

Metadata Templates and Gap Handling

M. Fiebig

NILU – Norsk Institutt for Luftforskning

L. Ferrighi

Norwegian Meteorological Institute

Multiple Metadata Standards Exist: Examples

Dublin Core Element Set

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

ISO 19115/19139 Geographic information – metadata

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

Darwin Core

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

...

Choosing a Metadata Standard

- Choose the appropriate repository first
- Especially for topic repositories, these will tell you which metadata to report, and how.
- 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

WMO Global Atmosphere Watch
World Data Centre for Aerosols

GLOBAL ATMOSPHERE WATCH

Web Size Search

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

Home September 25, 2017

News & Events

AAAR 30th Annual Conference
October 16-20, 2017
Kaleigh Convention Center
Raleigh, NC

EGU General Assembly 2018
General Assembly of the
European Geosciences Union,
17-19 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 wave-lengths)
 - light hemispheric backscattering coefficient (various wave-lengths)
 - light absorption coefficient (various wave-lengths)
- Chemical Properties:
 - mass concentration of major chemical components (two size fractions)
- Column and Profile:
 - aerosol optical depth (various wave-lengths)
 - 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 - GAW SIS](#)
- [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 - WDC-GG](#)
- [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 are logos for NILU, emep, WMO Global Atmosphere Watch, ACTRIS, AMAP, OSPAR, HELCOM, and the European Union. Below the logos is a navigation bar with links for Home, Acknowledgment, Data policy, a username input field, and a Login button. The main content area features six 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 are 'From' and 'To' dropdowns set to '>>All', a 'Reset' button, and a 'List datasets' button. The text 'Available datasets: 122286' is displayed. At the bottom left, there is a 'Map (Populate) (Show large)' section with a terrain map of Europe and surrounding regions. At the bottom right, there is an 'Additional resources' section with a list of links and social media icons for Facebook and Twitter.

<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

WMO Global Atmosphere Watch
World data centres on
- Aerosols (GAW-WDCA)
- Reactive Gases (GAW-WDCRG)

AMAP
GUAN
HELCOM

OSPAR
SIOS

ACTRIS

Home Acknowledgment Data policy You are logged in as: markus Logout

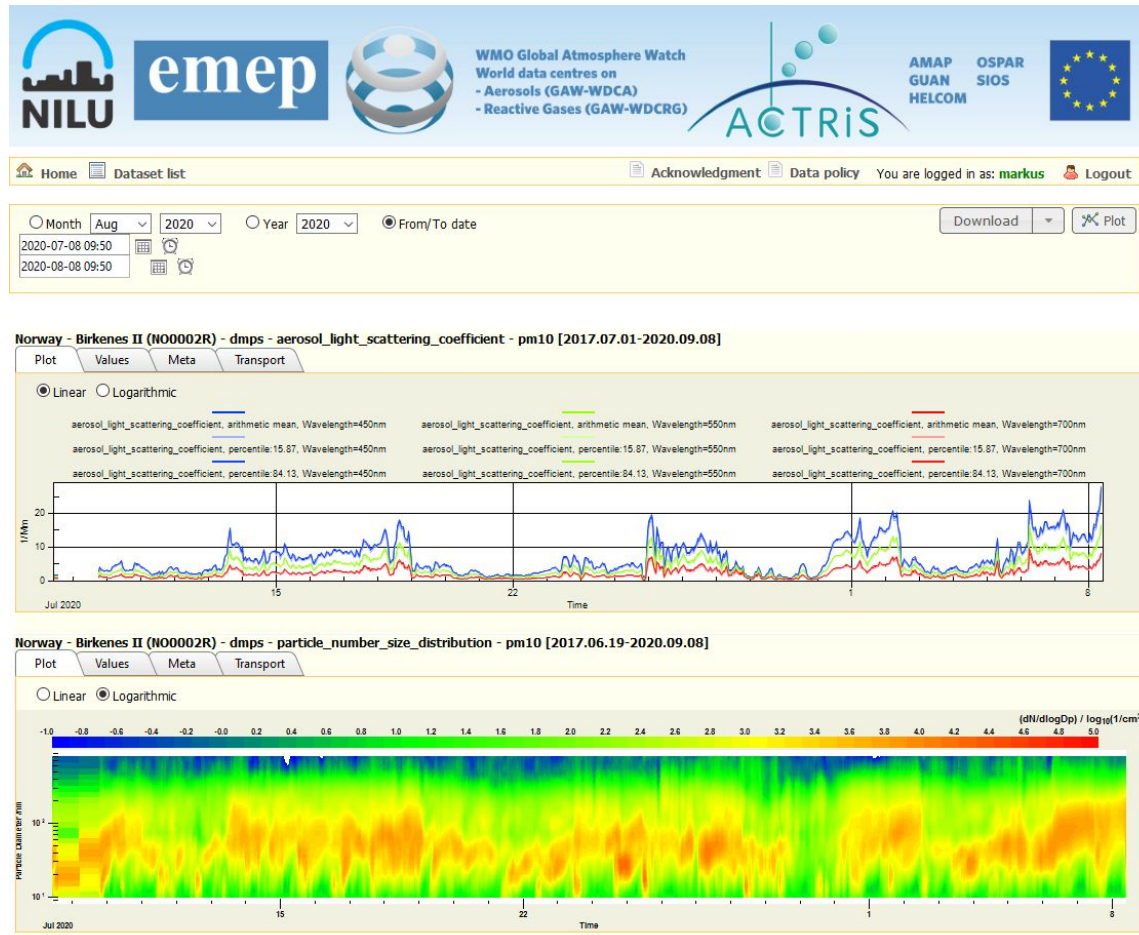
Month: Dec 2020 Year: 2020 From/To date: 2020-08-08 09:50 to 2020-09-08 09:50 Download Plot

