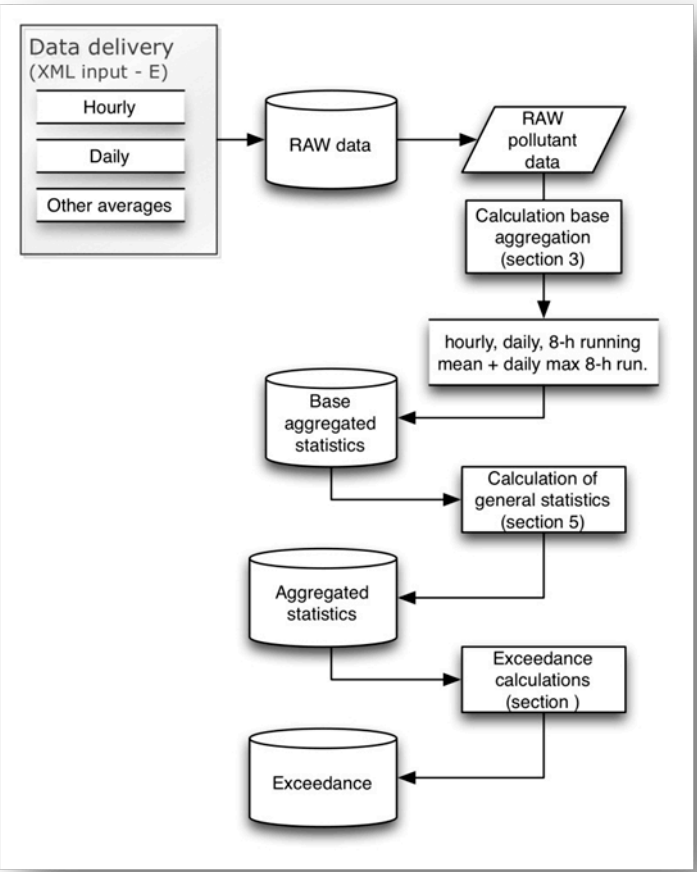
Aggregation rules for e-reporting

(Following Decision 2011/850/EU)

v 0.8.2



March 2017

Working Paper

Subtask 1.1.1.3 in ETC/ACM AP2017
February 2016
Jaume Targa, Tony Bush, Wim Mol

European Topic Centre
on Air Pollution and
Climate Change Mitigation

The European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM)
is a consortium of European institutes under contract of the European Environment Agency
RIVM UBA-V ÖKO AEAT EMISIA CHMI NILU VITO INERIS 4Sfera PBL CSIC

DISCLAIMER

This ETC/ACM Working Paper has not been subjected to European Environment Agency (EEA) member country review. It does not represent the formal views of the EEA.

© ETC/ACM, 2017.

