
Towards a semantic approach for environmental timeseries data reusability

FAIR Data Science for Green Life Sciences
Session B: Intelligent data infrastructures in agrifood

Argyris Samourkasidis, Ioannis N. Athanasiadis
INF Group



Motivation

- Semantic heterogeneity
- Environmental timeseries
- Re-usability
- E-scientists



Assumptions - What is a dataset?

- A collection of observations
- with arbitrary:
 - format/syntax
 - vocabulary



A specific problem: DSSAT to APSIM

```
dssat.txt — inputs
! SRAD    daily Insolation Incident On A Horizontal Surface (MJ/m^2/day)
! T2M     Average Air Temperature At 2 m Above The Surface Of The Earth (degrees C)
! TMIN    Minimum Air Temperature At 2 m Above The Surface Of The Earth (degrees C)
! TMAX    Maximum Air Temperature At 2 m Above The Surface Of The Earth (degrees C)
! RH2M    Relative Humidity At 2 m (%)
! TDEW    Dew/Frost Point Temperature At 2 m (degrees C)
! RAIN    Average Precipitation (mm/day)
! WIND    Wind Speed At 10 m Above The Surface Of The Earth (m/s)
*WEATHER DATA: NASA

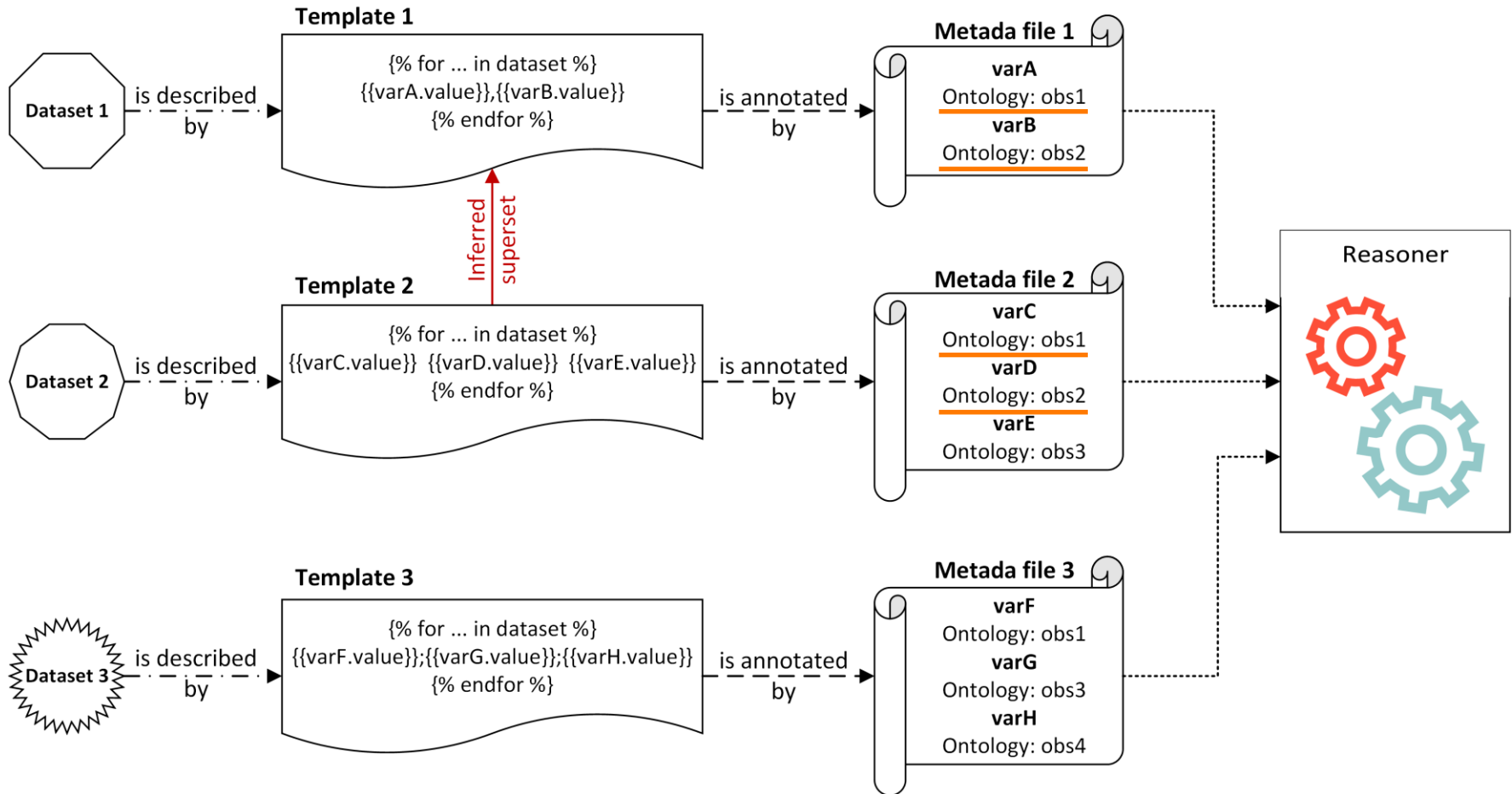
@ INSI  WTHLAT  WTHLONG  WELEV  TAV  AMP  REFHT  WNDHT
   NASA  33.500   -80.750    39          10

@ WEYR  WEDAY  SRAD   TMAX   TMIN  RAIN  WIND  TDEW   T2M  RH2M
2000   1    8.0  16.8   6.5  0.4  0.8  10.1  11.2  92.9
2000   2   10.7  19.6   9.2  0.0  1.6  12.7  13.7  93.8
2000   3   12.2  21.7  12.3  0.0  3.0  14.6  16.4  89.6
2000   4    4.6  21.8  13.9  0.0  5.1  15.9  17.9  88.4
2000   5   13.6  10.1   4.0  0.0  4.5  1.0   7.2  64.9
2000   6    9.8   9.8   1.3  0.0  2.5  3.5   6.2  82.8
2000   7   12.0  15.4   6.7  0.0  3.0  7.7  10.8  80.7
2000   8    9.3  14.8   1.7  0.0  2.8  2.4   8.4  65.6
2000   9    4.8  18.6   6.0  0.0  1.9  11.3  12.9  90.0
2000  10    9.8  20.9  10.9  0.8  4.8  13.3  17.2  77.6
