New Developments in ACTRIS Surface In-Situ Data Flow and Handling

M. Fiebig, C. Lund Myhre, A.M. Fjæraa, T. Hamburger

NILU - Norsk institutt for luftforskning
Norwegian Institute for Air Research
Proposed ACTRIS VOC Data Flow in a Nutshell

1. Data is quality assured at the station, including tools provided by ACTRIS.
2. Data are submitted to ACTRIS DC / EBAS via submission portal (ensures homogeneous format low in errors), associated with project “ACTRIS_preliminary”.
3. Data are moved to a read-only FTP site at ACTRIS DC for download by the ACTRIS VOC QA group.
4. ACTRIS VOC QA group checks data, and either:
   1. Logs issues found in issue tracker, and associates issue with responsible PI / submitter. Data are corrected by submitter, change documented in issue tracker. Back to 2.
   2. Finds the data to pass all QA tests, i.e. closes all issues on dataset (or creates a closed issue if pass at first submission). Go to 5.
5. Data are taken from FTP site, and imported into EBAS with project association “ACTRIS”.

Note: for NOx, the whole QA procedure will be handled by NOx QA centre. Will NOx QA centre give feedback via issue tracker?
Requirements for Organising Data QA and Feedback

2 related requirements when organising data QA and data feedback:

1. **VOC:**
   Issues found with submission of a station by QA group need to be recorded, station PI notified, correction followed up, DC notified when issues are solved.

2. **Aerosol**
   Issues are found when data are screened for larger studies (e.g. trend analysis, model inter-comparison). Issue needs to be documented, assigned to responsible PI, correction followed up, and corrected in archive.

→ Issue tracker at [http://ebas-feedback.nilu.no](http://ebas-feedback.nilu.no). To be used for project / framework QA, but also general feedback on EBAS / ACTRIS system.
What is an issue tracker?
• Web portal for organising and keeping track of issues and tasks in a project.
• Issues are described in text, attachments allowed for illustration & documentation.
• Issues can have 4 statuses reflecting work flow: open, assigned, resolved, closed.
• All users need to have an account!
• Each user has a “role” (set of rights): anybody can report issues or set issue to “resolved”
• Only “managers” (members of QA group) can assign or close an issue.

HOMEWORK:
• Every station / data PI & submitter to create account at ebas-feedback.nilu.no
• Use syntax for login name: <First Name>_<Last Name>
EBAS Feedback: Organising Quality Assurance and Feedback on Data
EBAS Feedback: Organising Quality Assurance and Feedback on Data
EBAS Feedback: Organising Quality Assurance and Feedback on Data
EBAS Feedback: Organising Quality Assurance and Feedback on Data
EBAS Feedback: Organising Quality Assurance and Feedback on Data
Quality Assurance Metadata: Why Do We Need That?

• So far, quality metadata has focused on uncertainty of individual data points.
• A huge fraction of ACTRIS’ work is around QA measures:
  - inter-comparison workshops
  - round-robin experiments
  - on-site inter-comparison
  - on-site audit
• ACTRIS markets itself as RI for high-quality data – we need to “show-off” and document that.
• Motivation for stations to participate in QA measures.
Metadata Documenting Quality Assurance

• Measure master data:
  - Unique ID of QA measure (issued by EBAS)
  - Type of QA measure (on site intercomparison, off-site intercomparison, round-robin, on site audit)
  - Entity issuing QA measure (name, address, URL)
  - Date of QA measure
  - Reference to test procedure used (title, URL).

• Requested from data provider:
  - Unique ID of QA measure (issued by EBAS)
  - Date of QA measure
  - Outcome:
    ▪ pass / no pass,
    ▪ uncertainty found (number, unit)
  - reference document (title, date, URL)
  - Will depend on component, at least for gases

• Consistent with ISO19115 DQ metadata items.

To Do: Defined Interface between QA Centre and Data Centre:

• Public documentation of QA measures and results at QA centre website.
• Documentation in machine-readable format
## Data Levels Defined in ACTRIS: Use for VOC & NOx?

<table>
<thead>
<tr>
<th>Data Level</th>
<th>Description</th>
<th>Used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>• Annotated raw data&lt;br&gt;• format instrument specific&lt;br&gt;• all data / information for processing to final value.</td>
<td>• contains all parameters provided by instrument as provided&lt;br&gt;• “native” time resolution</td>
</tr>
<tr>
<td>1</td>
<td>• processed to final parameter&lt;br&gt;• invalid data removed</td>
<td>• “native” time resolution&lt;br&gt;• format property specific</td>
</tr>
<tr>
<td>1.5</td>
<td>• aggregated to hourly averages&lt;br&gt;• variability quantified&lt;br&gt;• format property specific&lt;br&gt;• STP correction if necessary</td>
<td>auto-processed</td>
</tr>
<tr>
<td>2</td>
<td>• SOP describes steps from one to the next level.&lt;br&gt;• All levels use EBAS NASA-Ames format.</td>
<td>manual QA&lt;br&gt;regular collection</td>
</tr>
</tbody>
</table>

All levels use EBAS NASA-Ames format.
Why Do We Have Data Levels?

