Training session on data formatting and data submission to GAW/EBAS data centre

Arranged within the frame of SAMLAC in November 2018

Ann Mari Fjæraa (amf@nilu.no)

With contribution from Markus Fiebig, Cathrine Lund Myhre and the EBAS-team

Outline

- About NILU, GAW, EBAS and ACTRIS
- Motivation for a central database and data reporting
- Data formatting and -submission



NILU - Norsk institutt for luftforskning

Norwegian Institute for Air Research

NILU is a private independent, nonprofit institution established in 1969.

- Research Council of Norway
- Norwegian and international industry
- Government agencies
- The EU's research programs, Ca 40% of the institute is international activity

More than 180 employees

- Main office in Oslo, Norway
- Tromsø
- Poland, The United Arab Emirates and in South Africa.

Chemical analysis, atmospheric observatories, modelling, consulting services

- Laboratories are among the most advanced in Europe
- Observatories in the Arctic, in Antarctica and in Norway
- Atmospheric data centre for ESA EVDC, EMEP, GAW-WDCA, EU-projects and others



GAW-WDCA and **GAW-WDCRG** data centre

Integrating information and measurements from all parts of the globe to obtain new information and facilitate easy access to the data

Main services of the data centre:

Collect and archive all project data in a long term sustainable system.

Provide free and open access to all data resulting from project.

Complement with data from other relevant networks.

Provide new products and tools for analysis of atmospheric composition.



http://www.gaw-wdca.org

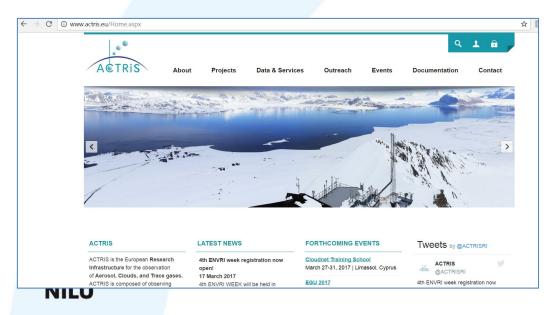


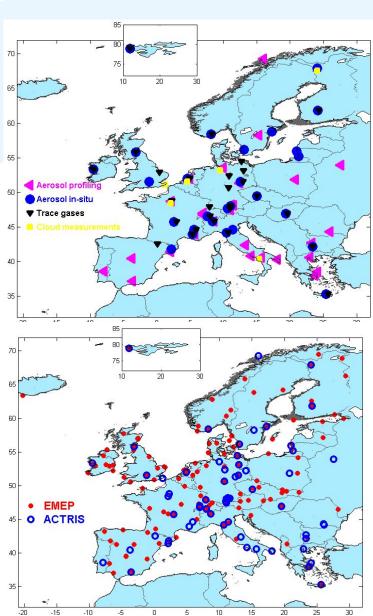
ACTRIS: Aerosol, Cloud and Trace Gases Research

Infrastructure Network

ACTRIS is an EU infrastructure project improving atmospheric observations, developing new methods and protocols, and harmonizing existing observations. http://www.actris.eu

We focus on ground based in situ data and EBAS during this course





EBAS Web portal

Placeholder for all GAW-WDCA, GAW-WDCRG, EMEP and ACTRIS groundbased in-situ data, among data from a wide range of other projects.

http://ebas.nilu.no





EBAS Database history

Originally the data archive of the European part of the UN Convention for Long-Range Transport of Air Pollution (CLRTAP), the European Monitoring and Evaluation Programme (EMEP) – first version in 1979.

Placeholder for all GAW-WDCA and GAW-WDCRG data.

New web interface in 2009, upgraded in 2017, linking also to **other tools**.

http://ebas.nilu.no

Currently ~110 000 datasets in the archive, and growing.





Data reporting, sharing and ownership







Background

Why is data reporting important?

Added value of reporting and sharing data.

Who owns the data in the data base?

Association to projects and the labelling of data:

- GAW-WDCA data
- GAW-WDCRG data
- ACTRIS data
- EMEP data

Data policies, Fair and regulated use of data.







Why is data reporting important?

Lot's of effort behind atmospheric observations, important that data are used!

Atmospheric measurements are hard work by many people and considerable amount of money is invested



Reporting makes the data available for various users, now and in the future, to contribute to solve the central environmental questions within air quality

and climate

Goal to provide access to data for many users for

- ACTRIS aerosol and trace gas activities
- ✓ EMEP
- ✓ GAW-Aerosol (GAW-WDCA)
- ✓ GAW-Reactive Gases (GAW-WDCRG)

Central data base with common standards for import and export facilitate easy access to data for a wide range of users

The value of reporting and sharing data

Make data available for various users now and in the future in sustainable data base.

Access to data for variety of users can hopefully also result in: Improved funding situation by demonstrating the use and value of data.

Improved data quality by improving methods and measurement practice as more data are used.

Facilitate collaboration and interactions between measurement communities.

