

Metadata Templates and Gap Handling

M. Fiebig

NILU – Norsk Institutt for Luftforskning

L. Ferrighi

Norwegian Meteorological Institute

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)

...

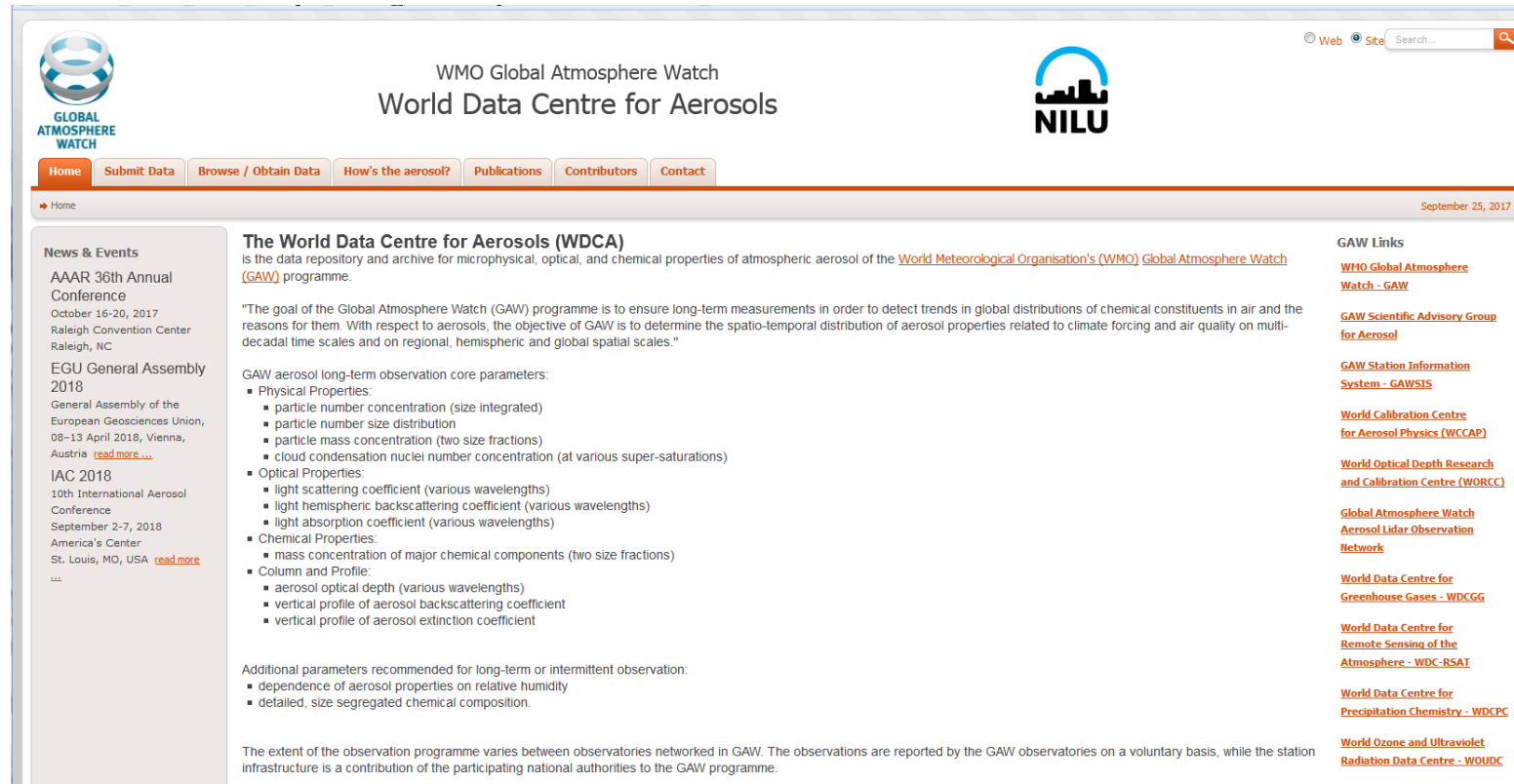
Choosing a Metadata Standard

- 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!