2000  11   13.9  18.1   6.4  0.0  4.1  4.7  11.2  63.7
2000  12   13.7  19.4   3.0  0.0  2.4  1.3  10.5  53.2
2000  13    9.0  21.1   9.5  0.0  5.3  6.8  14.3  60.5
2000  14   14.3   7.3  -0.3  0.0  5.4  -9.9   3.8  36.4
2000  15  12.9   8.0  -4.2  0.0  1.6 -11.2  1.1  39.6
2000  16    8.4  15.9   0.2  0.0  4.2  -0.9   8.2  52.6
2000  17   13.3  10.6   3.0  0.0  3.7  -1.1   7.5  54.1
2000  18   2.7   9.8   4.2  6.1  2.8  2.5   6.0  78.2
2000  19    9.7   6.7   1.6  1.0  2.6  -0.0   3.8  76.1
2000  20  12.5  11.7   2.4  0.4  6.1  3.9   7.9  75.5
2000  21  15.6   4.4  -3.5  0.0  3.7  -9.2  -0.5  52.2
2000  22    6.6   3.0  -5.1  0.9  2.3  -8.2  -0.8  57.1
2000  23    1.9   6.2   1.9  53.3  2.3  3.2   3.2  100.0
2000  24   2.2   2.0  -0.2  39.0  5.3  -0.4  0.8  92.2
```



```
Yucheng.met — inputs
[weather.met.weather]
Latitude   = 36.68
longitude  = 116.98
tav       = 14.6 (oC) ! annual average ambient temperature
amp       = 28.2 (oC) ! annual amplitude mean monthly temperature
!!!! 1/01/1961 to 31/12/2005
day year radn maxt mint rain wind RH
273 2002 17.5 27.2 14.6 0 3.5 54
274 2002 13.6 23.1 14.7 0 5.3 40
275 2002 15.8 27.1 11.1 0 5.5 29
276 2002 15.5 25.8 16.5 0 3.8 39
277 2002 14.9 25.5 14.6 0 2.5 63
278 2002 15.2 23.1 15.2 0 3 47
279 2002 13.4 19.9 10.9 0 3 38
280 2002 15.7 19.3 8.2 0 2.5 47
281 2002 15.3 22.9 8.6 0 3.8 41
282 2002 15.4 26 14.3 0 5.8 30
283 2002 10.9 24.1 16.1 0 4.3 47
284 2002 14.1 26.2 17.6 0 3.3 54
285 2002 9.5 24.1 19.4 0 4.8 50
286 2002 14.7 24.2 13.5 0 3.8 22
287 2002 11.7 25.8 19 0 6.3 43
288 2002 15.7 27.3 12.8 0 2.8 42
289 2002 15.2 30.7 22 0 7.3 34
290 2002 12.2 30.2 21.3 0 5.5 58
291 2002 4.2 23.4 9.5 4.9 5.5 89
292 2002 4.2 10.8 6.4 0.7 5.3 79
293 2002 9.1 13.8 6.4 0.3 3.5 60
294 2002 15.8 10.3 5.2 0 4.3 51
295 2002 8.1 11.6 3.5 0 2.8 52
296 2002 14.6 15.9 5.6 0 5.5 46
```

