darwin core developed for museum specimens.
- Can also be used for biodiversity samples.
- Home: https://dwc.tdwg.org/
- NorDataNet / SIOS template generator: https://www.nordatanet.no/cgi-bin/darwinsheet/?setup=aen
- Reference of terms: https://dwc.tdwg.org/terms/

NorDataNet
Norwegian Scientific Data Network

Example 2: WMO GAW World Data Centre for Aerosol



https://www.gaw-wdca.org/

NorDataNet
Norwegian Scientific Data Network

Observations with Reporting Support

Regular / Advanced (traceable):

- Particle number concentration
- Particle number size distribution (sub-micron)
- Cloud Condensation Particle Number Conc. / Size dist.
- Scattering Coefficient
- Absorption Coefficient

Regular only:

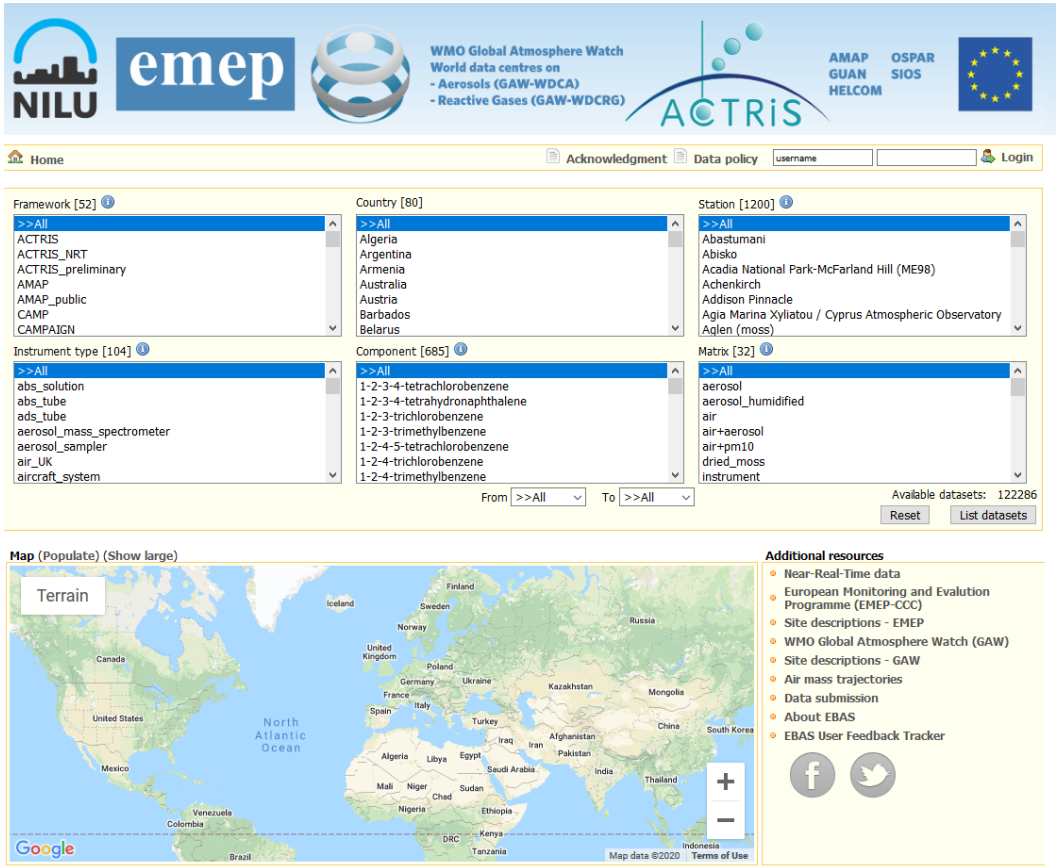
- Aerosol optical depth
- PM mass (gravimetric)
- PM mass (online)
- Aerosol Chemical Composition (GAW standard)
- Aerosol Chemical Speciation (online, AMS / ACSM)
- Particle number size distribution (super-micron, OPC, APS)

Further:

