



What Is Data Quality?




What is "Data Quality"?
and
WHEN is data "quality" ?

A personal view

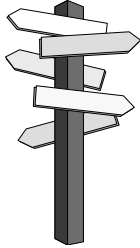
Theo Hafkenscheid
NMi van Swinden Laboratory
NL – DELFT



AAMG Dec 2006




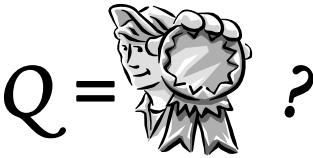
Roadmap of presentation




- ▶ Introduction
 - ▶ What is quality ?
- ▶ EU AAQ Legislation framework
 - ▶ What are requirements for data ?
 - ▶ What instruments are available ?
 - ▶ Implications for practice
- ▶ Summary
- ▶ Perspective

AAMG Dec 2006



Q =  **?** **What is "Quality" ?**


AAMG Dec 2006



Definitions (1)

- ▶ "Fitness for purpose"
- ▶ "Conformance to requirements"
- ▶ "Conformance to needs and expectations of customers"
- ▶ "Measure of conformance of a product or service to certain specifications or standards"

AAMG Dec 2006




Definitions (2)

- ▶ **ISO 9000**
"The degree to which a set of inherent characteristics fulfils requirements"
- ▶ **ISO 8402**
"The totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs"

→ "Entity" = datum / data
→ "Stated and implied needs" = ???

AAMG Dec 2006



Stated needs

- ▶ **Framework: EU Ambient Air Quality Legislation - CAFE Directive COM(2005)447**
- ▶ **Aims include**
 - ▶ assessing ambient air quality in Member States on the basis of common methods and criteria (**AQ assessment**)
 - ▶ providing information on ambient air quality ... and monitoring long-term trends and improvements; ensuring that this information is made available to the public (**information & reporting**)
 - ▶ maintaining ambient air quality where it is good and improving it in other cases (**AQ management**)

AAMG Dec 2006

What Is Data Quality?

NMi

Stated needs

▶ **Related to use / applications of data:**

- ▶ **Air Quality Assessment**
 - ▶ Measurements, modeling, objective estimation of concentrations of pollutants
- ▶ **Information and Reporting**
 - ▶ To the public (at least on a daily, and preferably hourly, basis for most pollutants)
 - ▶ To the European Commission (annually)
- ▶ **Air Quality Management**
 - ▶ Exceedances of limit values
 - ▶ Trend monitoring and analysis
 - Basis for Action Plans/Programmes

AAMG Dec 2006

NMi

Implied needs

▶ **Air Quality Assessment**

- ▶ Data need to be representative (for covered space and time)
- ▶ Data need to be "accurate"
- **But no more than needed for specific assessment regimes !**
- ▶ Data need to be comparable throughout EU

→ Use of uniform assessment strategies, consisting of combinations of measurement and modeling

→ Use of properly validated measurement, modeling and estimation methods under appropriate regimes of QA/QC

AAMG Dec 2006

NMi

Implied needs

▶ **Information and Reporting**

- ▶ Monitoring and modeling methods must have
 - ▶ a high time resolution (preferably much less than the required information interval)
 - ▶ a high time coverage (preferably no gaps in information provided)
 - ▶ a "near-direct" availability of accurate results (for information of the public on an hourly basis)

→ Application of accurate, automated, continuous measurement methods

→ Application of validated models to increase spatial resolution and to provide forecasting

AAMG Dec 2006

NMi

Implied needs

▶ **Air Quality Management**

- ▶ Exceedances of limit values
 - ▶ Data need to be representative
 - ▶ Data need to be "accurate" over stated averaging periods (e.g. 1 hour and 1 year for NO₂; 1 day and 1 year for PM₁₀)
- ▶ Basis for Action Plans/Programmes
 - ▶ Data need to be internally consistent over time (no leaps in concentration levels), e.g. for modeling purposes

→ Can one method fulfill all needs ?

→ Requires full traceability of results to accepted standards

AAMG Dec 2006

NMi

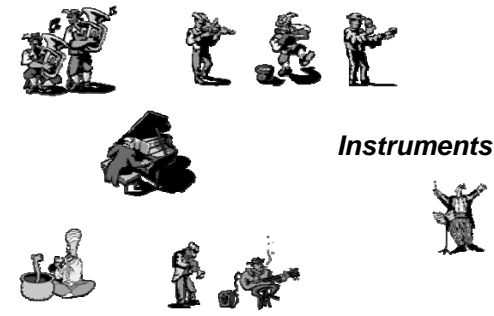
Summary of needs

▶ **Data need to**

- ▶ Be representative for time and space covered
- ▶ Have a high time resolution
- ▶ Have a high time coverage
- ▶ Be "directly" available
- ▶ Be "correct upon acquisition"
- ▶ Be accurate over different periods linked to reference periods of limit values and information frequencies
- ▶ Be comparable throughout EU
- ▶ Be internally consistent over prolonged time periods

AAMG Dec 2006

NMi



Instruments

AAMG Dec 2006

What Is Data Quality?