Illustrate project progress: The data centre is a prominent PR instrument for projects.

EU commission and other bodes judges projects by number and rank of data users/data sets etc.

Who owns the data in the data base?

Every dataset created within ACTRIS, EMEP, GAW (and other programs) is owned by the partner/data providers who created this dataset.

Public data with easy access is the goal, but it has to be regulated...

The conditions of use of data is regulated in data policy documents for various programs/networks.



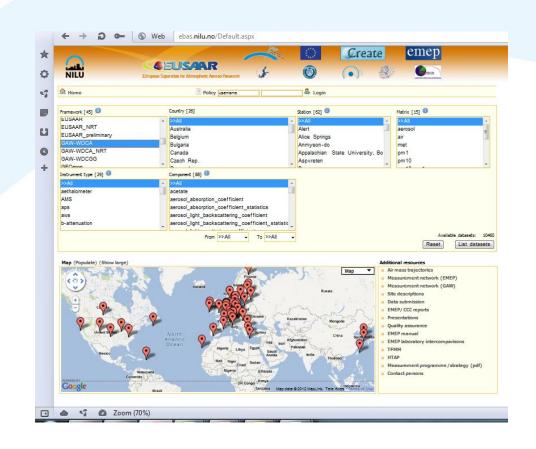




Data associated to WMO-GAW

All measurements performed within the GAW-program at GAW sites will have the GAW-WDCA and GAW-WDCRG label in EBAS.

Some of the sites overlapping with both ACTRIS and EMEP.





Users of EBAS

Data Providers:

- Policy Frameworks: CLRTAP, EMEP, HTAP, HELCOM, AMAP
 Data is owned by country or agency contributing to framework.
- Scientific Networks: GAW (WDCA, WDCRG)
 Data is owned by data provider or PI.
- Research Projects: EUSAAR, EUCAARI, GEOmon, ACTRIS
 Varying data ownership, but usually data provider / PI.

Data Users:

- The providing frameworks themselves.
- Modellers, EEA, ECMWF, Aerocom, ...



Data use has to be fair and regulated

Lot off effort behind, visibility to the data providers.

Facilitate the involvement of the data providers to ensure proper use of data when necessary.

Reduce misinterpretations (balance between data use, data analysis, depending on use etc).

Make the funding source visible, also important for future funding situation.



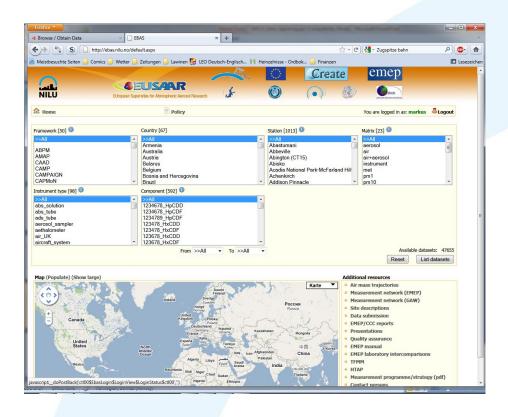
Data formatting and submission







The EBAS web interface 1 / 3



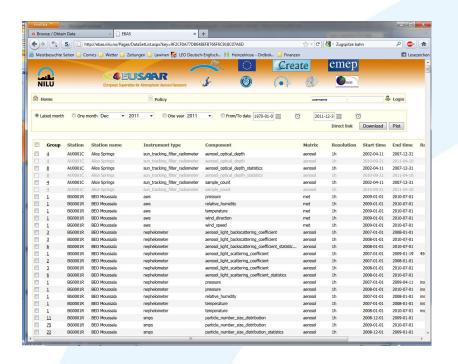
http://ebas.nilu.no

EBAS web-interface functions:

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



The EBAS web interface 2 / 3

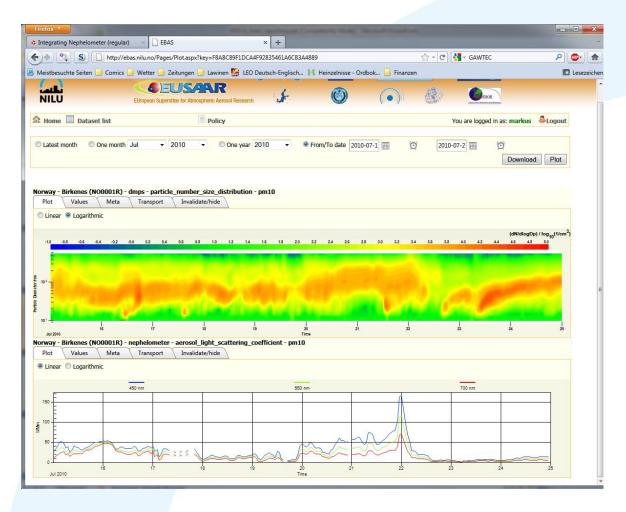


Search result page of EBAS web-interface:

- 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!).