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/>

Example 2: WMO GAW World Data Centre for Aerosol



The screenshot shows the homepage of the WMO Global Atmosphere Watch World Data Centre for Aerosols. The header features the WMO logo, the title "WMO Global Atmosphere Watch World Data Centre for Aerosols", and the NILU logo. A navigation bar includes links for Home, Submit Data, Browse / Obtain Data, How's the aerosol?, Publications, Contributors, and Contact. The main content area is divided into three columns. The left column, "News & Events", lists recent conferences like the AAAR 36th Annual Conference and the EGU General Assembly 2018. The middle column, "The World Data Centre for Aerosols (WDCA)", describes the data repository and lists GAW aerosol long-term observation core parameters, including Physical Properties (particle number concentration, size distribution, mass concentration, cloud condensation nuclei), Optical Properties (light scattering, backscattering, absorption coefficients), Chemical Properties (mass concentration of major chemical components), and Column and Profile (aerosol optical depth, backscattering, extinction coefficients). It also mentions additional recommended parameters like dependence on relative humidity and detailed chemical composition. The right column, "GAW Links", provides links to various GAW-related resources, including the WMO Global Atmosphere Watch - GAW, GAW Scientific Advisory Group for Aerosol, GAW Station Information System - GAWSYS, World Calibration Centre for Aerosol Physics (WCCAP), World Optical Depth Research and Calibration Centre (WORCC), Global Atmosphere Watch Aerosol Lidar Observation Network, World Data Centre for Greenhouse Gases - WDCGG, World Data Centre for Remote Sensing of the Atmosphere - WDC-RSAT, World Data Centre for Precipitation Chemistry - WDCPC, and World Ozone and Ultraviolet Radiation Data Centre - WOUDC. The date "September 25, 2017" is displayed in the top right corner.

WMO Global Atmosphere Watch
World Data Centre for Aerosols

Home Submit Data Browse / Obtain Data How's the aerosol? Publications Contributors Contact

Home September 25, 2017

News & Events

AAAR 36th Annual Conference
October 16-20, 2017
Raleigh Convention Center
Raleigh, NC

EGU General Assembly 2018
General Assembly of the European Geosciences Union, 08-13 April 2018, Vienna, Austria [read more ...](#)

IAC 2018
10th International Aerosol Conference
September 2-7, 2018
America's Center
St. Louis, MO, USA [read more ...](#)

The World Data Centre for Aerosols (WDCA)
is the data repository and archive for microphysical, optical, and chemical properties of atmospheric aerosol of the [World Meteorological Organisation's \(WMO\) Global Atmosphere Watch \(GAW\)](#) programme.

"The goal of the Global Atmosphere Watch (GAW) programme is to ensure long-term measurements in order to detect trends in global distributions of chemical constituents in air and the reasons for them. With respect to aerosols, the objective of GAW is to determine the spatio-temporal distribution of aerosol properties related to climate forcing and air quality on multi-decadal time scales and on regional, hemispheric and global spatial scales."

GAW aerosol long-term observation core parameters:

- Physical Properties:
 - particle number concentration (size integrated)
 - particle number size distribution
 - particle mass concentration (two size fractions)
 - cloud condensation nuclei number concentration (at various super-saturations)
- Optical Properties:
 - light scattering coefficient (various wavelengths)
 - light hemispheric backscattering coefficient (various wavelengths)
 - light absorption coefficient (various wavelengths)
- Chemical Properties:
 - mass concentration of major chemical components (two size fractions)
- Column and Profile:
 - aerosol optical depth (various wavelengths)
 - vertical profile of aerosol backscattering coefficient
 - vertical profile of aerosol extinction coefficient

Additional parameters recommended for long-term or intermittent observation:

- dependence of aerosol properties on relative humidity
- detailed, size segregated chemical composition.

The extent of the observation programme varies between observatories networked in GAW. The observations are reported by the GAW observatories on a voluntary basis, while the station infrastructure is a contribution of the participating national authorities to the GAW programme.