ETC/ACM working Paper 2017

European Topic Centre on Air Pollution and Climate Change Mitigation

PO Box 1

3720 BA Bilthoven

The Netherlands

Phone +31 30 2748562

Fax +31 30 2744433

Email etcacm@rivm.nl

Website *http://acm.eionet.europa.eu/*

Document history

Version	Authors/ Reviewers	Date	Comments
0.0	Jaume Targa	07/05/2012	This is a working paper for EEA, EEA developers and ETC/ACM to describe aggregation and statistics to be calculated under the new e-Reporting process. This document initially describes the aggregation for SOR *-First stable draft following initial feedback*
0.1	Jaume Targa	21/05/2012	Incorporation of suggestions and comments following internal review
0.2	Jaume Targa	05/06/2012	Incorporation of suggestion and comments following comments from Wim Mol, Frank de Leeuw, Sheila Cryan, Libor Cernosky, Peder Gabrielsen.
0.3	Wim Mol/Jaume Targa	25/06/2012	Re-draft to incorporate wider aggregation rules
0.4	Wim Mol	27/06/2012	Comments/track changes processed in version 0.3
0.41	Jaume Targa	28/06/2012	Minor adds on general statistics tables + re-drafting of some sections for e-reporting
0.5	Wim Mol/Frank de Leeuw	28/06/2012	Some changes and comments processed in version 0.41
0.51	Tony Bush	12/07/2012	Re-structuring of some sections + re-drafting some sections + consistency of style
0.52	Jaume Targa	13/07/2012	Minor modifications
0.60	Tony Bush	13/07/2012	Minor modifications
0.62	Tony Bush	26/07/2012	Minor modifications, inclusion of weighted mean commentary from Wim Mol
0.63	Wim Mol	27/07/2012	Improvement formulas in 4.2.1, 4.2.2 and 5.1.2
0.64	Wim Mol/Frank de Leeuw	30/07/2012	Some small modifications
0.65	Tony Bush	08/08/2012	Minor modifications actioning EEA comments
0.66	Tony Bush	31/08/2012	Further minor modifications actioning EEA comments
0.67	Jaume Targa Tony Bush	24/09/2012	Minor modifications + commenting
0.68	Tony Bush	*Updates in 2013*	Further modifications
0.69	Jaume Targa	Updates in 2013	Further modifications
0.7	Jaume Targa	24/05/2014	Updated to include codelist for aggregation, XML example updated + update on other code list added
0.7.1	Jaume Targa	03/06/2014	Minor updates
0.8_draft	Jaume Targa	21/07/2015	Updated document to include: SOMO35, AOT40c-5 years, better description on how to calculate data capture, data coverage & time coverage, management of validity codes,
0.8	Jaume Targa	23/12/2015	Final
0.8.1	Jaume Targa	19/02/2016	Add clarification on SO2 winter mean

Version	Authors/ Reviewers	Date	Comments
0.8.2	Jaume Targa	22/03/2017	Add 99 percentile for PM_{10} and $PM_{2.5}$

Contents

1 Introduction 1
1.1 Structure of this document 2
1.2 Scripts for the calculation of aggregations and statistics 2
2 Configuration of the data delivery 3
2.1 Configuration of the data array 3
2.1.1 Validity and verification status flags 4
2.1.2 Encoding for automated measurements / observations 4
2.1.3 Encoding for sampler based multiday measurements / observations 5
3 Processing of raw data by EEA system 9
3.1 Storage of primary data & key attributes on records 9
3.1.1 Key meta-information 9
3.1.2 Pollutant & unit 9
3.1.3 Start and end time of the record 9
3.1.4 om:resultTime (time when record was generated by provider) 9
3.1.5 Validity and verification flag 9
3.1.6 Origin of the value (CDR official repository vs FTP/SOS) 10
3.2 Database storage of primary data 10
4 Calculation of base aggregations for hourly and daily values 12
4.1 Daily average (24-hour mean) 12
4.2 8-hour running mean 13
4.3 Daily maximum 8-hour running mean (daymax) 13
5 Calculation of data capture, time coverage & data coverage 15
5.1 Definition 15
5.1.1 Data capture (Dcap) 15
5.1.2 Time coverage (TC) 15
5.1.3 Data coverage (DC) 15
5.2 Data capture rates 16
5.2.1 Data capture calculations for hourly and daily data types 16
5.2.2 Data capture calculations from base aggregation data 17
5.2.3 Missing data fractions arising from maintenance and calibration 17
5.3 Data capture for multi day sampler based data types integrated over an averaging period 18
5.3.1 Data capture for multi day sampler based data types 18
5.3.2 Time coverage for multi day sampler based data types 19
6 Calculation of statistics 20
6.1 Averages and means 24
6.1.1 For hourly and daily data types 24
6.1.2 For multi-day sampler based data types 25
6.1.3 Winter mean 26
6.2 Other statistics for hourly and daily data 27
6.2.1 Percentiles 27
6.2.2 Maxima 27
6.2.3 k^{th} highest value (maxk) 28
6.2.4 AOT40 (only O_3) 29