Abstract architectural design



Workflow

The user:

- Drafts template
 - One template per syntax
- Annotates observables with terms from an ontology
- Defines the corresponding units of measurement
 - *Metadata file* is bound to a template

The reasoner:

- Parses *metadata files*
- Creates dataset axioms
- Infers compatibility



APSIM example: Input template

```
APSIM.MET — inputs
Add License

1 [weather.met.weather]
2
3 Latitude = 36.68
4 longitude = 116.98
5 tav = 14.6 (oC) ! annual average ambient temperature
6 amp = 28.2 (oC) ! annual amplitude in mean monthly temperature
7 !!!! 1/01/1961 to 31/12/2005
8 day year radn maxt mint rain wind RH
9 273 2002 17.5 27.2 14.6 0 3.5 54
10 274 2002 13.6 23.1 14.7 0 5.3 40
11 275 2002 15.8 27.1 11.1 0 5.5 29
12 276 2002 15.5 25.8 16.5 0 3.8 39
13 277 2002 14.9 25.5 14.6 0 2.5 63

Line: 12:47 | Plain Text | Tab Size: 4
```

```
APSIM.tmpl — templates
Add License

1 [weather.met.weather]
2
3 Latitude = {{station.latitude}}
4 longitude = {{station.longitude}}
5 tav = {{station.tags.tav}}
6 amp = {{station.tags.amp}}
7 !!!! 1/01/1961 to 31/12/2005
8 day year radn maxt mint rain wind RH
9 {%for timestamp, radn,maxt,mint,rain,wind,RH in chunk%}
10 {{timestamp.dayofyear}} {{timestamp.year}} {{radn.value}} {{maxt.value}} {{mint.value}} {{rain.value}} {{wind.value}} {{RH.value}}
11 {%endfor%}

Line: 1 | Plain Text | Tab Size: 4
```