GAW Links

[WMO Global Atmosphere Watch - GAW](#)

[GAW Scientific Advisory Group for Aerosol](#)

[GAW Station Information System - GAWSYS](#)

[World Calibration Centre for Aerosol Physics \(WCCAP\)](#)

[World Optical Depth Research and Calibration Centre \(WORCC\)](#)

[Global Atmosphere Watch Aerosol Lidar Observation Network](#)

[World Data Centre for Greenhouse Gases - WDCGG](#)

[World Data Centre for Remote Sensing of the Atmosphere - WDC-RSAT](#)

[World Data Centre for Precipitation Chemistry - WDCPC](#)

[World Ozone and Ultraviolet Radiation Data Centre - WOUDC](#)

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

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 parameters

The EBAS / WDCA Web-Interface 1 / 3

The screenshot displays the EBAS / WDCA Web-Interface. At the top, there is a header with logos for NILU, emep, WMO Global Atmosphere Watch, ACTRIS, AMAP, OSPAR, HELCOM, and the European Union. Below the header, there is a navigation bar with links for Home, Acknowledgment, Data policy, a username field, and a Login button. The main content area features several search filters: Framework [52], Country [80], Station [1200], Instrument type [104], Component [685], and Matrix [32]. Each filter has a dropdown menu with a list of options. Below the filters, there is a 'From' and 'To' date range selector, a 'Reset' button, and a 'List datasets' button. The bottom section of the interface includes a map titled 'Map (Populate) (Show large)' showing the distribution of stations across the world, with a 'Terrain' button and a 'Google' logo. To the right of the map, there is an 'Additional resources' section with links to various data sources and a social media section with Facebook and Twitter icons.

<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.

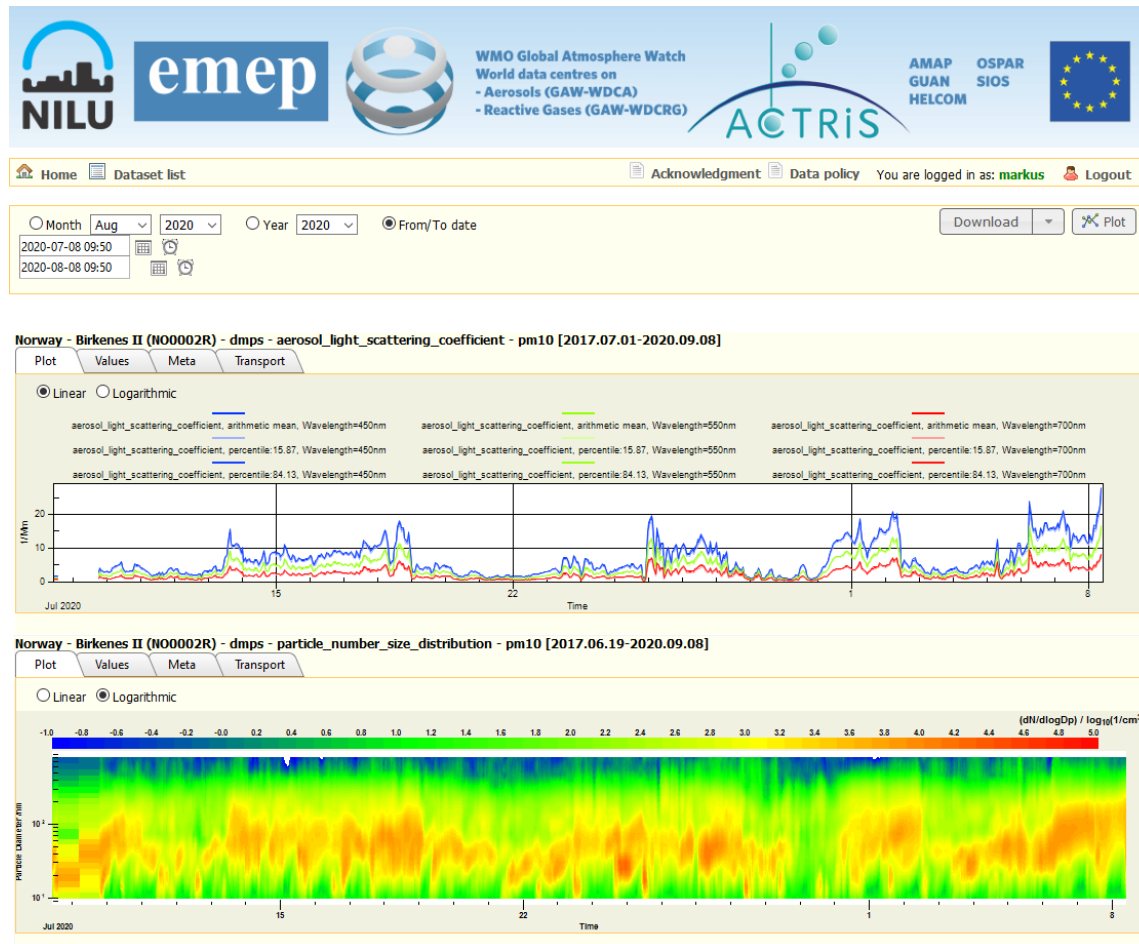
The EBAS Web-Interface 2 / 3

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

The EBAS Web-Interface 3 / 3



Plot page for selected datasets:

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

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.