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).



Modes of data submission

1. Regular, annual data submission

- Final, fully quality assured data, including uncertainty, and variability where applicable.
- Deadlines depend on framework reported to: EMEP / ACTRIS: 31 July following year.

2. Advanced data reporting

 Designed to establish traceability of data back to the time of measurement.

3. Near-Real-Time data reporting

- Data should be available to the user within max. 3 hours of measurement.
- Data is being processed and screened automatically, lower quality and higher uncertainty as with regularly reported data accepted.



Why do we ask providers to format the data?

1. Avoid errors

 Reformating 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 receives over 6000 datasets annually. Submitting formatted data frees resources for other services, e.g. dissemination.



Online formatting templates

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

Templates

Please browse the different categories for a list of templates, or select from the complete list below:

NOx

Cloud condensation nuclei

Particle Number Concentration

Particle Number Size Distribution

Particle light absorption coefficient

Particle light scattering coefficient

Aerosol Optrical Depth

Particulate Mass Conc., gravim

Particulate Mass Conc., online

Particulate chemical composition, online (ACSM)

VOC

Inorganic air/aerosol chemistry (filter-based)

Inorganic Precipitation Chemistry

EC OC

Heavy metals in aerosol particle-phase

Heavy metals in precipitation

Meteorology

Ozone

Coarse Mode Particle Size Distribution

NMH

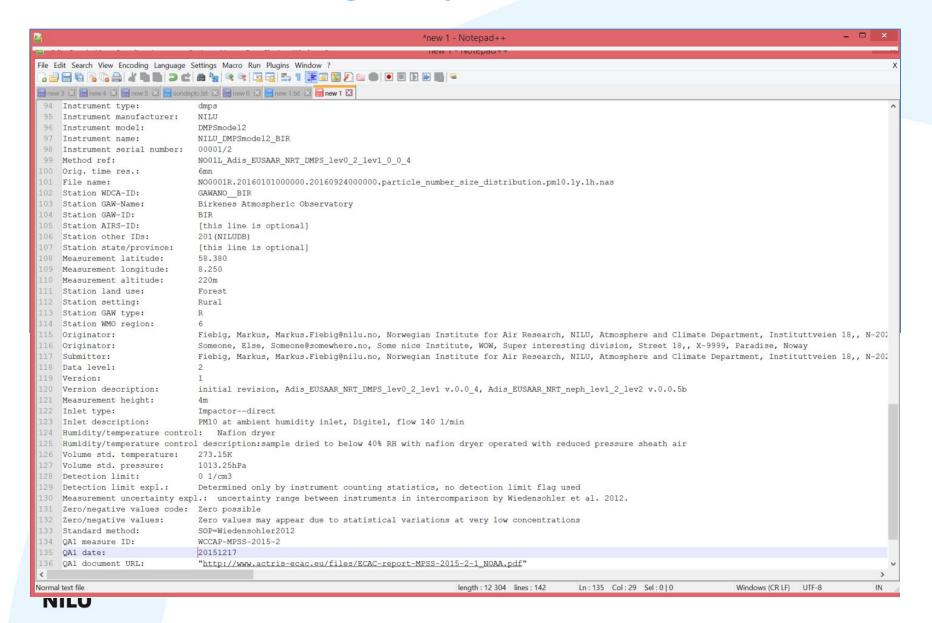
Mercury in air or aerosols

Mercury in precipitation

EBAS Master Template



Online formatting templates

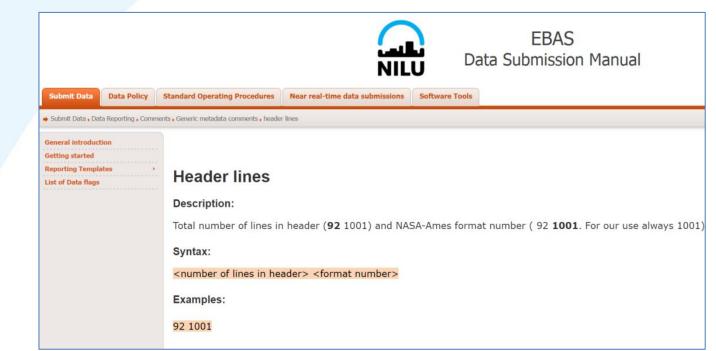


Online formatting templates

Metadata is "data about data"

Will ensure that your data will be understood and interpreted by any user.

Click on each line for **detailed explanations** of each metadata item.