- Met. Base paramaters

NorDataNet
Norwegian Scientific Data Network

The EBAS / WDCA Web-Interface 1 / 3



http://ebas.nilu.no

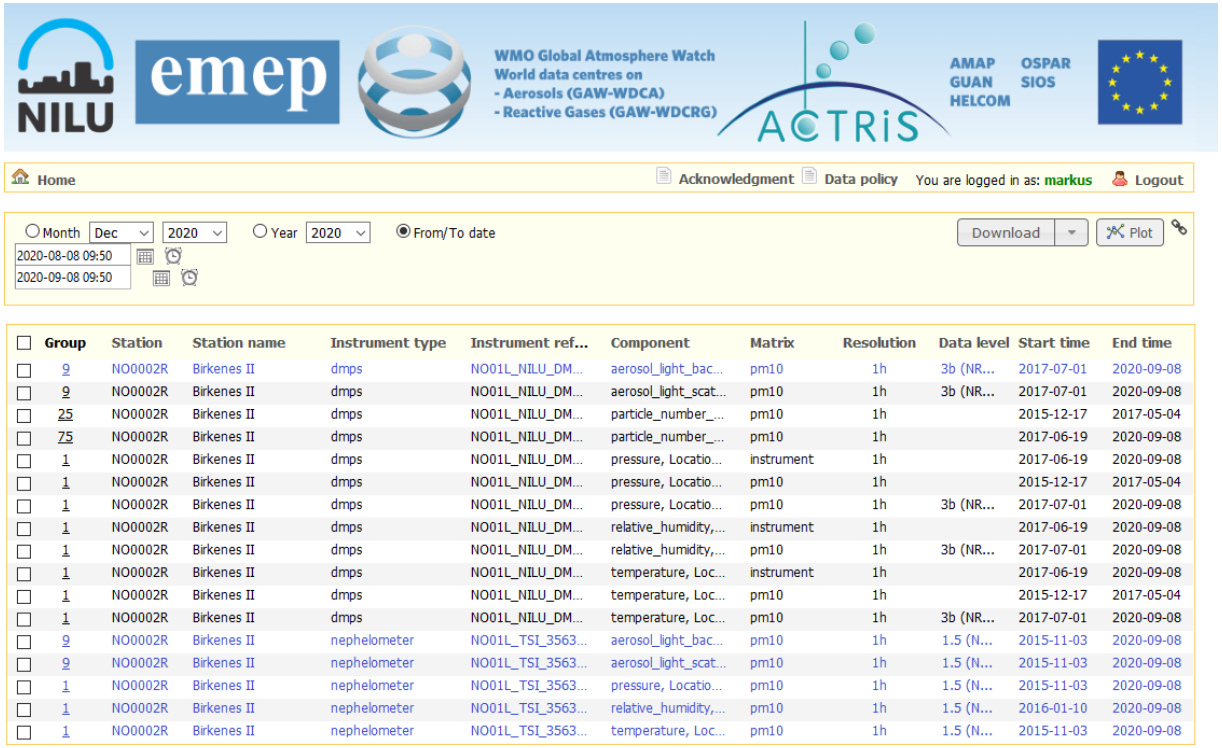
EBAS web-interface functions:

- Search datasets by criteria: Framework, country, station, matrix, instrument type, component.
- Visualise distribution of stations on map.
- Manage access restricted data.
- Links to other resources, e.g. trajectory calculations for station

- Plot, browse, compare datasets
- Download data.