- **Traceability**: Whole chain of data acquisition / processing / QA can be traced back to the time of measurement.
- Allows to reprocess the data.
- Separates DAQ / processing / QA chain into well defined steps, great tool for finding the cause of failing intercomparison.
- Data is documented also for a user in 15 years from now.
- Higher level frameworks are moving to requiring this feature.
How to Make Use of Data Levels for NOx & VOCs

NOx:
• Similar to online aerosol observations: make use of all levels?
• Draft templates exist, expert / stake holder overhaul needed.

VOCs:
• Offline measurement, levels normally not used.
• Calibration & target gas measurements should be documented, but may confuse end-user.
• Use level 0 & 2, almost identical templates, except:
  Level 0: contains all calibration / zero / target gas measurements, respective data points use respective flags.
  Level 2: same data, but excluding calibration / zero / target gas measurements.
• Level 0 archived offline, level 2 accessible in EBAS-web.
• Expert / stake holder input needed.
Use of Flags in Data Submission: The General Philosophy

So far, data archives are often single purpose – ACTRIS is multi-purpose:

• Data are flagged invalid only with non-quantifiable instrument malfunction or severe local contamination (paint job on station, smoking under inlet) – non-representative even for close surroundings, can’t be corrected.

• Other conditions of local influence are to be flagged with respective flag, but data are reported as valid, e.g.:
  - Farming activity
  - Dust storm
  - Biomass burning episode
  - Volcanic eruption
## Comparison of Data Flags, 1/2

<table>
<thead>
<tr>
<th>Flag</th>
<th>V/I/M</th>
<th>Description</th>
<th>VOC use</th>
<th>NOx use</th>
<th>Aerosol use</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>V</td>
<td>Valid measurement</td>
<td>no obvious local influence or technical/instrumental problems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>V</td>
<td>Episode data checked and accepted by data originator.</td>
<td>-----</td>
<td>-----</td>
<td>Used to validate outlier</td>
</tr>
<tr>
<td>147</td>
<td>V</td>
<td>Below theoretical detection limit or formal Q/A limit, but a value has been measured and reported and is considered valid</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>185</td>
<td>V</td>
<td>Possible local contamination indicated by wind direction or velocity</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>186</td>
<td>V</td>
<td>Possible local contamination indicated by single scattering albedo</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>187</td>
<td>V</td>
<td>Possible local contamination indicated by occurrence of new particles.</td>
<td>-----</td>
<td>-----</td>
<td>Used in a setting where new particle formation indicates contamination)</td>
</tr>
<tr>
<td>188</td>
<td>V</td>
<td>Possible local contamination indicated by low wind speed.</td>
<td>-----</td>
<td></td>
<td>Used in case of auto-contamination by station</td>
</tr>
<tr>
<td>189</td>
<td>V</td>
<td>Possible local contamination indicated by wind from contaminated sector.</td>
<td>-----</td>
<td></td>
<td>Used if flag 185 is too unspecific.</td>
</tr>
<tr>
<td>390</td>
<td>V</td>
<td>Data completeness less than 50%</td>
<td>Needed for averaged values, SOP uses 66%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>392</td>
<td>V</td>
<td>Data completeness less than 75%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>394</td>
<td>V</td>
<td>Data completeness less than 90%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>410</td>
<td>V</td>
<td>Saharan dust event</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>459</td>
<td>I</td>
<td>Extreme value, unspecified error</td>
<td>-----</td>
<td></td>
<td>Unexplained extreme values, technical problem suspected.</td>
</tr>
</tbody>
</table>
## Comparison of Data Flags, 2/2

