



Creating standardized and FAIR data in AnaEE Denmark

Klaus Steenberg Larsen

Associate Professor, AnaEE Denmark Coordinator
Department of Geosciences and Natural Resource Management
University of Copenhagen

Katrine Korsgaard Vendelboe

Centre Manager, AnaEE Technology Centre
Department of Plant and Environmental Sciences
University of Copenhagen



AnaEE Denmark and potential users of data

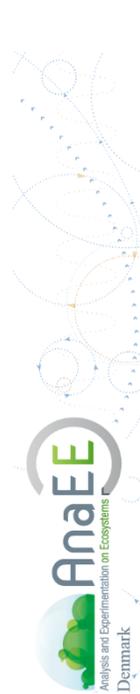
AnaEE (Analysis and Experimentation on Ecosystems)

- A pan-European research infrastructure for experimental field-scale research facilities (<https://www.anaee.eu/>)
- Partners in AnaEE Denmark are UCPH, AU, DTU and RUC

Potential users of data

- Research institutions
- Companies / industry
- Governmental institutions

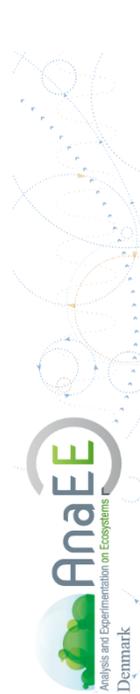
AnaEE Denmark will share all data together with (the international) AnaEE ERIC community through a central data and modelling center (with decentralised data storage).



AnaEE Denmark FAIR data strategy

- Decentralised data storage – no common database (same at international level)
- Go for DOIs – resources and repositories are now becoming available at the Universities, e.g. ERDA at UCPH, LOAR (Royal Library, AU), and DTU Data at DTU
- Develop standardized metadata descriptions across platforms
- Data in several categories:
 - "Easy": Standard meteorological data from experiments
 - "Difficult": Measured response data (highly variable)

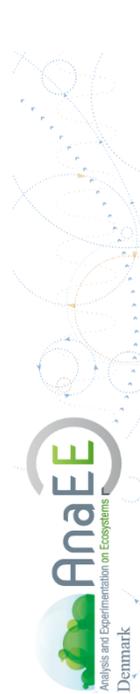
Start here!



AnaEE Denmark has had two FAIR data projects:

- co-funded by DeiC (Danish e-infrastructure Cooperation)

- Project 1: 10 MM in total in 2019
- Project 2: 5 x 2-3 days workshops in June-October 2020 with DeiC + GOFAIR. No MM, but free workshops



Implementing FAIR – project 1 proces

July-August 2019

Project plan

Project group

Selection of people from all 3 universities (KU, AU, DTU) within AnaEE DK

1st meeting

Introductions by project participants and FAIR status at each university
Introduction to FAIR by DeIC

September 2019

Identification of variables and metadata standards

Each participant shared standards and variables from their common practice, the group searched for already established standards in the field and several discussions led to a common ontology, which leans towards one already established and well known standard (ICOS).

October 2019

Creating metadata

Participants started creating their own metadata based on the agreed standards.

Finding repositories

The group started searching for appropriate repositories – within and outside their universities.

November 2019

Creating metadata

Participants started creating their own metadata based on the agreed standards.

Uploading datasets

Participants started uploading datasets to a common repository (at KU), registering metadata and creating DOIs.

December 2019 ->

3 FAIRfied datasets

3 complete datasets were uploaded to KUs ERDA with DOIs

Learning and reporting

Participants have continued following the practice in the pilotproject – the second project focused on making machine-readable metadata, but university infrastructure is not yet capable of supporting this.

Challenges

Difficulties understanding the overall concepts (metadata, ontology, variables) and level of detail needed

Challenges

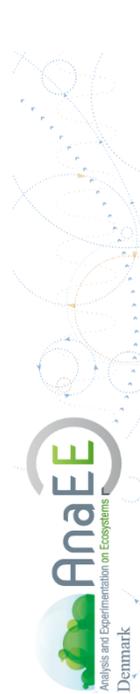
Different levels of repository options at different institutions – AU did not have a repository, KU had one but with very low level of search- and findability options.



