# SMEAR database

Micromet field course
Hyytiälä 5.9.2018

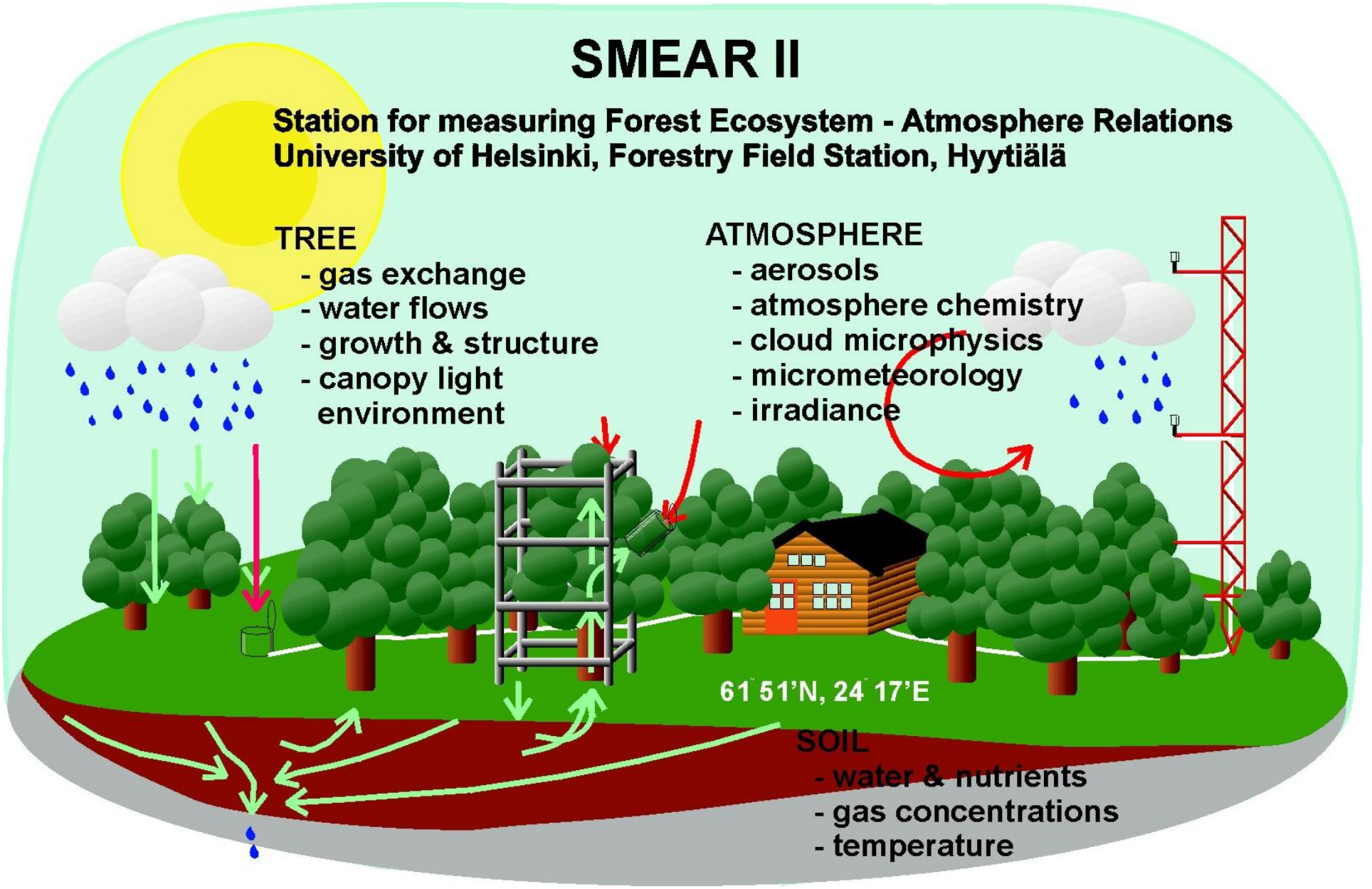
Pasi Kolari

# What is database?

- Any organized collection of data can be called database
- "Proper" relational database:
  - Data in table form (not as files!)
  - Built-in methods for cross-linking and searching different items in and across the tables and related to values of any other variables (such as time or quality flag):
    "select data in columns a,b from table x"
    "select data in columns a,b from table x where values of column c are between 1 and 2"
- Communication using relatively simple sets of commands
  - Structured Query Language (SQL)