NorDataNet
Norwegian Scientific Data Network

The EBAS Web-Interface 2 / 3



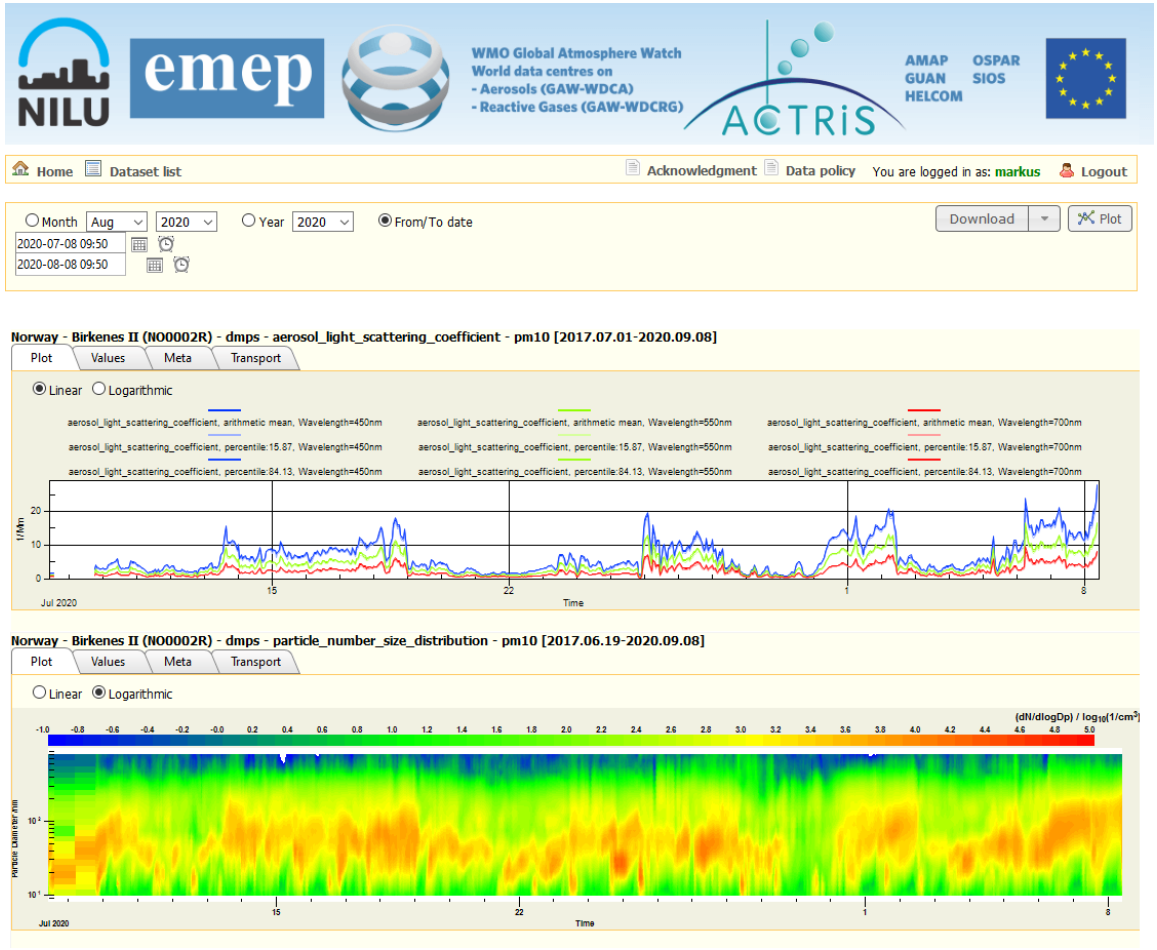
Group	Station	Station name	Instrument type	Instrument ref...	Component	Matrix	Resolution	Data level	Start time	End time
9	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	aerosol_light_bac...	pm10	1h	3b (NR...	2017-07-01	2020-09-08
9	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	aerosol_light_scat...	pm10	1h	3b (NR...	2017-07-01	2020-09-08
25	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	particle_number_...	pm10	1h		2015-12-17	2017-05-04
75	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	particle_number_...	pm10	1h		2017-06-19	2020-09-08
1	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	pressure, Locatio...	instrument	1h		2017-06-19	2020-09-08
1	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	pressure, Locatio...	pm10	1h		2015-12-17	2017-05-04
1	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	pressure, Locatio...	pm10	1h	3b (NR...	2017-07-01	2020-09-08
1	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	relative_humidity,...	instrument	1h		2017-06-19	2020-09-08
1	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	relative_humidity,...	pm10	1h	3b (NR...	2017-07-01	2020-09-08
1	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	temperature, Loc...	instrument	1h		2017-06-19	2020-09-08
1	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	temperature, Loc...	pm10	1h		2015-12-17	2017-05-04
1	NO0002R	Birkenes II	dmps	NO01L_NILU_DM...	temperature, Loc...	pm10	1h	3b (NR...	2017-07-01	2020-09-08
9	NO0002R	Birkenes II	nephelometer	NO01L_TSI_3563...	aerosol_light_bac...	pm10	1h	1.5 (N...	2015-11-03	2020-09-08
9	NO0002R	Birkenes II	nephelometer	NO01L_TSI_3563...	aerosol_light_scat...	pm10	1h	1.5 (N...	2015-11-03	2020-09-08
1	NO0002R	Birkenes II	nephelometer	NO01L_TSI_3563...	pressure, Locatio...	pm10	1h	1.5 (N...	2015-11-03	2020-09-08
1	NO0002R	Birkenes II	nephelometer	NO01L_TSI_3563...	relative_humidity,...	pm10	1h	1.5 (N...	2016-01-10	2020-09-08
1	NO0002R	Birkenes II	nephelometer	NO01L_TSI_3563...	temperature, Loc...	pm10	1h	1.5 (N...	2015-11-03	2020-09-08

Search result page of EBAS web-interface:

- Lists datsets that meet search criteria set on home page.
- Datasets that are present, but access restricted, are displayed in grey.
- Time period for plotting or download to be selected on top (select appropriate radio button!).