Available templates

Reporting formats for

- Particle number size distribution (DMPS / SMPS)
- Aerosol scattering coefficient (integ. nephelometer)
- Aerosol absorption photometer (filter abs. photometer)
- Particle number concentration (CPC)
- Parameters with reporting formats updated with additional metadata:
- PM mass concentration (gravimetric)
- PM mass concentration (online methods)

- Cloud condensation nucleus number concentration (CCNC)
- Cloud condensation nucleus number size distribution (DMPS / CCNC)
- NOx (chemiluminescence photometer)
- NOy
- VOCs
- Online aerosol chemical speciation (ACSM / AMS)
- EC / OC
- Air ion spectrometer
- Meteorological parameters

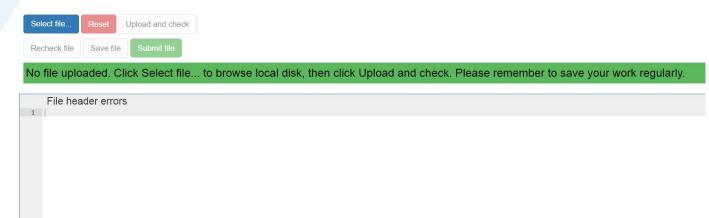


http://ebas-submit-tool.nilu.no/

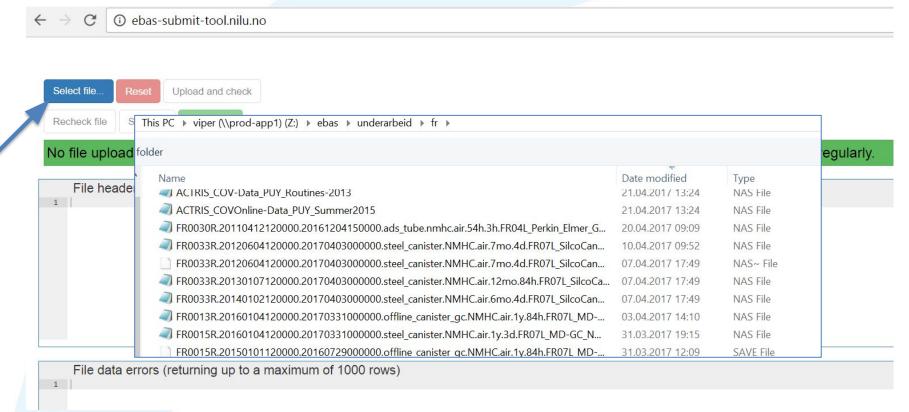
The EBAS Data Submission Tool is an **online file format checker** and a **data submission system**.

It is designed to give the data submitters direct feedback on the formatted NASA Ames files and to deliver files through online data submission.

It gives you information on how to best troubleshoot validation errors.









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

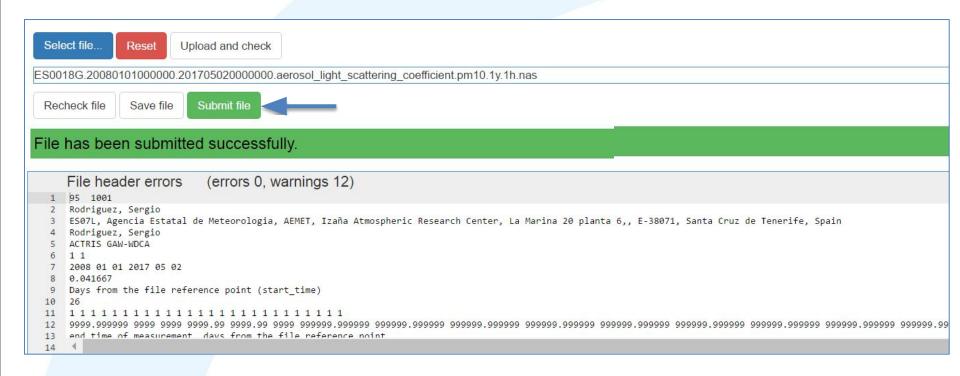


No file uploaded. Click Select file... to browse local disk, then click Upload and check. Please remember to save your work regularly.