# SMEAR II

**Station for measuring Forest Ecosystem - Atmosphere Relations**
**University of Helsinki, Forestry Field Station, Hyytiälä**



# SMEAR database

- Continuous field data from SMEAR stations, including satellite stations
- MySQL database ’smear’ with multiple tables
  - HYY_*, VAR_*, KUM_*, SII1_*, KVJ_*, ...
- Time is the primary key, time step in most tables 1 min
  - Turbulent fluxes (*_EDDY tables) are always 30 min averages
  - Some tables are constructed so that sample or instrument ID is also mandatory column value
  - Few infrequently measured variables with variable name itself as column value (table HYY_SLOW)
  - So far no geospatial data
- Quality level ([variablename]_EMEP) for each data record
  - 1=online processed, 2=processed and checked by expert

# SMEAR database

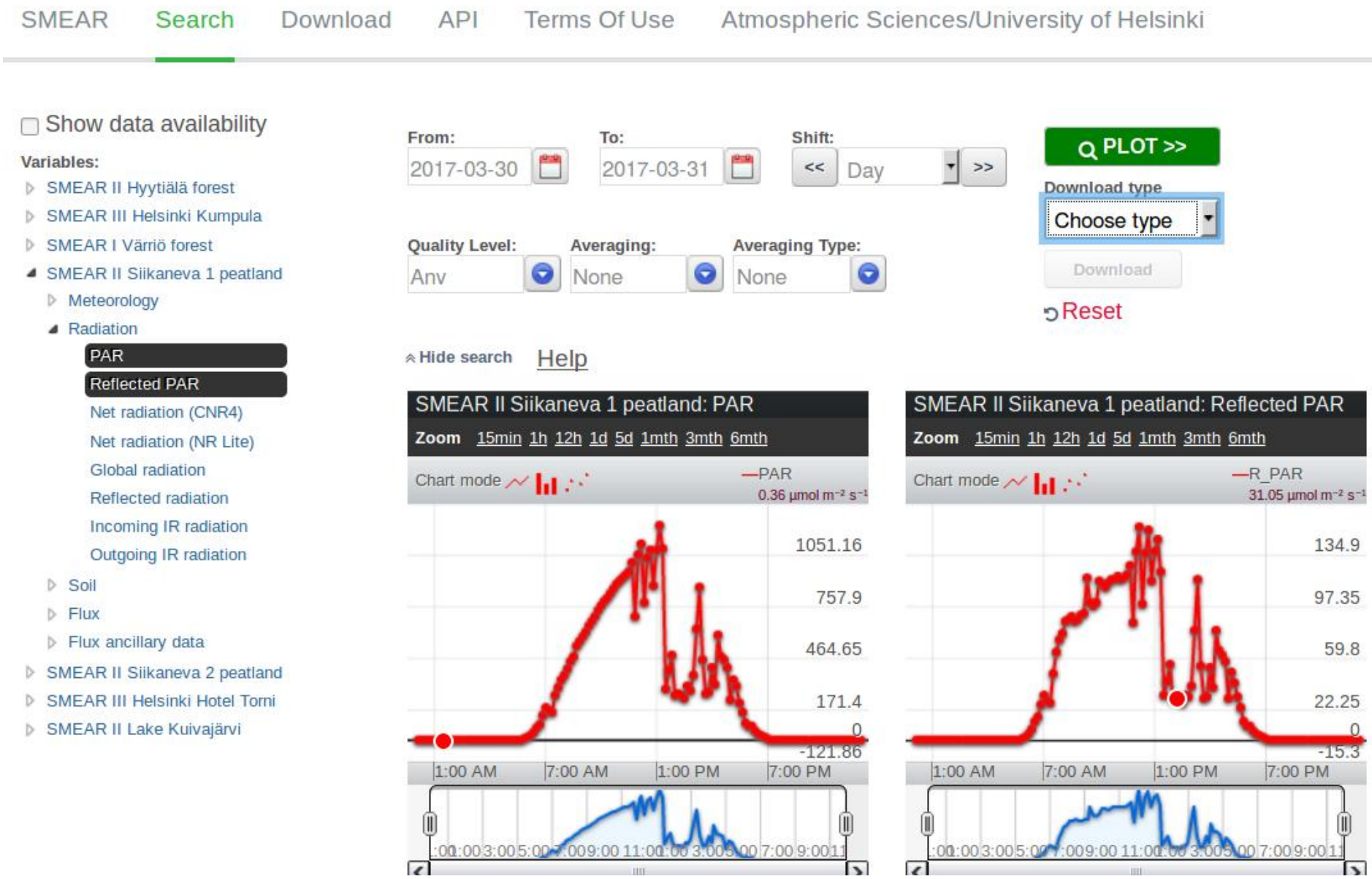
- Time coverage of data:
  - Instantaneous observation (most met & soil measurements)
  - Average or accumulation over 1 min or 30 min
- Not all variables are measured every minute
  à empty value field (NULL) in the table
- The database engine can apply basic math to the data
  - SELECT MIN(samptime),AVG(T168) FROM HYY_META WHERE samptime>=’2017-01-01’ AND samptime<’2017-01-02’ GROUP BY HOUR(samptime);
- SmartSMEAR employs native MySQL math and additional functionality
  - Different temporal averaging functions (be careful!)
- Basic documentation (metadata) for all variables, part of it can be embedded in downloaded data