NorDataNet
Norwegian Scientific Data Network

The EBAS Web-Interface 3 / 3



Plot page for selected datasets:

- Screen, evaluate, compare between instruments, compare between stations, …
- Download datasets (data is automatcally grouped by instrument).

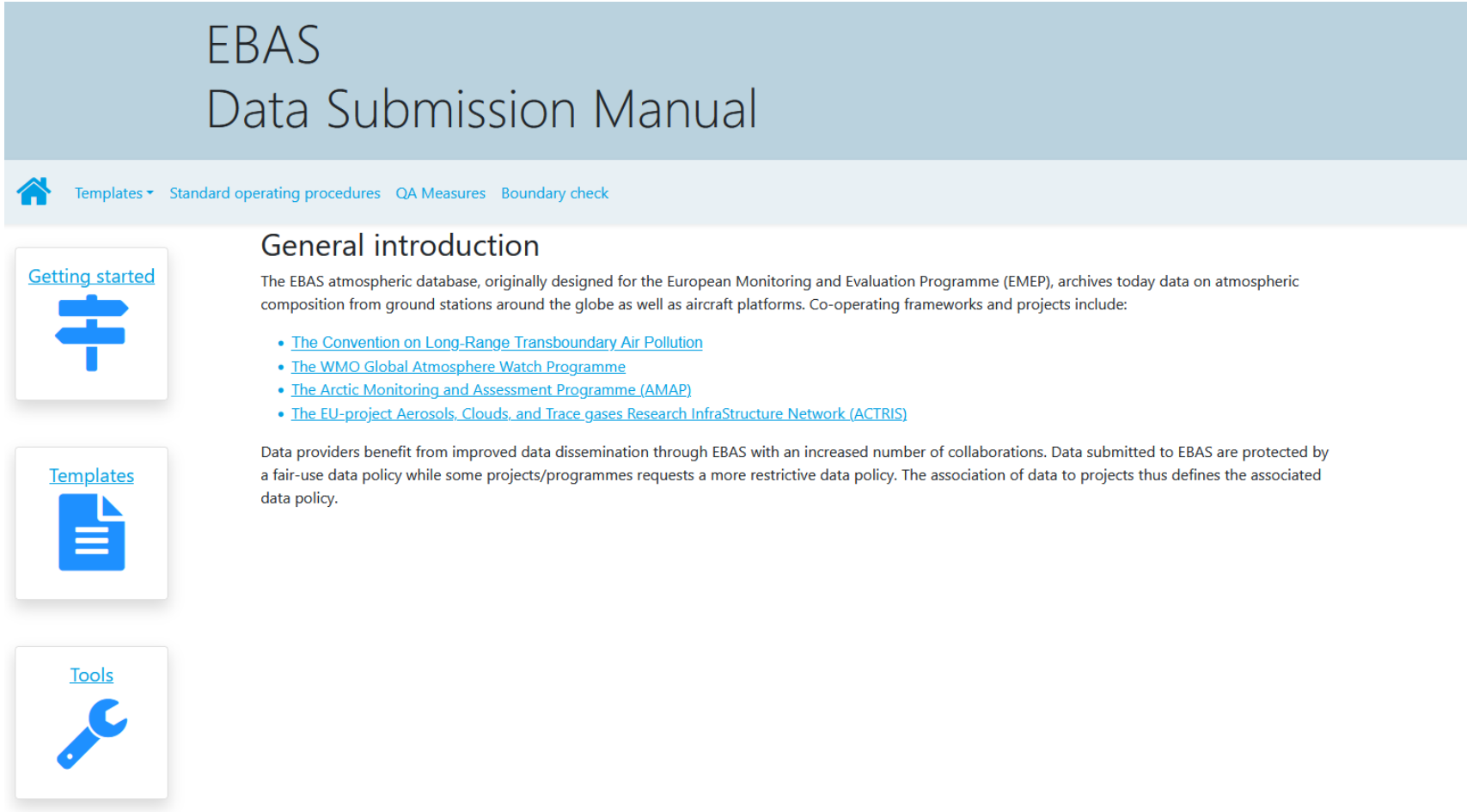
NorDataNet
Norwegian Scientific Data Network

Why Using NASA Ames 1001 Format for Reporting?