The steps we went through (in short):

- Create standardized templates for data and meta data
- Find a persistent data repository, i.e. ERDA (UCPH) and LOAR (Royal Library for AU)
- Get a PID (in our case DOI – ERDA and Royal Library can provide)
- Create machine-readable metadata (M4M)

Link to data: <https://anaee.dk/access/>



AnaEE DK FAIR datasets

Brandbjerg/CLIMAITE

Data from CLIMAITE (Brandbjerg), Copenhagen

University: doi.org/10.17894/ucph.e58a99c2-da7b-444a-b1c0-11f00e70041c

For more information about the site: [CLIMAITE/Brandbjerg \(Copenhagen University\)](https://CLIMAITE/Brandbjerg)

Højbakkegård

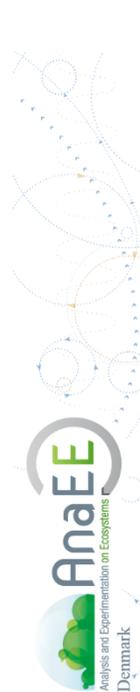
Data from Højbakkegård, Copenhagen University: doi.org/10.17894/ucph.8d941a14-b098-4ca5-b177-412f50be1731

For more information about the site: [Højbakkegård \(Copenhagen University\)](https://Højbakkegård)

Foulumgård

Data from Foulumgård, Aarhus University: dx.doi.org/10.21994/loar4109

For more information about the site: [Foulum \(Aarhus University\)](https://Foulum)



Data files are "flat" csv files – time series of data

FOULUM_20180525_20191231_HOURLY.csv - Excel

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do... Klaus Steenberg Larsen Share

Clipboard Font Alignment Number Conditional Formatting Styles Cells Editing

A1 # TITLE: Meteorological data -FOULUM_20180525_20191231

1 # TITLE: Meteorological data -FOULUM_20180525_20191231
2 # FILE NAME: FOULUM_20180525_20191231_HOURLY.csv
3 # DATA FORMAT: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
4 # TOTAL LINES: 14088
5 # HEADER LINES: 35
6 # STATION CODE: FOULUM
7 # STATION NAME: FOULUM
8 # OBSERVATION CATEGORY: Meteorology
9 # COUNTRY/TERRITORY: Denmark
10 # CONTRIBUTOR: Aarhus University
11 # LATITUDE: 56.29 N
12 # LONGITUDE: 9.34 E
13 # ALTITUDE: 54 m
14 # NUMBER OF SAMPLING HEIGHTS: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
15 # CONTACT POINT: JWM Pullens jwmp@agro.au.dk
16 # COVERING PERIOD: 2018-05-25T00:00 2019-12-31T23:00
17 # TIME INTERVAL: hourly
18 # LABEL MISSING DATA: NA
19 # MEASUREMENT PARAMETER: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
20 # MEASUREMENT UNIT: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
21 # MEASUREMENT METHOD: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
22 # SAMPLING TYPE: continuous
23 # TIME ZONE: UTC +1
24 # MEASUREMENT SCALE: NA
25 # DATA POLICY: Data is licensed under a Creative Commons Attribution 4.0 international licence (<https://creativecommons.org/licenses/by/4.0/>)
26 # DATA QUALITY ASSURANCE: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
27 # COMMENTS:
28 # - Times are UTC +1
29 # - Time-averaged values are reported at the end of the averaging interval.
30 # - flag '1' = original data. Quality checked to be OK
31 # - flag '2' = original data failed quality check and data were gapfilled by linear interpolation between data points with flag = 1 from the same sensor
32 # - flag '3' = original data failed quality check and data were gapfilled with Mean Daily Variance for the sensor
33 # - flag '4' = original data failed quality check and data were gapfilled with data from other site (e.g. nearest weather station)
34 # END OF HEADER