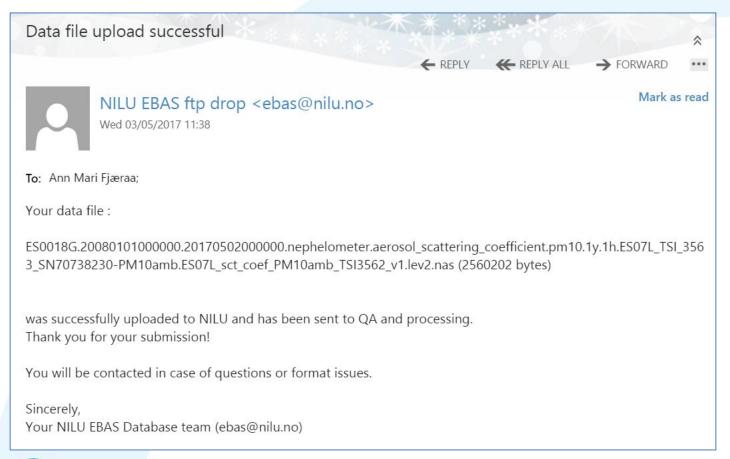






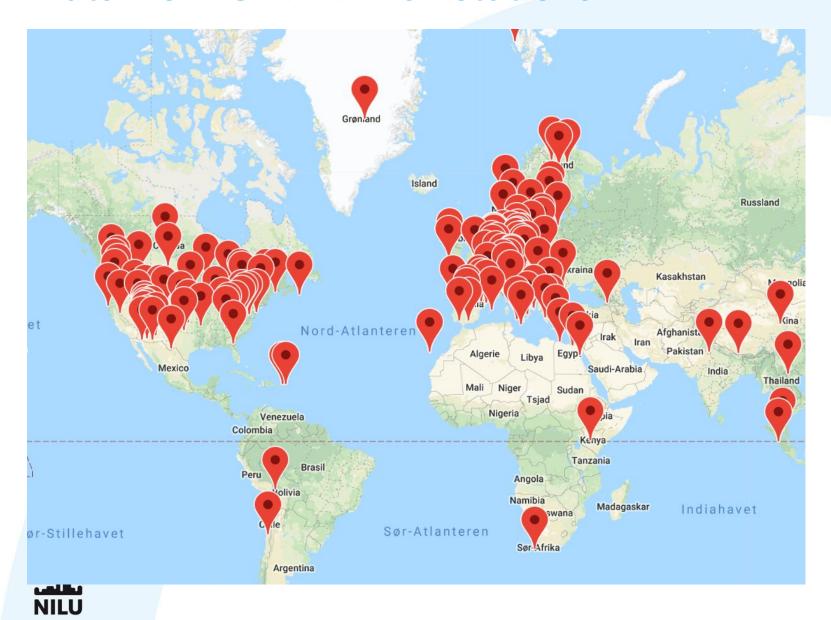








Data from GAW-WDCA stations



Practical examples

Open http://ebas.nilu.no/

Open https://ebas-submit.nilu.no/

Open https://ebas-submit-tool.nilu.no/

These are the links needed for data formatting, -checking and submission.

Make sure to have minimum a spreadsheet and a text file editor installed on your pc, e.g. MS Excel and Notepad ++

More advanced users might want to use programs such as python for the file formatting.

Practical examples

Collect your data in the spreadsheet

4	A	В	С	D	E	F	G	Н	1	J	K	L	M	N	0	Р	Q	R	S 4
1	Start date	End date	Start time	End time	P	RH T		sc1	sc2	sc3	bsc1	bsc2	bsc3	perc1_sc1	perc2_sc1	perc1_sc2	perc2_sc2	perc1_sc3	perc2_sc3
2	01.01.2017	01.01.2017	00:00	01:00	975,2	19,5	300,9	0,27187	-0,09045	0,61103	0,11978	-0,12408	0,3127	0,21548	-0,05436	0,49024	2,42093	1,70594	3,07579
3	91.01.2017	01.01.2017	01:00	02:00	975	19,1	300,9	0,12322	-0,30607	0,50115	0,16233	-0,02604	0,35287	0,1493	-0,09245	0,43942	2,48594	1,63732	3,32603
4	01.01.2017	01.01.2017	02:00	03:00	975,2	18,7	300,8	0,26316	-0,15821	0,59956	0,22416	0,06541	0,38332	0,03342	-0,19336	0,32035	2,41977	1,52134	3,21641
5	01.01.2017	01.01.2017	03:00	04:00	975,1	17,8	300,8	-0,03984	-0,43657	0,29166	0,12779	-0,03693	0,3035	0,16246	-0,05457	0,38271	1,53822	0,56632	2,48587
6	01.01.2017	01.01.2017	04:00	05:00	975,2	17,2	300,8	0,07285	-0,27135	0,49095	0,19372	0,01417	0,37442	0,03456	-0,29016	0,28149	1,30223	0,39315	2,24446
7	01.01.2017	01.01.2017	05:00	06:00	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	01.01.2017	01.01.2017	06:00	07:00	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	01.01.2017	01.01.2017	07:00	08:00	1075,3	95	1300,7	30,37	-26,584	33,494	10,811	-45,62	30,694	18,994	-11,363	50,267	104,284	-12,62	203,627
10	01.01.2017	01.01.2017	08:00	09:00	975,4	14,6	300,8	-0,02694	-0,54638	0,39194	-0,07726	-0,26191	0,07819	0,23279	-0,12984	0,55131	1,29487	0,49192	1,88645
11	01.01.2017	01.01.2017	09:00	10:00	976,4	13,6	300,8	0,50022	0,50022	0,50022	0,4067	0,4067	0,4067	0,25991	0,25991	0,25991	2,17815	2,17815	2,17815
12	01.01.2017	01.01.2017	10:00	11:00	976,4	13,5	300,8	0,21934	-0,15618	0,58582	0,18057	-0,03336	0,38659	-0,03386	-0,35864	0,2591	1,92234	1,10031	2,75977
13	01 01.2017	01.01.2017	11:00	12:00	97/6,4	13,6	300,9	0,19353	-0,18554	0,56217	0,13241	-0,05095	0,30615	0,17171	-0,15002	0,46601	1,71889	0,93046	2,40406
14	01.01.2017	01.01.2017	12:00	13:00	976,3	13,9	301	0,17879	-0,14647	0,50496	0,16256	-0,03532	0,34888	0,18183	-0,03816	0,4122	1,80356	0,86475	2,54504
1 =	01 01 2017	01 01 2017	12.00	1/1.00	076 5	1/	201 1	0 16761	0 35646	U ESSUE	0 0///70	N 17N70	A 21EE2	ก 1กวกว	0 17/6	רררנח	1 /0//00	000075	2 25051

You will need to modify the **start_time** and **end_time** parameters.