# Communication with SMEAR database

- Direct connection through db connector, construct SQL queries with Python, R, Matlab...
  - possible in UH network, including VPN
- Graphical user (=human) interface (GUI)
  - Search and Download pages
  - accessible from anywhere
- Application programming interface (API)
  - accessible from anywhere

# SmartSMEAR search page

https://avaa.tdata.fi/web/smart/smear/search



# SmartSMEAR download page

https://avaa.tdata.fi/web/smart/smear/download

SMEAR | Search | Download | API | Terms Of Use | Atmospheric Sciences/University of Helsinki | SMEAR IV

Help

**Station**

- SMEAR II Hyytiälä forest
- SMEAR III Helsinki Kumpula
- SMEAR I Värriö forest
- SMEAR II Siikaneva 1 peatland
- SMEAR II Siikaneva 2 peatland
- SMEAR III Helsinki Hotel Torni
- SMEAR II Lake Kuivajärvi
- SMEAR IV Puijo

**Select variable category**

Radiation

Calculate availability on update

**From:** 2017-03-30 **To:** 2017-03-31

**Shift:** << Day >>

**Quality Level:** Any

| Variable | Description | Source |
|---|---|---|
| PAR | Photosynthetically active radiation in wavelength range 400-700 nm | Li-Cor Li-190 |
| Reflected PAR | Reflected photosynthetically active radiation | Li-Cor Li-190 |
| Net radiation (CNR4) | Net radiation (Kipp&Zonen CNR4) | Kipp & Zonen |
| Net radiation (NR Lite) | Net radiation (Kipp&Zonen NR Lite 2) | Kipp&Zo |
| Global radiation | Global shortwave solar radiation in wavelength range 0.3-4.8 µm | Kipp & Zonen |
| Reflected radiation | Reflected shortwave radiation in wavelength range 0.3-4.8 µm | Kipp & Zonen |
| Incoming IR radiation | Incoming far-infrared radiation in wavelength range 5-50 µm | Kipp & Zonen |
| Outgoing IR radiation | Outgoing far-infrared radiation in wavelength range 5-50 µm | Kipp & Zonen |

# Machine communication with SMEAR database

- SmartSMEAR is made for "manual" human use. When processing SMEAR data, it's often more convenient to download the data "on the fly".
- You can access all data in SMEAR database using any software with MySQL interface
  - access from university network, including VPN
  - Instructions in SMEAR wiki https://wiki.helsinki.fi/display/SMEAR/SMEAR+database
- API provides access from anywhere and includes all SmartSMEAR functionality plus additional metadata access
  - Detailed documentation on AVAA API page https://avaa.tdata.fi/web/smart/smear/api

# SMEAR API

- API (Application Programming Interface) =
  piece of code for translating software commands to the native language of the database
- OS is API between different software and file/memory access
- SMEAR API can be used with any software that can retrieve data from the web
  - web browsers
  - command line tools (curl, wget, ...)
  - math/programming software (Python, Matlab, ...)