<table>
<thead>
<tr>
<th>Flag</th>
<th>V/I/M</th>
<th>Description</th>
<th>VOC use</th>
<th>NOx use</th>
<th>Aerosol use</th>
</tr>
</thead>
<tbody>
<tr>
<td>499</td>
<td>V</td>
<td>Inconsistent with another unspecified measurement.</td>
<td>-----</td>
<td>-----</td>
<td>Inconsistent e.g. scattering coefficient calculated from particle size distribution.</td>
</tr>
<tr>
<td>559</td>
<td>V</td>
<td>Unspecified contamination or local influence, but considered valid</td>
<td>unexpected values, no instrumental or technical failure.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>599</td>
<td>I</td>
<td>Unspecified contamination or local influence</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>640</td>
<td>V</td>
<td>Instrument internal relative humidity above 40%.</td>
<td>-----</td>
<td>-----</td>
<td>Used to signal possible hygroscopic growth of aerosol particles (neph, MPSS)</td>
</tr>
<tr>
<td>652</td>
<td>V</td>
<td>Construction / activity nearby.</td>
<td>-----</td>
<td>-----</td>
<td>Includes disturbance by other lab activity.</td>
</tr>
<tr>
<td>662</td>
<td>V</td>
<td>Too high sampling flow, data considered valid.</td>
<td>-----</td>
<td>-----</td>
<td>E.g. aerosol flow rate out of range but considered valid.</td>
</tr>
<tr>
<td>676</td>
<td>V</td>
<td>station inside cloud (visibility &lt; 1000 m).</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>677</td>
<td>I</td>
<td>Icing or hoar frost in the intake.</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>699</td>
<td>I</td>
<td>Mechanical problem, unspecified reason.</td>
<td>-----</td>
<td>-----</td>
<td>E.g. problems with flow, leaks.</td>
</tr>
<tr>
<td>797</td>
<td>V</td>
<td>Data element taken from co-located instrument</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>980</td>
<td>M</td>
<td>Missing due to calibration or zero/span check</td>
<td>-----</td>
<td></td>
<td>E.g. neph zero checks</td>
</tr>
<tr>
<td>999</td>
<td>M</td>
<td>Missing measurement, unspecified reason</td>
<td>missing measurements, outliers, technical problems, non-ambient contamination, definitely unusable for known reason (smoking, …)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>???</td>
<td>I</td>
<td>Missing due to zero/span check</td>
<td>Zero/span gas applied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>???</td>
<td>I</td>
<td>Missing due to target gas check</td>
<td>Target gas applied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>???</td>
<td>I</td>
<td>Missing due to calibration</td>
<td>Calibration gas applied</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2 Routes For Data Submission

Regular, annual submission:
• Submission of annual dataset by **30 June** the following year.
• Data shall have highest level of QA possible.
• Dataset assembled by data originator, because detail knowledge resides here.

Near-Real-Time Data Submission:
• Submission shall be collected and disseminated within max. 3 hours of measurement.
• Compromise on data quality due to auto-processing accepted.
• Alert service for provider if instrument “misbehaves”.
• Central data processing to ensure uniform dataset and implement alert service.
NRT data collection flow chart

**Data Centre:**
- check for correct data format (level 0).
- check whether data stays within specified boundaries (sanity check).

**Station:**
- auto-creates hourly data files (level 0).
- initiates auto-upload to NRT server.

**Sub-network data centre:**
- auto-creates hourly data files (level 0).
- initiates auto-upload to NRT server.

**Processing to level 1**
- Hourly level 1 data file

**Processing to level 1.5**
- Hourly level 1.5 data file

**EBAS database**

**Transfer**

**FTP transfer to data centre**

**Automatic feedback**

**Station:**
- collects raw data in custom format

**User access (restricted) via web-interface:**
- ebas.nilu.no

**User access via machine-to-machine web-service**
Near-Real-Time Reporting for Reactive Gases!

**NOx:**
- Technically, NOx data can be phased into the ACTRIS NRT infrastructure.
- Previous efforts (MACC) need to be considered.
- Who will be doing the processing, or specify the SOP for NRT processing?

**VOCs:**
- “NRT” VOC submission would in fact be rapid delivery, suitable for validation, but not operational services.
- NRT data submissions don’t receive feedback – import only when 100% correct.
- VOC rapid delivery would need dedicated tool for stations to submit 100% correctly formatted files.
Online Checks For Incoming (NRT) Data: Current Status

M. Fiebig

NILU - Norsk institutt for luftforskning
Norwegian Institute for Air Research
Online Outlier Check for Incoming Data

• Original plan: online check of VOC concentration ratios – proves difficult to specify.
• Plan B: implement outlier check for data submitted via submission portal
• Check will be based on comparison with running percentiles.
• Outliers found will need to be flagged valid (flag 110 - Episode data checked and accepted by data originator) or removed as invalid for file to be accepted.
• Fine-tuning threshold for outliers will be based on feedback from submitters.

BUT:
• How do we proceed with online check of VOC ratios?
Closure studies for particle light scattering coefficient

• Slowly degrading instrument performance often difficult to diagnose.
• Closure of scattering coefficient between nephelometer measurement and calculation from size distribution can help to diagnose instrument performance.
• Focus on NRT submissions, option for use on regular (annual) data submissions.
• Additional incentive to participate in NRT programme

Elaborate implementation options:
• Conditions at stations (observed size ranges, occurrence of coarse-mode particles) too heterogeneous to set strict thresholds for alerts.
• Instead, users receive weekly scatter plot of scattering coefficient measured over calculated from MPSS.
Why Is Increasing the Number of Near-Real-Time Stations All Important for ACTRIS?