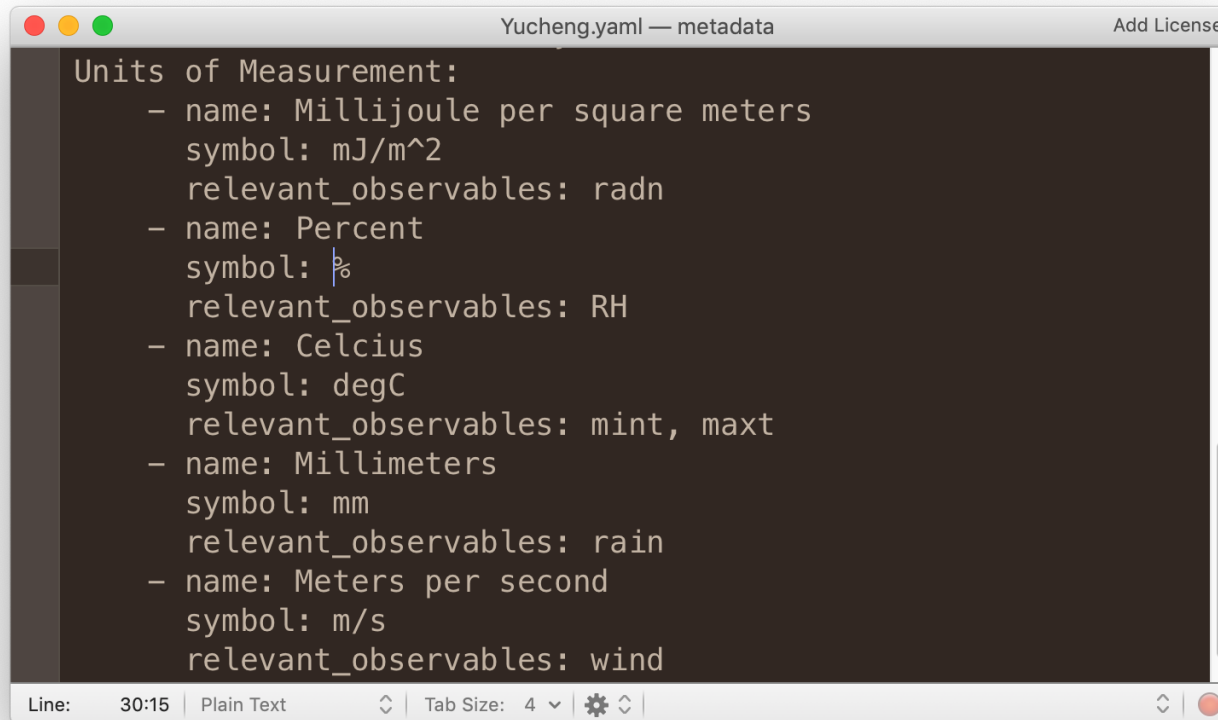


APSIM example: Metadata file (Observables)

```
Yucheng.yaml — metadata Add License
Station:
  name: Yucheng
  license: Attribution|
Observables:
  - observable_id: mint
    name: Temperature
    ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semadam.owl#Temperature
    qualifiers: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semadam.owl#min
  - observable_id: maxt
    name: Max Temperature
    ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semadam.owl#Temperature
    qualifiers: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semadam.owl#max
  - observable_id: rain
    name: Rain
    ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semadam.owl#Rain
  - observable_id: radn
    name: Solar radiation
    ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semadam.owl#SolarRadiation
  - observable_id: wind
    name: Wind
    ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semadam.owl#WindSpeed
  - observable_id: RH
    name: Relative humidity
    ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semadam.owl#RelativeHumidity
Line: 3:25 | Plain Text | Tab Size: 4 |  
```



APSIM example: Metadata file (UOMs)



The screenshot shows a text editor window titled "Yucheng.yaml — metadata" with a "Add License" button in the top right corner. The editor contains the following text:

```
Units of Measurement:  
- name: Millijoule per square meters  
  symbol: mJ/m^2  
  relevant_observables: radn  
- name: Percent  
  symbol: %  
  relevant_observables: RH  
- name: Celcius  
  symbol: degC  
  relevant_observables: mint, maxt  
- name: Millimeters  
  symbol: mm  
  relevant_observables: rain  
- name: Meters per second  
  symbol: m/s  
  relevant_observables: wind
```

The status bar at the bottom of the editor shows "Line: 30:15 | Plain Text | Tab Size: 4" and includes icons for search, settings, and a red circle.



Reasoner - Axioms and compatibility

Dataset A
ontology:observable1
ontology:observable1
ontology:observable2
ontology:observable3
ontology:observableN

Dataset B
ontology:observable1
ontology:observable2
ontology:observable3
...

ontology:observableN

- A is equal to B,
- A is a superset of B,
- A is **not compatible** to B



Implementation

- EDAM (`pip install edam`)
- Owlready2
 - Reasoner
- Pint
 - Units of measurement transformation



Demonstration

Observables	Datasets				
	APSIM	AgMiP	DSSAT	WOFOST	KNMI
Solar Radiation	radn (mJ/m^2)	SRAD (MJ/m^2)	SRAD (MJ/m^2)	irradiation (kJ/m^2)	Q (J/cm^2)
Avg Temperature	-	-	T2M ($^{\circ}C$)	-	TG ($d^{\circ}C$)
Max Temperature	maxt ($^{\circ}C$)	TMAX ($^{\circ}C$)	TMAX ($^{\circ}C$)	maxt ($^{\circ}C$)	TX ($d^{\circ}C$)
Min Temperature	mint ($^{\circ}C$)	TMIN ($^{\circ}C$)	TMIN ($^{\circ}C$)	mint ($^{\circ}C$)	TN ($d^{\circ}C$)
Precipitation	rain (mm)	RAIN (mm)	RAIN (mm)	precip (mm)	RH (dmm)
Wind speed	wind (m/s)	WIND (km/h)	WIND (m/s)	mwind (m/s)	FG (dm/s)
Relative Humidity	RH (%)	RHUM (%)	RH2M (%)	-	UG (%)
Dew Point Temperature	-	DEWP ($^{\circ}C$)	TDEW ($^{\circ}C$)	-	-
Vapor Pressure	-	vprs (hPa)	-	emvp (kPa)	PG ($dhPa$)