# SMEAR API example

- SQL:
  Select min(samptime),avg(LE)
  from HYY_EDDY233 where samptime>'2014-07-01' and samptime<'2014-07-31'
  group by hour(samptime),
  dayofyear(samptime),year(samptime);

- API:
  http://avaa.tdata.fi/smear-services/smeardata.jsp?variables=LE&table=HYY_EDDY233&from=2014-07-01 00:00:00&to=2014-07-31 23:59:59&quality=ANY&averaging=60MIN&type=ARITHMETIC

- Try the data query builder on API documentation page!

# Result of API data request:

```
Year,Month,Day,Hour,Minute,Second,HYY_EDDY233.LE
    2014,7,1,0,0,0,15.633785
    2014,7,1,1,0,0,3.0540042
    2014,7,1,2,0,0,9.237995
    2014,7,1,3,0,0,8.260468
    2014,7,1,4,0,0,26.01663
    2014,7,1,5,0,0,23.121483
    2014,7,1,6,0,0,37.650288
    2014,7,1,7,0,0,78.8799
    2014,7,1,8,0,0,NaN
    2014,7,1,9,0,0,NaN
    ...
```

# Data documentation in SMEAR database

- Metadata tables for
  - variable properties
  - tables
  - stations
  - data lifecycle events
  - keywords

# SMEAR VariableMetadata

variableID links the variables to other metadata tables

tableID indicates the location of the variable in the database

variable is the column name

short human-readable title, description, unit, type of variable

source instrument (primary measurement) or source variables (derivative)