6.2.5 SOMO35 (only O_3) *31*
6.2.6 AEI (only $PM_{2.5}$) (Average Exposure Indicator) *32*
6.2.7 National Exposure Reduction Target (NERT) *32*
6.3 Calculation of basic exceedance information 33
6.3.1 Exceedance of thresholds based on calendar year, winter mean and maximum daily eight hour mean aggregated statistics *34*
6.3.2 Number of hours or days with concentration > y µg/m³ *34*
6.3.3 Three consecutive hours with concentration > y µg/m³ *34*
6.4 XML encoding of aggregated statistics 35
7 Up-To-Date and Summer Ozone Reporting calculations 38
7.1 Data sources 38
7.2 Rounding off rules 38
7.3 Aggregations 38
7.4 Aggregating valid data 38
7.5 Verification of data. 38
7.6 Percentage of valid data 38
7.7 Calculation statistics on UTD data 39
7.7.1 ALT – NO2: Number of times when 3 consecutive hours have been above ALT environmental Objective of 400 µg/m³ *40*
7.7.2 ALT – SO2: Number of times when 3 consecutive hours have been above ALT environmental Objective of 500 µg/m³ *41*
7.7.3 LTO –Max. daily 8-hour mean - Days when max daily 8-hour mean exceeded the LTO in one year - 120 µg/m³ *42*
7.7.4 O3 - Health - INT - One hour - Hours in exceedance in a calendar year -180 µg/m³ *43*
7.7.5 O3 -> Health -> ALT -> One hour -> Hours in exceedance in a calendar year ->240 µg/m³ *44*
7.8 Additional aggregation on Summer Ozone Reporting and output report 44
7.8.1 LTO –Max. daily 8-hour mean – *46*
7.8.2 INT & ALT – hourly mean *47*
8 References 48
Annex A Environmental objectives and reporting metrics 49
Annex B Margin of tolerance and upper/lower assessment thresholds 52
Annex C Definition of the averaging time "var" 53
Annex D Historical methods for calculation of data capture and time coverage from multi day sampler based data types integrated over an averaging period 54
Annex E List of pollutants with Monitoring Objective 55

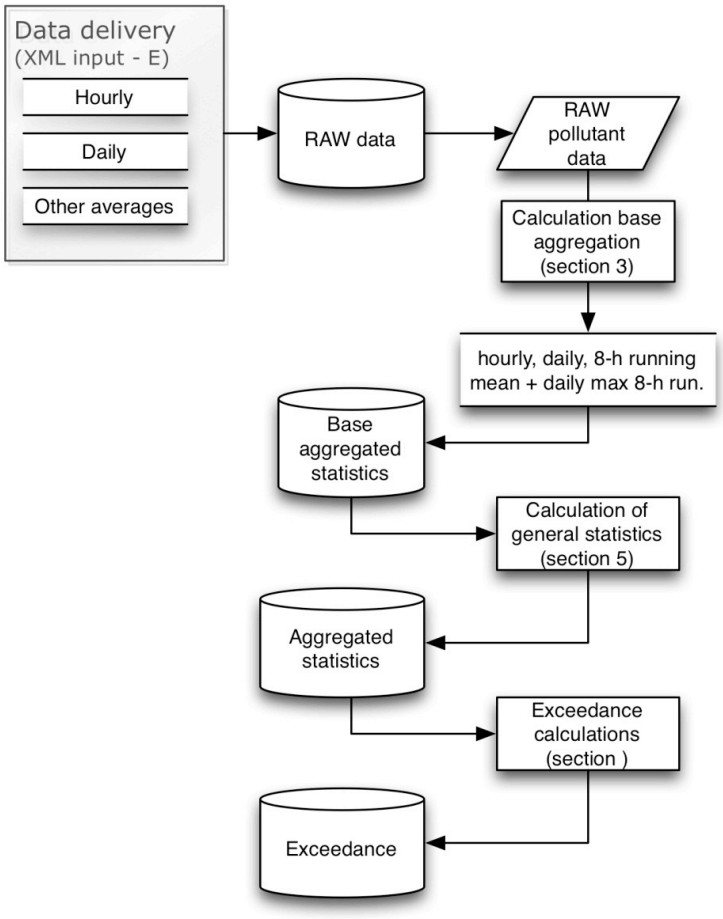
1 Introduction

This document provides the rules for calculation of aggregations and statistical values on air quality observations as specified by the Directives 2004/107/EC (EU, 2004) and 2008/50/EC (EU, 2008) and the derived Implementing Provisions 2011/850/EC (EU, 2011). It also provides the aggregation and statistical values to be calculated for the up-to-date data flows (UTD). The document is intended as a technical specification for developers who will write the software routines for EEA and therefore assumes no background knowledge of air quality legislation or air quality reporting.

The aggregations and statistics presented underpin the evidence base required to assess compliance with the Directives or demonstrate attainment of environment objectives. However, the scope of this document is limited to the calculation base aggregations and statistics, including some basic logic for the assessment of exceedance which is used for UTD in particular. It does not extend to the full logic and extensive rules for assessing of attainment of environment objectives; this will be the subject of a separate document.