M. Fiebig

NILU - Norsk institutt for luftforskning
Norwegian Institute for Air Research
“Quite simply, knowledge is the engine of our economy. And data is its fuel.”

(Neelie Kroes, former European Commission Vice President in charge of the Digital Agenda)

EU Commission’s Digital Agenda:

• Digital technologies are to be used for generating economic growth (next Facebook, Google, Air BnB, ...)

• Data has to flow as freely as possible to be available for new products and services.

• A whole General Directorate is working on this!

• Every project funded has to demonstrate how it fits into digital agenda (section on socio-economic impact).
Which will be the future users of Copernicus?

- EO Professionals
- Other Private Services (e.g. Transport, Tourism, ...)
- Public Authorities
- Policy and Decision Makers
- Research
- Education (including schools)
- ...
- Public (all citizens) ??

Open to new market opportunities – from few to many users

BUT

Each different category could have different requirements for data management

Source: Alessandro Annoni, JRC
Transforming observations into information

Satellite data
- Radiances and retrieved products from space agencies and collaborating institutions

In situ data
- From environmental agencies, EU projects and global networks

Other data relating to emissions

Acquisition of observations and pre-processing

Estimates of fire and other emissions

Global processing

Regional processing

User interface and supplementary services

Products
- Global records of the distributions, transport, sources and sinks of greenhouse and reactive gases, and aerosols
- Global forecasts of reactive gases and aerosols
- Detailed forecasts and assessments of air quality for Europe
- Stratospheric ozone, UV radiation and solar energy records and forecasts
- Support for policy and downstream services

Source: Vincent-Henri Peuch, ECMWF
The global observing system for atmospheric composition

Source: Vincent-Henri Peuch, ECMWF
Key elements of the data and information policy

- **Free, full and open access**

  - **No restriction on use nor on users**
    - Reproduction, redistribution with or without adaptation
    - Commercial and non-commercial purposes
  
  - **A free of charge version of any dataset is always available** (under pre-defined format on Copernicus dissemination platform)

  - **Worldwide without limitation in time**

Source: Astrid-Christina Koch, Copernicus
This policy applies to

A. **Data (and information) generated inside Copernicus**
   - Sentinel mission data
   - Service information

Does not apply to

B. **Data (and information) generated outside Copernicus**
   - Contributing Mission data
   - In situ and reference data and information

Copernicus sets the rule for A and follows (or negotiates) the rules for B set by the data providers.
Operational Services: What’s In It For The Station?

Success for ACTRIS-ESFRI:
• ACTRIS-ESFRI will establish ACTRIS (stations & facilities) as permanent infrastructure.
• Will be successful only if relevant (for digital agenda / Copernicus)

Funding & Visibility at national level:
• Your national funding agency can demonstrate they are contributing to Copernicus (with your station)
• Increased chances of getting national funding on ACTRIS ticket.

Visibility:
• Work towards visibility of network and station in final product.
Increasing the number of NRT stations and instruments

- Follow-up on stations that have NRT upload procedures ready.
- Take one instrument type at a time, focus first on nephelometers, then absorption photometers, then MPSS.

Within ACTRIS WP3, we should provide & share DAQ software that:
- collects data from instrument
- displays incoming data graphically,
- writes data to hard drive in level 0 format
- uploads hourly level 0 files to NRT account at NILU.
- auto-start of the software, e.g. after a power failure.

Need volunteers to write, provide, and maintain such software!!!
### Increasing the number of NRT stations and instruments

#### Volunteers:

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Responsible</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSI 3563 nephelometer</td>
<td>NILU (ready!)</td>
</tr>
<tr>
<td>ECOTECH nephelometer</td>
<td>U. Crete</td>
</tr>
<tr>
<td>MAAP</td>
<td>JRC?</td>
</tr>
<tr>
<td>Magee AE31</td>
<td>NILU</td>
</tr>
<tr>
<td>Magee AE33</td>
<td>NILU</td>
</tr>
<tr>
<td>Radiance Res. PSAP</td>
<td>NILU</td>
</tr>
<tr>
<td>TROPOS SMPS</td>
<td>TROPOS</td>
</tr>
<tr>
<td>Custom DMPS / SMPS</td>
<td>Stations themselves</td>
</tr>
</tbody>
</table>
AeroCom INSITU PNSD: Comparing Climate Model Particle Size Distributions with Surface InSitu Observations

M. Fiebig & S.M. Platt

NILU - Norsk institutt for luftforskning
Norwegian Institute for Air Research
Previous Work:

Western Europe, Mediterranean, and Arctic sites

ACTRIS-2 WP3 workshop Athens 10-12 November 2015