35	TIMESTAN	Year	Month	Day	Hour	TA	TA_flag	RH	RH_flag	WS	WS_flag	WD	WD_flag	P	P_flag	G	G_flag	PA	PA_flag	NETRAD	SW_IN	SW_IN_fli	TS_1_1	TS_1_1_fli	TS_2_1	TS_2_1_fli	TS_2_2	TS_2_2_flag
36	2.02E+11	2018	5	25	0	13.5	1	74.2	1	2.8	1	92.2	1	0	1	-2.6	1	1019.1	1	-113	0	1	11.8	1	16.2	1	15.3	1
37	2.02E+11	2018	5	25	1	13	1	74.6	1	2.6	1	83.9	1	0	1	-4.1	1	1019.1	1	-110.8	0	1	11.7	1	16	1	15.2	1
38	2.02E+11	2018	5	25	2	12.2	1	78.1	1	2.6	1	79	1	0	1	-5.3	1	1018.9	1	-108.8	0	1	11.1	1	15.7	1	15.2	1
39	2.02E+11	2018	5	25	3	11.5	1	78	1	1.5	1	116.1	1	0	1	-6.3	1	1018.8	1	-108.5	0	1	9.7	1	15.5	1	15.1	1
40	2.02E+11	2018	5	25	4	11.2	1	78.3	1	1.8	1	111.8	1	0	1	-7.6	1	1018.6	1	-100	9.6	1	9.3	1	15.3	1	15	1



Metadata (adapted from ICOS)

Additional metadata file follows each data file

- metadata on variables (adapted from ICOS)



FOULUM_REF_METADATA_20180525_20191231_HOURLY (3).csv - Excel

Klaus Steenberg Larsen

Variable_Code	Units	Description	Sensor_Height	Sensor_Type	Sensor_Measurement_Interval	Data_Analysis	Data_Quality_Assurance	Grib_code
TIMESTAMP	YYYYMMDD	Date	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	NA
Year	Year	Time (yearly)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2004
Month	Month	Time (monthly)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2003
Day	Day	Time (daily)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2002
Hour	Hour	Time (hourly)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2001
TA	°C	Air temperature	2 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.0.0.0
RH	%	Relative Humidity	2 m	Rotronic	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.0.1.1
WD	°	Wind direction	2 m	Vector instruments A101 M/L	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.0.2.0
WS	m s-1	Wind speed	2 m	Vector Instrumts W200P	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.0.2.1
P	mm	Precipitation	1.5 m	OTT Pluvio2	0.1 mm	1 minute values, 10 minute sum	Data were flagged for gapfilling by applying with vis	4.2.0.1.8
G	W m-2	Soil heat flux	5 cm depth	Campbell scientific: HFT3-L	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.2.3.26
PA	hPa	Air pressure	2 m	Vaisala PTB101B	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.0.3.0
NETRAD	W m-2	Net shortwave radiation	1.5 m	Kipp & Zonen CNR1	1 minute	1 minute values, 10 minute mean	Data were calculated and not flagged	4.2.0.4.0
SW_IN	W m-2	Global radiation	1.5 m	Kipp & Zonen CNR1	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.0.4.7
TS_1_1	°C	Soil temperature under grass	- 0.05 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.2.0.2
TS_2_1	°C	Soil temperature at 10 cm depth below bare soil	- 0.10 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.2.0.2
TS_2_2	°C	Soil temperature at 30 cm depth below bare soil	- 0.30 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying with vis	4.2.2.0.2
For WMO code tables. Grib code table FM_92 (Ver.23.0.0) see https://www.wmo.int/pages/prog/www/WMOcodes/WMO306_v12/LatestVERSION/LatestVERSION.html								

FOULUM_REF_METADATA_20180525_20

Metadata in data file and metadata file are not M4M! (M4M is next step...)