Figure 1 presents an overview the life cycle for air quality measurements (and other sequence of events for data processing) within the e-Reporting system from the start-point (reporting) to end-point (calculation of exceedances for assessment of compliance).

Figure 1 Overview calculation base values, statistics and exceedances



The calculations are made on all primary (base) measurement data (observations) delivered (reported) by the countries under the Air Quality e-Reporting system. The rules presented in this document are based on those in operation within AirBase. Note that the aggregation rules are also applicable for modeled air quality data.

Statistical parameters are initially calculated for all sampling points irrespective of data capture rates (the proportion of valid data available). The aggregated data and statistics so derived are stored in a database without rounding together with associated information on the data capture, time coverage and data coverage. To support end-user data needs and for compliance assessment the statistical parameters may be selected and downloaded from the aggregated database based on user defined data capture, time coverage and data coverage selections including those specified by the Directive(s) data quality objectives (DQOs).

1.1 Structure of this document

Chapter 2 of this document presents an overview of the configuration of the data delivery (the reported observational data in XML). Chapter 3 defines first level of data processing, the calculation of base aggregations for hourly and daily values. Chapter 4 documents how the calculation of data capture and time coverage rates are to be made, these are important for downstream calculations. Chapter 5 defines how air quality compliance and general statistics are to be calculated. Chapter 6 document calculation for Up-to-date (UTD) and summer ozone reporting (SOR).

1.2 Scripts for the calculation of aggregations and statistics

In the previous data exchange systems (Exchange of information, EoI) several scripts were developed for the calculation of the aggregations and statistical values. These may be used are reference material to assist in the design and development of the new e-Reporting data handling and processing systems.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

2 Configuration of the data delivery

Primary measurement data will be reported (delivered) in XML following the specifications set out at http://www.eionet.europa.eu/aqportal. This schema provides generic rules for encoding of observations based on the OGC's O&M and SWE encoding standards. At present 2 encoding formats are proposed to accommodate the particular needs for

1. Automatic measurements / observations
2. Sampler based multiday measurements / observations

Detailed examples of the these generic data deliveries are presented in Figure 2 and Figure 3 and describe how measurements will (a) be encoded as part of a swe:dataArray, (b) important metadata on the observations will be marked-up using the OGC and ISO Observations and Measurements (O&M) conceptual model.

The core data elements of the schema are summarised in Table 1. These need to be recognised and stored in the database systems in order cross reference any derived statistics calculated with metadata associated with the measurement. Please refer to the "User Guide to XML" at YYYY for specific IT description of the reported XML.

Table 1 A description of the core components of the XML data deliveries

XML element	Description
om:OM_Observation gml:id	Provides a unique identifier for the group of observations
om:phenomenonTime	Provides information on the time period over which observation have been taken
om:resultTime	Provides a time stamp for when the result of the observation was generated
om:procedure	Documents configuration details of the equipment performing observations (including analytical details in appropriate)
om:parameter	Allows declaration of the sampling point (Samplingpoint_GB0729A_sp2) to associate with the observations
om:observedProperty	Defines the component being observed
om:featureOfInterest	Provides detailed information on the properties parcel of air being sampled
swe:elementCount	Defines the number of records to expect in the swe:dataArray
swe:dataRecord	Provides a mechanism for configuring the swe:dataArray. (further details given below).

2.1 Configuration of the data array

The primary measurement data to be imported and processed may be delivered at a variety of time resolutions depending on the measurement technique used. Automatic measurements (for example) will routinely be resolved to hourly and daily measurement periods and are regularly sorted. For these measurements the data array will be composed of 6 fields; a timestamp for the start of the measurement, a timestamp for the end of the measurement, a validity flag, a verification

flag and the concentration value. The timestamp references will use the full (extended) ISO-8601 format.

Sample based multiday measurements (e.g. sampler based measurements for Heavy Metals and PAH) may use other averaging times. **These can be weekly, 2-weekly, 4-weekly, monthly, 3-monthly, yearly, variable**. For these measurements the data array will be configured as follows; measurement start date/time, measurement end date/time, validity flag, verification flag and the % of valid period for the period covered by the observation. The time coverage for the period covered by the observation is calculated from start and end time. **If the % of valid period is not provided it will be assumed that the % is 100.**

If the averaging time deviates too much from a constant period, the averaging time receives a "var" flag. For the definition see Annex C.