NMi

Instruments (1)

→ **CAFE Directive – COM(2005)447 specifies e.g.**

- ▶ Annexes III and V
- ▶ Criteria for siting of monitoring stations
 - ▶ Guidance available (Report “Guidance on Assessment”)
- ▶ Reference methods
 - ▶ Standardized and validated by CEN TC 264
- ▶ Possibility to use equivalent methods
 - ▶ Guidance available for demonstration of equivalence (Report “Guidance for the Demonstration”)

AAMG Dec 2006

NMi


Instruments (2)

→ **CAFE Directive – COM(2005)447 specifies e.g.**

- ▶ Data Quality Objectives for specific assessment regimes
 - ▶ Expanded uncertainty (in region of limit value)
 - ▶ Time coverage
 - ▶ Data capture
 - ▶ Required proportion (fraction) of valid data (ozone)
- ▶ Requirement to have “competent authorities” for a number of tasks specific to AQ assessment and management (article 3)

AAMG Dec 2006

NMi



Practical implementation

AAMG Dec 2006

NMi

Uncertainty

- ▶ Defined cf. ISO-GUM
- ▶ Determined cf. ISO-GUM and/or ISO 5725
- ▶ Documented in EN standards
- **But what is “the region of the limit value” ?**
- ▶ Usually interpreted as “at the level of the limit value”
- ▶ What about “above the UAT”, at least for usual measurement averaging periods ?

AAMG Dec 2006

NMi

Time coverage

- ▶ No formal definition
- ▶ = fraction of limit value reference period covered by measurement data (cf. Guidance Report on Preliminary Assessment) ?
- **No time coverage requirement for fixed measurements except for benzene**
- ▶ What’s the relationship with “data capture” ?

AAMG Dec 2006


NMi

Data capture

- ▶ No formal definition
- ▶ = 1 – fraction of instrument failure (Guidance Report on Preliminary Assessment) ?
- ▶ = availability (e.g. EN 14211) ?
- ▶ = 75% of usual averaging time (e.g. EN 14211) ?
- ▶ But Directives specify 90% !
- ▶ Should ≥90% of averaged values be available, e.g. ≥7884 hourly values per year for NO₂ ?
- **What about interpretation as for ozone (“proportion of valid data”) ?**
- **What are valid data ?**


AAMG Dec 2006

What Is Data Quality?

 **Valid data**


- ▶ No general criteria given
- ▶ Related to instrument status
- Frequent performance of span and zero checks in order to maximize data capture
- ▶ Negative values to be included for gases (EN 14211 etc.)
- ▶ But how negative ? $> -3 \cdot u_{zero}$?
- ▶ What about negative values for PM, e.g. large negative hourly values for β -monitors ?
- ▶ Strategies for dealing with aberrant values ?

AAMG Dec 2006

 **Data capture**


- ▶ Regular calibration and normal maintenance excluded
- ▶ What about span checks if concentration of calibration compound is traceable to accepted standards ?
- ▶ What is "normal" maintenance ? Preventive maintenance ?

AAMG Dec 2006

 **Relation between DQO**



- ▶ **Synergism/antagonism between data quality objectives ?**
 - ▶ Order of priority between objectives ?
 - ▶ Practice observed: time coverage/data capture are given priority
 - ▶ QC regimes are adapted to maximize TC/DC
 - ? Validation criteria ?

AAMG Dec 2006


 **Competence**

- EN-ISO 17025 accredited for measurements (i.e. testing) of National Reference Laboratories
- ▶ Other regional laboratories ?
- ▶ Certification of span gases ?
- ▶ Type approval testing (CEN/TC264 WG22 proposes EN-ISO 17025 accreditation) ?
- ▶ Equivalence testing ? Modeling ?
- **Harmonization of the concept of competence over all relevant areas**
- **How does "regional competence" compare throughout the EU over all relevant areas ?**

AAMG Dec 2006

  **Status summary**

AAMG Dec 2006

 **Air Quality Assessment**

- ! **Quality needs relatively well covered for measurements through**
 - ▶ (guidance for) assessment strategies
 - ▶ availability of standardized and validated implementations of reference methods
 - ▶ guidance for equivalence demonstration
 - ▶ regional, bi-lateral and EU-wide initiatives for organization of measurement comparisons
 - ▶ proposed accreditation, at least for National Reference Laboratories
 - ▶ existence of the Aquila NRL network

AAMG Dec 2006

What Is Data Quality?

NMi

Air Quality Assessment

- ? **Requirements for other “competent authorities”**
 - Regional laboratories
 - Type approval of instruments (CEN/TC264 WG22)
 - Modeling, objective estimation
 - Equivalence testing
- ? **Validity and QA/QC of models**
- ? **Accuracy of results for PM produced by automated continuous methods**
 - Type approval
 - Validation
 - Ongoing QA/QC

AAMG Dec 2006

NMi

Information and Reporting

- ! **Monitoring and modeling methods mostly fulfill requirements of high time resolution and time coverage**
- ! **Results are usually made available to the public on an hourly basis**
- ? **Quality of directly-published hourly results may be doubtful, especially for PM**
- ? **Differences in validation concepts lead to differences in reported (validated) results**

AAMG Dec 2006

NMi

Air Quality Management

- ! **Data generated for gases are generally representative and accurate, both over short-term and long-term periods**
- ? **Data generated for PM10 may not be accurate, and may suffer from leaps in time due to changes in equipment and correction factors**
- ? **Lack of harmonization/uniformity in data validation concepts leads to inconsistent EU datasets**

AAMG Dec 2006

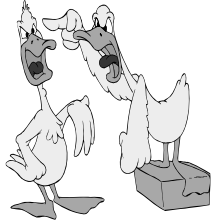
NMi

Focal points

- Uniform explanation of concepts of time coverage and data capture, also in relation to uncertainty
- Data validation
- Common concepts for competence and their implementations
- Type approval and QA/QC of automated continuous methods for PM
- Validation and harmonization of modeling

AAMG Dec 2006

NMi



Discussion ?

AAMG Dec 2006