Practical examples

The level 2 data must be in 1h averages, continues measurements

The year start at time 0, and after one full day of 24hours time is 1/24 = 0.041667

4	Α	В	С	D	E	F	G	Н	1	J
4	01.01.2017	01.01.2017	02:00	03:00	0,083333	0,125000	\	975,2	18,7	30
5	01.01.2017	01.01.2017	03:00	04:00	0,125000	0,166667		975,1	17,8	30
6	01.01.2017	01.01.2017	04:00	05:00	0,166667	0,208333		975,2	17,2	30
7	01.01.2017	01.01.2017	05:00	06;00	0,208333	0,250000		0	0	
8	01.01.2017	01.01.2017	06:00	07:00	0,250000	0,291667		0	0	
9	01.01.2017	01.01.2017	07:00	08:00	0,291667	0,333333		1075,3	95	130
10	01.01.2017	01.01.2017	08:00	09:00	0,333333	0,375000		975,4	14,6	30
11										
12										
13	31.12.2017	31.12.2017	22:00	23:00	364,943667	364,958333		980,2	13,1	30
14	31.12.2017	01.01.2018	23:00	00:00	364,958333	365,000000		981,2	12,8	30
15						/				

Missing data, suspicious data

How to deal with missing data

4	Α	В	С	D	E	F	G	Н	1	J	K	L	M	N	0	P	Q	R	S 4
1	Start date	End date	Start time	End time	P	RH T		sc1	sc2	sc3	bsc1	bsc2	bsc3	perc1_sc1	perc2_sc1	perc1_sc2	perc2_sc2	perc1_sc3	perc2_sc3
2	01.01.2017	01.01.2017	00:00	01:00	975,2	19,5	300,9	0,27187	-0,09045	0,61103	0,11978	-0,12408	0,3127	0,21548	-0,05436	0,49024	2,42093	1,70594	3,07579
3	01.01.2017	01.01.2017	01:00	02:00	975	19,1	300,9	0,12322	-0,30607	0,50115	0,16233	-0,02604	0,35287	0,1493	-0,09245	0,43942	2,48594	1,63732	3,32603
4	01.01.2017	01.01.2017	02:00	03:00	975,2	18,7	300,8	0,26316	-0,15821	0,59956	0,22416	0,06541	0,38332	0,03342	-0,19336	0,32035	2,41977	1,52134	3,21641
5	01.01.2017	01.01.2017	03:00	04:00	975,1	17,8	300,8	-0,03984	-0,43657	0,29166	0,12779	-0,03693	0,3035	0,16246	0,05457	0,38271	1,53822	0,56632	2,48587
6	01.01.2017	01.01.2017	04:00	05:00	975,2	17,2	300,8	0,07285	-0,27135	0,49095	0,19372	0,01417	0,37442	0,03456	-0,29016	0,28149	1,30223	0,39315	2,24446
7	01.01.2017	01.01.2017	05:00	06:00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	01.01.2017	01.01.2017	06:00	07:00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	01.01.2017	01.01.2017	07:00	08:00	1075,3	95	1300,7	30,37	-26,584	33,494	10,811	-45,62	30,694	18,994	-11,363	50,267	104,284	-12,62	203,627
10	01.01.2017	01.01.2017	08:00	09:00	975,4	14,6	300,8	-0,02694	-0,54638	0,39194	-0,07726	-0,26191	0,07819	0,23279	-0,12984	0,55131	1,29487	0,49192	1,88645
11	01.01.2017	01.01.2017	09:00	10:00	976,4	13,6	300,8	0,50022	0,50022	0,50022	0,4067	0,4067	0,4067	0,25991	0,25991	0,25991	2,17815	2,17815	2,17815
12	01.01.2017	01.01.2017	10:00	11:00	976,4	13,5	300,8	0,21934	-0,15618	0,58582	0,18057	-0,03336	0,38659	-0,03386	-0,35864	0,2591	1,92234	1,10031	2,75977
13	01.01.2017	01.01.2017	11:00	12:00	976,4	13,6	300,9	0,19353	-0,18554	0,56217	0,13241	-0,05095	0,30615	0,17171	-0,15002	0,46601	1,71889	0,93046	2,40406
14	01.01.2017	01.01.2017	12:00	13:00	976,3	13,9	301	0,17879	-0,14647	0,50496	0,16256	-0,03532	0,34888	0,18183	-0,03816	0,4122	1,80356	0,86475	2,54504
1 =	01 01 2017	01 01 2017	12.00	1/1.00	076 E	1.1	201 1	0 16761	0 35646	U ESSUE	0 0///70	0 1 7 N 7 O	N 21EE2	ก 1กวกว	0 17/6	רכרככ ח	1 /0/00	0 00575	2 25051

Missing values are always combinations of 999s

Missing values must be to a power of 10 higher than [max value] for each parameter.