1. **Simplicity**
   - Pure ASCII text, human readable, readily opened or edited by simple means (any editor or spreadsheet application).
   - Explanation relatively short, yet contains necessary metadata.
2. **Reduce Format Confusion**
   - Don’t increase number of existing formats (NASA Ames, Narsto, NetCDF, HDF, …) unnecessarily as long as metadata can be transported in old format, even though some features are old-fashioned.
   - Existing libraries can be used to handle files.
3. **Keep threshold low**
   - More modern, binary formats exist (NetCDF, HDF), but need special editors and steep learning curve to assemble.
   - NASA Ames can be assembled with simple tools rather quickly.

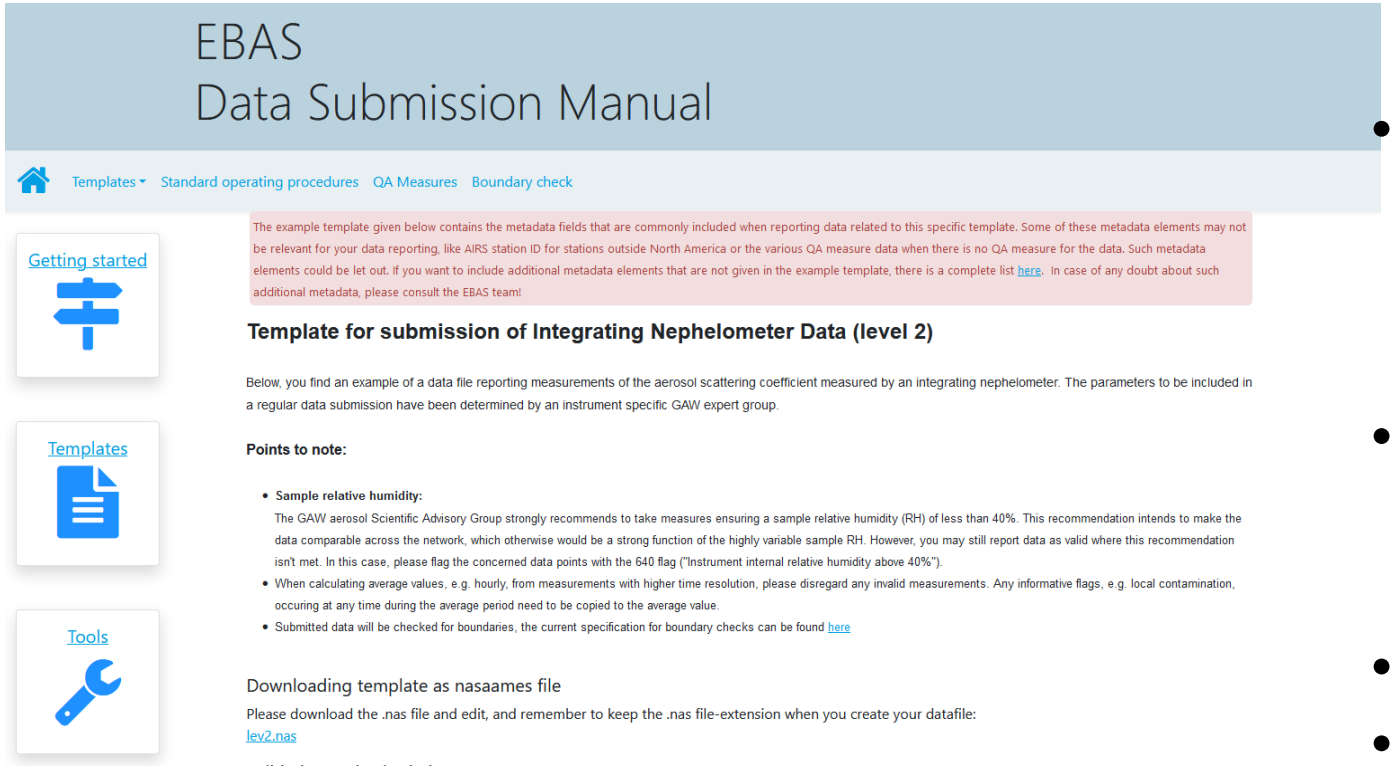
NorDataNet
Norwegian Scientific Data Network