Reasoner in action - DSSAT and APSIM

The screenshot shows a web browser displaying an ontology editor interface. The browser address bar shows the URL: `edam (https://raw.githubusercontent.com/BigDataWUR/EDAM/master/edam.owl) : [/Users/argyris/.edam/edam.owl]`. The page title is `edam (https://raw.githubusercontent.com/BigDataWUR/EDAM/master/edam.owl)`. The interface includes a search bar and navigation tabs for `Active Ontology`, `Entities`, `Classes`, `Object Properties`, `Individuals by class`, and `DL Query`. The `Classes` tab is active, showing a class hierarchy for `APSIM`. The hierarchy is as follows:

- owl:Thing
 - Observables
 - Qualifiers
 - Templates
 - APSIM
 - DSSAT
 - KNMI
 - AgMIP
 - WOFST
 - KNMI
 - AgMIP

A green callout box with a black border is overlaid on the `DSSAT` class, containing the text: **DSSAT is APSIM superset**.

The right-hand pane shows the `APSIM` class description: `APSIM`. It lists several axioms under the `Equivalent To` section:

- Equivalent To
 - Templates
 - and (hasObservable some Rain)
 - and (hasObservable some RelativeHumidity)
 - and (hasObservable some SolarRadiation)
 - and (hasObservable some WindSpeed)
 - and (hasObservable some maxTemperature)
 - and (hasObservable some minTemperature)

A green callout box with a black border is overlaid on this section, containing the text: **Axioms**.

Other sections in the right pane include `SubClass Of` (with `Templates` listed), `General class axioms`, `SubClass Of (Anonymous Ancestor)`, `Instances`, `Target for Key`, `Disjoint With`, and `Disjoint Union Of`.

At the bottom right of the interface, it indicates `Reasoner active` and `Show Inferences` is checked.



Demonstration

Observables	Datasets				
	APSIM	AgMiP	DSSAT	WOFOST	KNMI
Solar Radiation	<u>radn (mJ/m^2)</u>	SRAD (MJ/m^2)	<u>SRAD (MJ/m^2)</u>	irradiation (kJ/m^2)	Q (J/cm^2)
Avg Temperature	-	-	T2M ($^{\circ}C$)	-	TG ($d^{\circ}C$)
Max Temperature	maxt ($^{\circ}C$)	TMAX ($^{\circ}C$)	TMAX ($^{\circ}C$)	maxt ($^{\circ}C$)	TX ($d^{\circ}C$)
Min Temperature	mint ($^{\circ}C$)	TMIN ($^{\circ}C$)	TMIN ($^{\circ}C$)	mint ($^{\circ}C$)	TN ($d^{\circ}C$)
Precipitation	rain (mm)	RAIN (mm)	RAIN (mm)	precip (mm)	RH (dmm)
Wind speed	wind (m/s)	WIND (km/h)	WIND (m/s)	mwind (m/s)	FG (dm/s)
Relative Humidity	RH (%)	RHUM (%)	RH2M (%)	-	UG (%)
Dew Point Temperature	-	DEWP ($^{\circ}C$)	TDEW ($^{\circ}C$)	-	-
Vapor Pressure	-	vprs (hPa)	-	emvp (kPa)	PG ($dhPa$)



Demo



WAGENINGEN UNIVERSITY

WAGENINGEN UR

Findings

- Integrating semantic heterogeneous timeseries is a manual and custom process
- Related modelling solutions use different vocabularies
- Declarative approaches can address semantic heterogeneity
- In EDAM, we took first steps towards:
 - Compatibility inference,
 - Unit transformation,
 - Automatic transformation



Available on pip
`pip3 install edam`

Questions?

Thank you!



Argyris Samourkasidis
argysamo@gmail.com