If [max pressure] is 1030.00 hPa \rightarrow missing value for pressure is 99999.99

Missing data, suspicious data

How to deal with missing data

E	F	G	Н		J	K	L	IVI	IN	U
0,083333	0,125000		975,2	18,7	300,8	0,26316	-0,15821	0,59956	0,22416	0,06541
0,125000	0,166667		975,1	17,8	300,8	-0,03984	-0,43657	0,29166	0,12779	-0,03693
0,166667	0,208333		975,2	17,2	300,8	0,07285	-0,27135	0,49095	0,19372	0,01417
0,208333	0,250000		99999,9	9999,9	9999,9	99999,99999	99999,99999	99999,99999	99999,99999	99999,99999
0,250000	0,291667		99999,9	9999,9	9999,9	99999,99999	99999,99999	99999,99999	99999,99999	99999,99999
0,291667	0,333333		1075,3	95	1300,7	30,37	-26,584	33,494	10,811	-45,62
0,333333	0,375000		975,4	14,6	300,8	-0,02694	-0,54638	0,39194	-0,07726	-0,26191

										•
E	F	G	H		J	K	L	M	N	0
0,083333	0,125000		975,2	18,7	300,8	0,26316	-0,15821	0,59956	0,22416	0,06541
0,125000	0,166667		975,1	17,8	300,8	-0,03984	-0,43657	0,29166	0,12779	-0,03693
0,166667	0,208333		975,2	17,2	300,8	0,07285	-0,27135	0,49095	0,19372	0,01417
0,208333	0,250000		99999,9	9999,9	9999,9	99999,99999	99999,99999	99999,99999	99999,99999	99999,99999
0,250000	0,291667		99999,9	9999,9	9999,9	99999,99999	99999,99999	99999,999999	99999,99999	99999,99999
0,291667	0,333333		1075,3	95	1300,7	30,37000	-26,58400	33,49400	10,81100	-45,62000
0,333333	0,375000		975,4	14,6	300,8	-0,02694	-0,54638	0,39194	-0,07726	-0,26191



Flagging of data

How to flag your data

Full list of data flags at

https://ebas-submit.nilu.no/Submit-Data/List-of-Data-flags

List of flags used in ebas

All flags are grouped in four categories: V (valid measurement), I (invalid measurement), M (missing measurement) or H (hidden and invalid measurements).

Flag	V/I/H	Description
Group 9	: Missing fl	ags
999	M	Missing measurement, unspecified reason
990	M	Precipitation not measured due to snow-fall. Needed for historic data, should not be needed for new data
980	M	Missing due to calibration or zero/span check
900	H	Hidden and invalidated by data originator
Group 8	: Flags for	undefined data elements
899	M	Measurement undefined, unspecified reason
890	M	Concentration in precipitation undefined, no precipitation
Group 7	: Flags used	l when the value is unknown
799	I	Measurement missing (unspecified reason), data element contains estimated value
798	V	Measurement missing (unspecified reason), data element contains estimated value. Considered valid.
797	V	Data element taken from co-located instrument
784	I	Low precipitation, concentration estimated
783	M	Low precipitation, concentration unknown
782	V	Low precipitation, concentration estimated
781	V	Value below detection limit, data element contains detection limit
780	V	Value below detection or quantification limit, data element contains estimated or measured value. Use of flag 147 is encouraged.
771	V	Value above range, data element contains upper range limit
770	V	Value above range, data element contains estimated value
760	V	Value estimated by summing up the constituents measured
750	M	H ⁻ not measured in alkaline sample
741	V	Non refractory AMS concentrations. Don't include compounds that volatalises above 600 deg C
740	V	Probably biased gas/particle ratio
701	I	Less accurate than usual, unspecified reason. (Used only with old data, for new data see groups 6 and 5)
Group 6	: Mechanic	al or instrumental problem
699	I	Mechanical problem, unspecified reason

Flagging of data

How to flag your data

Full list of data flags at

https://ebas-submit.nilu.no/Submit-Data/List-of-Data-flags

Say something about the quality of the data

	perc2_bsc1	perc1_bsc2	perc2_bsc2	perc1_bsc3	perc2_bsc3	flag
ţ	1,31303	2,18575	1,63356	1,12328	2,03477	0,000
i	1,79935	2,44452	1,43118	1,09692	1,84614	0,000
	1,91803	2,60812	1,64051	1,16969	2,08107	0,000
	1,10412	1,94116	1,0576	0,63175	1,43576	0,000
,	0,76698	1,73577	0,95629	0,52078	1,36632	0,000
ì	99999,99999	99999,99999	99999,99999	99999,99999	99999,99999	0,999
)	99999,99999	99999,99999	99999,99999	99999,99999	99999,99999	0,999
,	800,28	1653,41	66,296	311,09	1025,36	0,599
,	0,75819	1,48232	1,28995	0,85156	1,71366	0,000