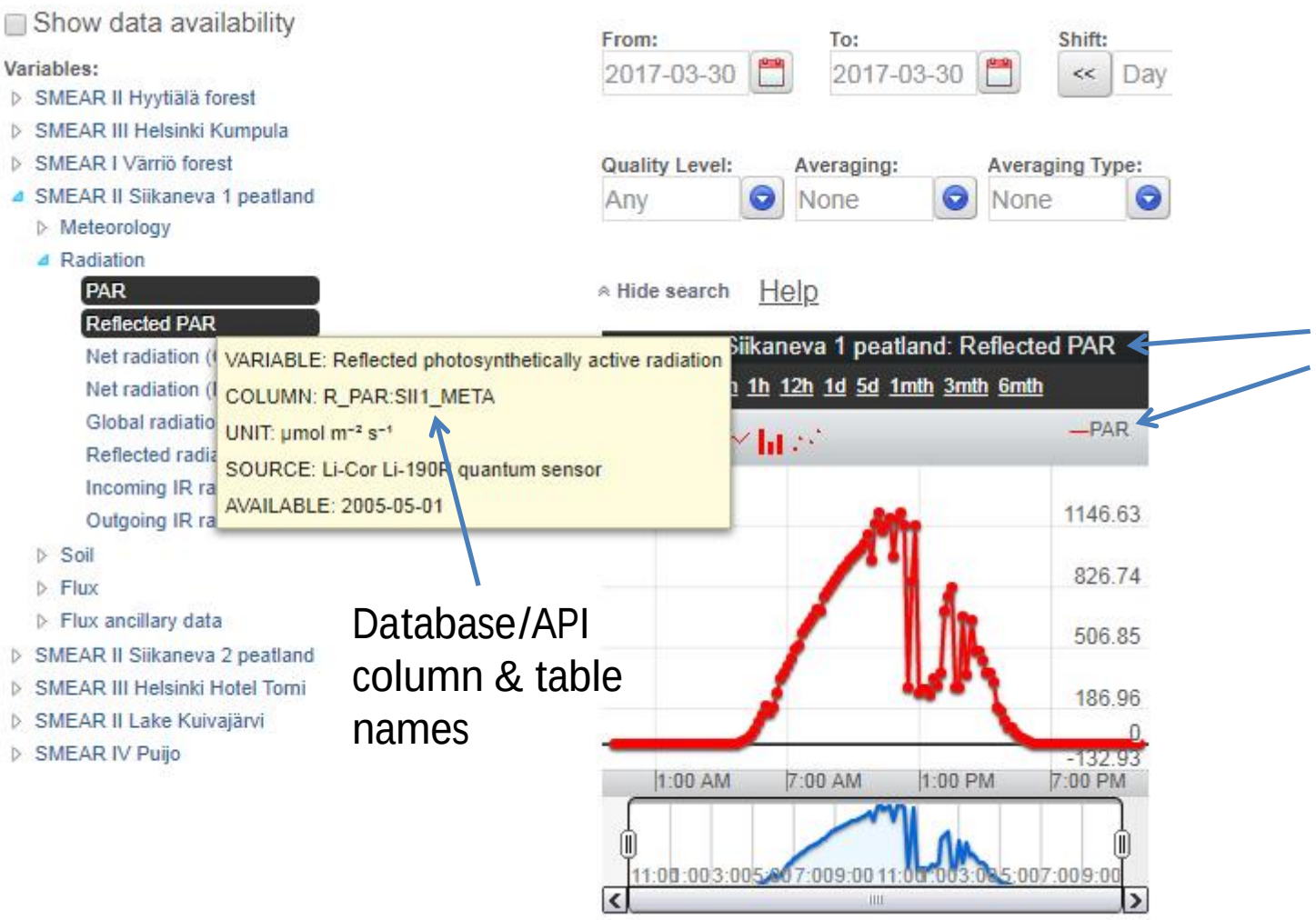
period_start/end; time coverage

| | |
|---|---|
| variableID | 81 |
| tableID | 4 |
| variable | WD504 |
| description | Horizontal wind direction at 50.4 m height, wind vane -8/2003 |
| type | double |
| unit | ° |
| title | Wind dir 50.4 m (vane) |
| source | Vector W200P wind vane |
| period_start | 1997-01-01 00:00:00 |
| period_end | 2003-08-27 23:59:59 |
| coverage | 0 |
| rights | public |
| category | Meteorology |
| mandatory | 0 |
| derivative | 0 |
| ui_avg_type | vectormean |
| vtimestamp | 2017-07-04 09:09:06 |

# Other metadata tables

- TableMetadata
  - table names and IDs
  - links to stations by stationID
- Station
  - stationID
  - coordinates
- Events
  - data lifecycle events
    - instrument changes and calibrations
    - data updates
- Tags
  - keywords in machine-readable format to help searching and interpreting the metadata

# Get variable name and metadata in GUI




Database/API column & table names

Lower title panel shows the correct variable!

# Get metadata via API

- One variable, need to know which:

  https://avaa.tdata.fi/smart-smear-portlet/variablemeta.jsp?tablevariables=HYY_META.WD504&allmeta=true

- Whole table HYY_META:

  https://avaa.tdata.fi/smart-smear-portlet/variablemeta.jsp?tableid=4

# Metadata via API request:


```
[{
    "_variableID":254,
    "_tableID":6,
    "_variable":"F_c_radtow",
    "_description":"Carbon dioxide flux, radiation tower 23 m height, secondary flux measuring
      setup",
    "_unit":"µmol m⁻² s⁻¹",
    "_title":"CO2 flux (2)",
    "_source":"Gill Solent HS1199 anemometer/thermometer & LI-COR LI-6262 gas analyzer",
    "_period_start":"Aug 5, 2001 8:15:00 PM",
    "_coverage":0,
    "_rights":"public",
    "_mandatory":false,
    "_derivative":true,
    "_vtimestamp":"Jan 15, 2015 7:34:33 AM",
    "_category":"Flux",
    "_cachedModel":false,
    "_new":false
}]
```