FOULUM_20180525_20191231_HOURLY.csv - Excel

TITLE: Meteorological data -FOULUM_20180525_20191231
 # FILE NAME: FOULUM_20180525_20191231_HOURLY.csv
 # DATA FORMAT: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
 # TOTAL LINES: 14088
 # HEADER LINES: 35
 # STATION CODE: FOULUM
 # STATION NAME: FOULUM
 # OBSERVATION CATEGORY: Meteorology
 # COUNTRY/TERRITORY: Denmark
 # CONTRIBUTOR: Aarhus University
 # LATITUDE: 56.29 N
 # LONGITUDE: 9.34 E
 # ALTITUDE: 54 m
 # NUMBER OF SAMPLING HEIGHTS: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
 # CONTACT POINT: JWM Pullens jwmp@agro.au.dk
 # COVERING PERIOD: 2018-05-25T00:00 2019-12-31T23:00
 # TIME INTERVAL: hourly
 # LABEL MISSING DATA: NA
 # MEASUREMENT PARAMETER: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
 # MEASUREMENT UNIT: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
 # MEASUREMENT METHOD: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
 # SAMPLING TYPE: continuous
 # TIME ZONE: UTC +1
 # MEASUREMENT SCALE: NA
 # DATA POLICY: Data is licensed under a Creative Commons Attribution 4.0 International License
 # DATA QUALITY ASSURANCE: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
 # COMMENTS:
 # - Times are UTC +1
 # - Time-averaged values are reported at the end of the averaging interval.
 # - flag '1' = original data. Quality checked to be OK
 # - flag '2' = original data failed quality check and data were gapfilled by linear interpolation
 # - flag '3' = original data failed quality check and data were gapfilled with meteorological model
 # - flag '4' = original data failed quality check and data were gapfilled with meteorological model

metadata

FOULUM_REF_METADATA_20180525_20191231_HOURLY (3).csv - Excel

Variable_Code

Variable_Code	Units	Description	Sensor_Height	Sensor_Type	Sensor_Measurement_Interval	Data_Analysis	Data_Quality_Assurance	Grib_code
1	Variable_Code	Description	Sensor_Height	Sensor_Type	Sensor_Measurement_Interval	Data_Analysis	Data_Quality_Assurance	Grib_code
2	TIMESTAMP	YYYYMMDD Date	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	NA
3	Year	Year Time (yearly)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2004
4	Month	Month Time (monthly)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2003
5	Day	Day Time (daily)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2002
6	Hour	Hour Time (hourly)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2001
7	TA	°C Air temperature	2 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.0.0.0	
8	RH	% Relative Humidity	2 m	Rotronic	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.0.1.1	
9	WD	° Wind direction	2 m	Vector Instruments A101 M/L	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.0.2.0	
10	WS	m s-1 Wind speed	2 m	Vector Instruments W200P	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.0.2.1	
11	P	mm Precipitation	1.5 m	OTT Pluvio2	0.1 mm	1 minute values, 10 minute sum	Data were flagged for gapfilling by applying vith vis 4.2.0.1.8	
12	G	W m-2 Soil heat flux	5 cm depth	Campbell scientific: HFT3-L	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.2.3.26	
13	PA	hPa Air pressure	2 m	Vaisala PTB101B	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.0.3.0	
14	NETRAD	W m-2 Net shortwave radiation	1.5 m	Kipp & Zonen CNR1	1 minute	1 minute values, 10 minute mean	Data were calculated and not flagged	4.2.0.4.0
15	SW_DOWN	W m-2 Global radiation	1.5 m	Kipp & Zonen CNR1	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.0.4.7	
16	TS_1_1	°C Soil temperature under grass	- 0.05 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.2.0.2	
17	TS_2_1	°C Soil temperature at 10 cm depth below bare soil	- 0.10 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.2.0.2	
18	TS_2_2	°C Soil temperature at 30 cm depth below bare soil	- 0.30 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis 4.2.2.0.2	

All metadata





Thank you!