<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.

Why Do We Ask Providers to Format the Data?

1. Avoid Errors

- Station is the authoritative source of metadata!
- Reformatting data and frequent iterations with provider induce misunderstanding and errors.

2. Scientific Standard of Provider

- Yearly submission is essence of a year's work, data (often) remains property of PI, they are responsible for the quality (policy dependent).

3. Work Load at Data Centre:

- EBAS (hosting WDCA) receives over 6000 datasets annually. Submitting formatted data frees resources for other services, e.g. dissemination.

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

EBAS
Data Submission Manual

September 25, 2017

General introduction

The EDAS 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 - UNEP](#)
- [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.

Submission Format

Data submitted to EBAS need to be formatted in the EBAS NASA Ames format by the data provider. The EBAS NASA Ames format is based on the ASCII text NASA-Ames 1001 format, but contains additional metadata specifications ensuring proper documentation, and is designed to be easily understandable (see [reasoning behind this setup](#)). This page provides links to data reporting templates for reporting data to EBAS.

Submission Procedure

The normal mode of submitting data to EDAS is the regular, annual data submission (see left margin, top). The deadline for a submission depends on the framework or project the dataset is associated with. EBAS also offers advanced data reporting that establishes complete traceability of the measurement and data analysis process. Participation in the advanced data reporting scheme is voluntary unless required by the associated project or framework. The usual steps for submitting data for the first time include:

1. **Initial contact with EBAS:** Please establish the initial contact with EBAS by writing an e-mail to ebas@nilu.no. In your mail, please indicate the station you are intending to report data for. If your station is part of the Global Atmosphere Watch programme, please also provide its GAW/SIS station ID (if you wish to register a new GAW site, please follow the instructions available at <https://gaw.wsl.ch>). See also the GAW Implementation Plan for further details: <http://www.wmo.int/pages/prog/era/cees/cew/documents/IPFinal11a1May11.pdf>
2. **IDs:** We will provide three further IDs: 1) the EDAS station code; 2) the EDAS platform code; 3) a code for your lab analysing the data, which you will need for the metadata in your submission. The three letter GAW IDs and the IDs used in the CLRTAP UNEP database EDAS were introduced independently and are maintained for consistency.
3. **Selection of time series to report:** For sites that have been in operation back in time, we encourage that the full historic time series is submitted to the database (multi-year submission is easy to accommodate, but bimetric needs to be split according to any significant changes with regards to instrumentation etc).

EBAS Data Submission Portal:

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

EBAS Data Submission Tool

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/smps, NOx, meteorology, NMHC and OVOC.

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

No file uploaded. Click Select file... to browse local disk, then click Upload and check. Please remember to save your work regularly.

File header errors

File data errors (returning up to a maximum of 1000 rows)

- 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.

Augmenting vocabulary: 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.