Instructions for Submitting Aerosol Data to EBAS, i.e. WDCA: http://ebas-submit.nilu.no/



NorDataNet
Norwegian Scientific Data Network

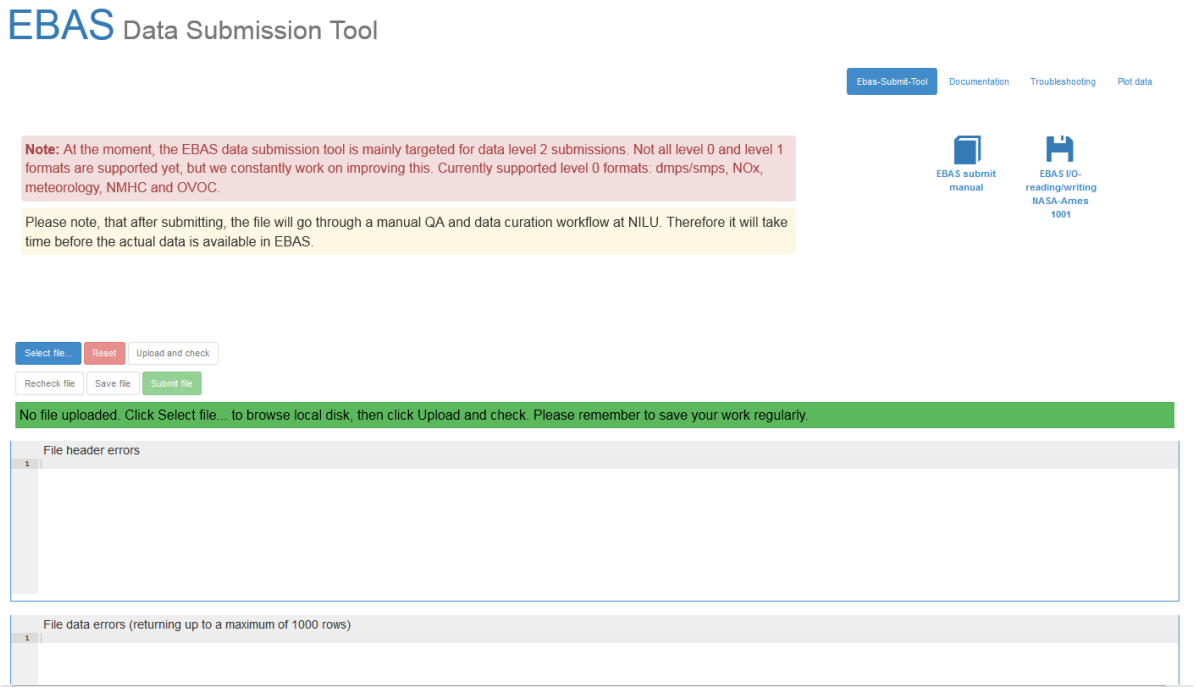
Find Template, Assemble Data in Submission Format



- Find template corresponding to your data, and copy header as a start.
- Go through the line-by-line instructions, and update content to your station and measurement, making sure to match syntax.
- Format the data section (white-space separated table, observe field length, missing values flag 0.999).
- Join header and data section
- Upload to submission portal

NorDataNet
Norwegian Scientific Data Network

EBAS Data Submission Portal: http://ebas-submit-tool.nilu.no/



- Dedicated web-page for checking data files before submission.
- Page shows uploaded test file with header / metadata errors highlighted and commented.
- Also data section of the file is checked.

NorDataNet
Norwegian Scientific Data Network

Gaps in Metadata: what if you don’t know or lack the words?

- Discovery and attribution metadata need to be present!
- Else, don’t state information you are uncertain about. Leave empty.
- Avoid missing information by planning. Check needed metadata before measurement.
- Try to augment vocabulary you are using. Work with data repository!

NorDataNet
Norwegian Scientific Data Network

Augmenting vocabulary (if you need): CF Standard Names

- Vocabularies inspire each other.
- CF standard names have drawn from GCMD.
- Do request new standard names, send request to mailing list: http://mailman.cgd.ucar.edu/mailman/listinfo/cf-metadata
- Count in a few weeks – months before request is discussed and adopted.

NorDataNet
Norwegian Scientific Data Network