Instructions for Submitting Aerosol Data to EBAS, i.e. WDCA:

<http://ebas-submit.nilu.no/>


EBAS Data Submission Manual

 [Templates](#) ▾ [Standard operating procedures](#) [QA Measures](#) [Boundary check](#)


[Getting started](#)



[Templates](#)



[Tools](#)



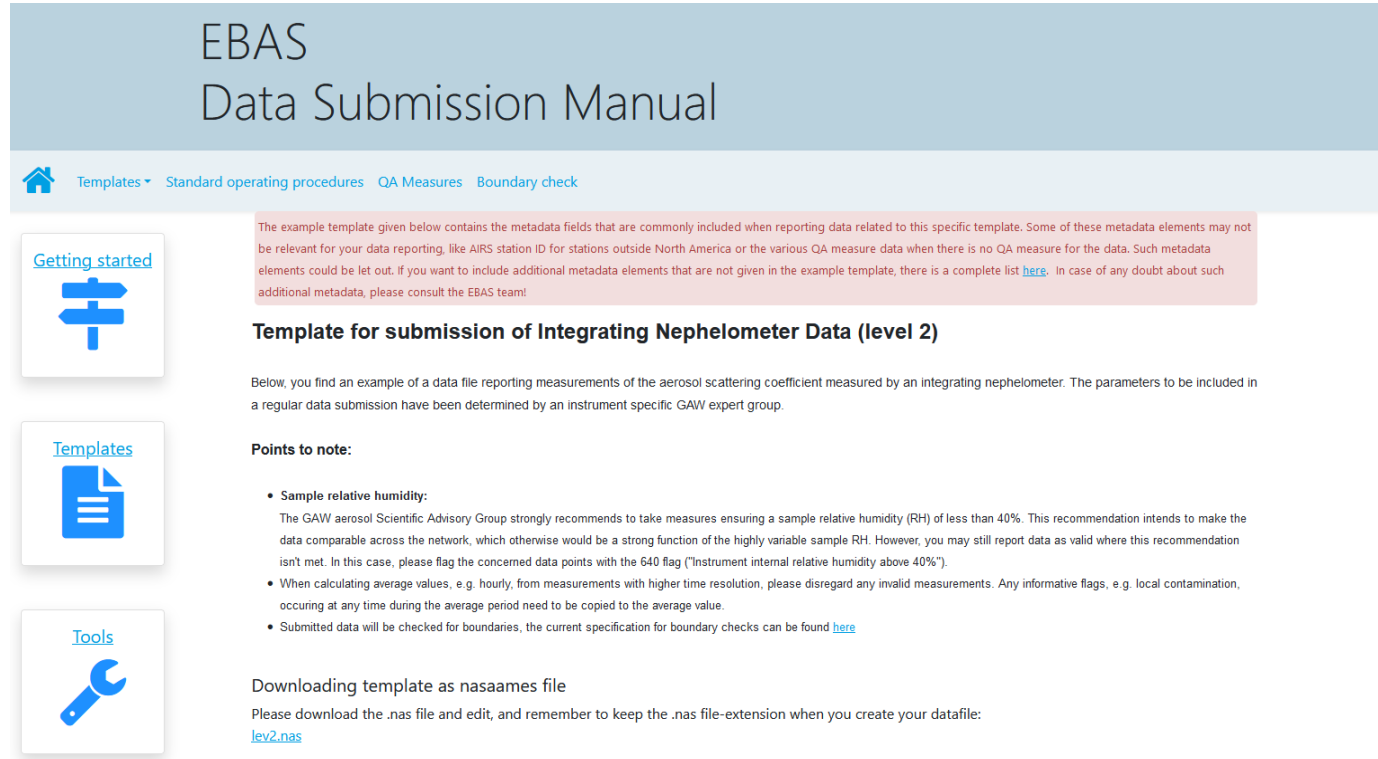
General introduction

The EBAS atmospheric database, originally designed for the European Monitoring and Evaluation Programme (EMEP), archives today data on atmospheric composition from ground stations around the globe as well as aircraft platforms. Co-operating frameworks and projects include:

