

# Metadata Needs for Ground Based In-Situ Measurements

Paul Eckhardt, NILU

Workshop for Air Quality and Atmospheric Composition (Dublin, 5-7 September 2012)



# Introduction

EBAS Database: Database for ground based in-situ measurements.

Initially designed as database for EMEP measurements, now used for dozens of different frameworks (EMEP, GAW-WDCA, ACTRIS, AMAP, HELCOM, OSPAR).

Reported files (NASA-Ames 1001 with extensions) ingested into relational database. All further data management as well as dissemination is done using this relational database.



# Scope of Presentation

Identifying needs for ground based in-situ measurements in general

- Data model
- Key metadata

Identifying specific needs of different communities we are serving (data provider view):

- Precipitation chemistry
- Aerosol chemistry
- Optical/microphysical aerosol properties
- Various trace gases (ozone, greenhouse gases, POP's)

Consider different data user points of view

# Data Model

## Dataset:

- A single timeseries of one parameter of an atmospheric property, considered consistent and comparable over time (no ruptures)
- Common core metadata for the whole dataset (may not change; neither for different periods nor through updates):

# Data Model

Time series generally consist of several submissions (for different periods).

For each submission, metadata may change to a certain extent.

Frequently version, version date (date of latest change), framework (network, restrictions, ...), any personal information (originator, submitter, ...)

Other metadata may change over time as well (as long as the constraint about consistency and comparability of data is not violated)

Metadata standards do not support variable metadata over time.

Ways to address this? Dataset series?

# Metadata Classification

## Types of metadata by usage:

- Discovery
  - Key descriptive (external) metadata
  - Strict conventions needed in order to enable automated discovery
- Determining fitness for use
  - Additional descriptive metadata
- Data access
  - Administrative metadata (technical metadata, revision info, access restrictions)
- Data usage
  - Additional descriptive metadata
  - Variable metadata (e.g. flags, other datasets) [metadata as part of the content]
  - Administrative metadata (provenance metadata, acknowledgement details)

# Discovery Metadata

## What?

- regime (i.e. ground based in-situ immission observation)
- matrix  
matrix (e.g. precipitation, air, aerosol, pm10, ...)
- variable
- statistical properties
- additional characteristics (multiple wavelengths, sizes etc)

## Where?

- country
- station code
- position (lat/lon/alt/height)

## When?

- start and end of timeseries
- time resolution

## How?

- instrument type

- Method



# Determining fitness for use

- Additional station metadata:
  - station name, codes in different frameworks, country, operator organization, type, setting, land-use, lat/lon/alt
- Data quality
- Detection limit
- Zero/negative values
- Instrument manufacturer / model / serial no.
- Standard conditions (temperature, pressure)
- Optical/microphysical aerosol measurements:
  - Inlet type (particle size)
  - Humidity and temperature control
- Aerosol filter measurements:
  - Sample preparation
  - Filter medium
  - Coating or solution

# Data Quality

## How to express data quality?

- Specification of uncertainty (precision, accuracy). This can be constant or variable (in extreme cases changing with each sample)
- Data Level
- Reference to SOP: possibility for very detailed description of data quality.
- Aerosol filter: blank correction

# Data Levels

## EBAS Data Levels:

- 0: instrument raw data, native resolution
- 1: processed physical parameters, native resolution
- 1.5: averaged to appropriate resolution (usually 1h)
- 2: manually quality assured
- 3: derived products

# Administrative Metadata

## Technical Metadata:

- file type
- access point

## Revision Information:

- Revision number
- Revision Timestamp

## Provenance:

- Organization
- Persons: PI, Submitter, DC Contact
- Instrument Operator Organization
- Laboratory or Processing Organization
- Framework

Data Policy, Restrictions, Acknowledgement

# Add. metadata for data usage

- Unit
- Missing value representation
- Data time zone, time zone offset
- Local time zone, time zone offset
- Flags (sample resolution)
  - Currently 130 different flags
  - 3 classes: valid, invalid, missing
- Associated Datasets, e.g.:
  - statistics, uncertainties
  - precipitation amount needed for concentrations
  - auxiliary data like (status information, flow rate, pressure, temperature, humidity)

# More Information:

<http://ebas.nilu.no>

<http://actris.nilu.no>

<http://www.gaw-wdca.org>

<http://www.nilu.no/projects/ccc/>