Flagging of data

How to flag your data

Full list of data flags at

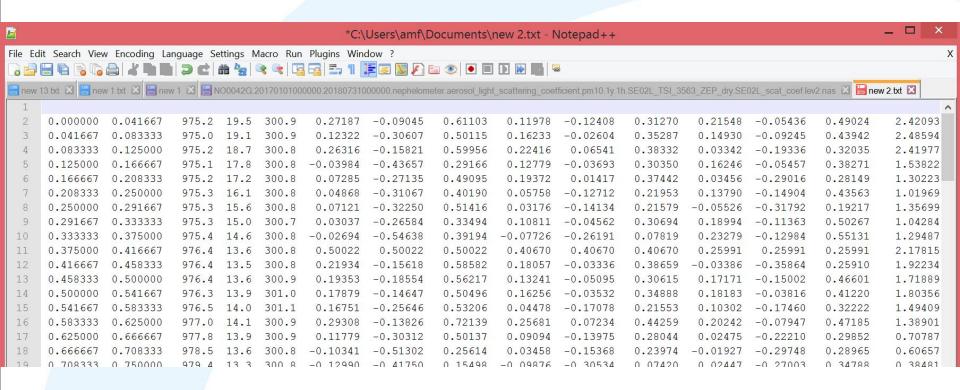
https://ebas-submit.nilu.no/Submit-Data/List-of-Data-flags

E	F	G	Н		J	K	L	M	N	0
0,083333	0,125000		975,2	18,7	300,8	0,26316	-0,15821	0,59956	0,22416	0,06541
0,125000	0,166667		975,1	17,8	300,8	-0,03984	-0,43657	0,29166	0,12779	-0,03693
0,166667	0,208333		975,2	17,2	300,8	0,07285	-0,27135	0,49095	0,19372	0,01417
0,208333	0,250000		99999,9	9999,9	9999,9	99999,99999	99999,99999	99999,99999	99999,99999	99999,99999
0,250000	0,291667		99999,9	9999,9	9999,9	99999,99999	99999,99999	99999,99999	99999,99999	99999,99999
0,291667	0,333333		1075,3	95	1300,7	30,37000	-26,58400	33,49400	10,81100	-45,62000
0,333333	0,375000		975,4	14,6	300,8	-0,02694	-0,54638	0,39194	-0,07726	-0,26191



Save your file and add header

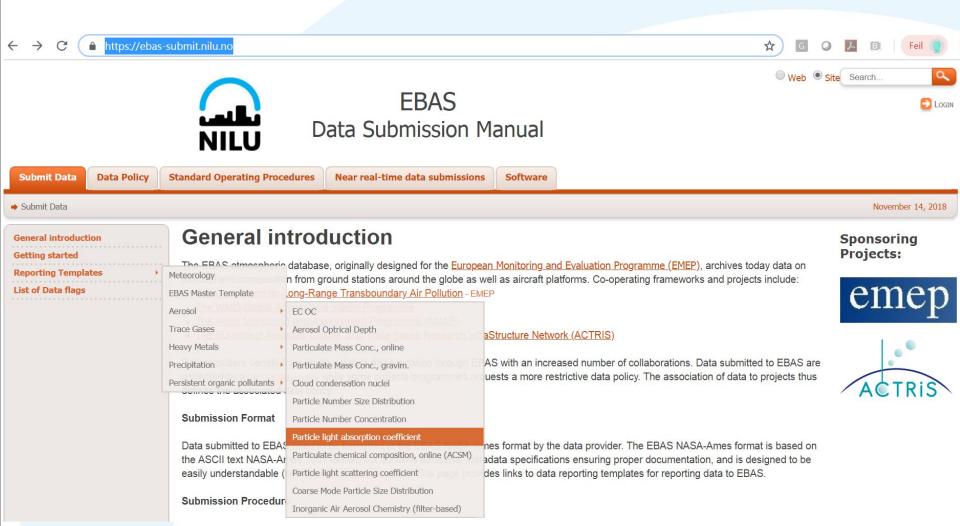
Save spreadsheet file as .txt and open in text editor





Save your file and add header

Find your template at https://ebas-submit.nilu.no/



Save your file and add header

Find your template at https://ebas-submit.nilu.no/
Add header, save and submit.

Let's try in practice!



EBAS in Social Media

Facebook: EBAS data centre

Twitter: @EBAS_NILU





Follow us for info about training courses, updates, events.



Thank you!