- [The Convention on Long-Range Transboundary Air Pollution](#)
- [The WMO Global Atmosphere Watch Programme](#)
- [The Arctic Monitoring and Assessment Programme \(AMAP\)](#)
- [The EU-project Aerosols, Clouds, and Trace gases Research InfraStructure Network \(ACTRIS\)](#)

Data providers benefit from improved data dissemination through EBAS with an increased number of collaborations. Data submitted to EBAS are protected by a fair-use data policy while some projects/programmes requests a more restrictive data policy. The association of data to projects thus defines the associated data policy.

Find Template, Assemble Data in Submission Format



The screenshot shows the EBAS Data Submission Manual website. The header is light blue with the text "EBAS Data Submission Manual". Below the header is a navigation bar with a home icon and links: "Templates", "Standard operating procedures", "QA Measures", and "Boundary check". On the left side, there are three icons: "Getting started" (a blue arrow pointing right), "Templates" (a blue document icon), and "Tools" (a blue wrench icon). The main content area has a pink box with text: "The example template given below contains the metadata fields that are commonly included when reporting data related to this specific template. Some of these metadata elements may not be relevant for your data reporting, like AIRS station ID for stations outside North America or the various QA measure data when there is no QA measure for the data. Such metadata elements could be let out. If you want to include additional metadata elements that are not given in the example template, there is a complete list [here](#). In case of any doubt about such additional metadata, please consult the EBAS team!". Below this is the section "Template for submission of Integrating Nephelometer Data (level 2)". The text below this section says: "Below, you find an example of a data file reporting measurements of the aerosol scattering coefficient measured by an integrating nephelometer. The parameters to be included in a regular data submission have been determined by an instrument specific GAW expert group." Under "Points to note:", there are three bullet points: "Sample relative humidity: The GAW aerosol Scientific Advisory Group strongly recommends to take measures ensuring a sample relative humidity (RH) of less than 40%. This recommendation intends to make the data comparable across the network, which otherwise would be a strong function of the highly variable sample RH. However, you may still report data as valid where this recommendation isn't met. In this case, please flag the concerned data points with the 640 flag ("Instrument internal relative humidity above 40%").", "When calculating average values, e.g. hourly, from measurements with higher time resolution, please disregard any invalid measurements. Any informative flags, e.g. local contamination, occurring at any time during the average period need to be copied to the average value.", and "Submitted data will be checked for boundaries, the current specification for boundary checks can be found [here](#)". At the bottom, there is a section "Downloading template as nasaames file" with the text: "Please download the .nas file and edit, and remember to keep the .nas file-extension when you create your datafile: [lev2.nas](#)".

- 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

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

EBAS Data Submission Tool

The screenshot shows the EBAS Data Submission Tool interface. At the top, there are navigation links: "Ebas-Submit-Tool", "Documentation", "Troubleshooting", and "Plot data". Below these, there are two icons: "EBAS submit manual" and "EBAS I/O-reading/writing NASA-Ames 1001". A red note box states: "Note: At the moment, the EBAS data submission tool is mainly targeted for data level 2 submissions. Not all level 0 and level 1 formats are supported yet, but we constantly work on improving this. Currently supported level 0 formats: dmpps/smpps, NOx, meteorology, NMHC and OVOC." Below the note, a yellow box says: "Please note, that after submitting, the file will go through a manual QA and data curation workflow at NILU. Therefore it will take time before the actual data is available in EBAS." The main interface has a top bar with buttons: "Select file...", "Reset", "Upload and check", "Recheck file", "Save file", and "Submit file". Below this, a green message box says: "No file uploaded. Click Select file... to browse local disk, then click Upload and check. Please remember to save your work regularly." The interface also has two sections for errors: "File header errors" and "File data errors (returning up to a maximum of 1000 rows)". Both sections are currently empty.

- 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.

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!

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.