2.1.1 Validity and verification status flags

Input data will be flagged according to the numbers below. The validity status specifies whether a value is acceptable to be used or whether it is invalid or missing. Verification status specifies whether the data has been checked and provides its status.

Table 2 Validity and verification status flags

Name	Code	Code number
Validity	Valid	**1**
	Valid below detection limit and the measurement value is given	**2**
	Valid below detection limit and number replaced by 0.5*detection limit	**3**
	Not valid due to station maintenance or calibration	**-99**
	Not valid due to other reasons	**-1**
Verification status	Not verified	**3**
	Preliminary verified	**2**
	Verified	**1**

See http://dd.eionet.europa.eu/vocabulary/aq/observationvalidity/view
http://dd.eionet.europa.eu/vocabulary/aq/observationverification/view

Editors note: Still to be discussed with EEA

Each value (measurement / observation) will have a validity flag and verification flag. The verification status (flag) for validated assessment datasets (Data flows E1a – measurements and E1b - modelling) must only be 1 (verified) when data has been fully verified and it is ready for compliance reporting. The verification status for UTD datasets (Data flows E2a) may be of any type.

Section 3 provides further information on how and where raw primary data is processed including validity and verification flags.

2.1.2 Encoding for automated measurements / observations

An example of a valid XML delivery for automatic measurements is presented in Figure 2.

This example clearly shows how measurements (observations) will be encoded as part of a swe:dataArray and also how important metadata related to the observations will be marked-up

using the O&M. The core data elements described in Table 1 are implemented and the data array configured to the specific requirements of automatic measurements and observations which have regularly spaced sampling/observation periods. As a result the swe:DataRecords within the array may be configured as shown in Figure 2 also at

EXAMPLE TO BE PROVIDED.

Where:

<swe:field name="startTime">	Specifies the time stamp of the start of the measurement period. All time references are reported using UTC and encoded in accordance with ISO 8601:2004 (E).
<swe:field name="endTime">	Specifies the time stamp of the end of the measurement period. All time references are reported using UTC and encoded in accordance with ISO 8601:2004 (E).
<swe:field name="Verification">	Allows for the declaration of the verification flag associated with the observation, , see section 2.1.1.
<swe:field name="Validity">	Allows for the declaration of the validity flag associated with the observation, see section 2.1.1.
<swe:field name="Value">	Allows for the declaration of the primary observation and the measurement unit.

2.1.3 Encoding for sampler based multiday measurements / observations

For sampler based observations or multiday observations, the observing period(s) may be irregularly spaced. As a result, the swe:dataArray is configured slightly differently to include:

1. the percentage of valid data in the sampling period

The core data elements described in Table 1 are implemented in the same way as for automatic measurements, as are the swe:DataRecords for validity, verification and the measurand quantity.

Where:

<swe:field name="startTime">	Specifies the time stamp of the start of the measurement period. All time references are reported using UTC and encoded in accordance with ISO 8601:2004 (E).
<swe:field name="endTime">	Specifies the time stamp of the end of the measurement period. All time references are reported using UTC and encoded in accordance with ISO 8601:2004 (E).
<swe:field name="Verification">	Allows for the declaration of the verification flag associated with the observation, , see section 2.1.1.
<swe:field name="Validity">	Allows for the declaration of the validity flag associated with the observation, see section 2.1.1.
<swe:field name="Value">	Allows for the declaration of the primary observation and the measurement unit.
<swe:field name="DataCapture">	Allows for the declaration of the data capture integrated over an observation (averaging) period as the proportion of valid measurement time relative total measured time

in the averaging period, expressed as an percentage. See section 5.3.1.This information is essential in calculating time weighted aggregated statistics e.g. the annual mean, see section 6.1.2.

See also;

http://cdr.eionet.europa.eu/gb/eu/aqd/e1a

Editorial note: Example UK data file will be provided when this data set is released at end Sept 2012.

Figure 2 An example XML encoding of a UTD data delivery from an automatic monitor observations measuring SO_2

```
<gml:featureMember>
        <om:OM_Observation gml:id="PrimaryUTDobservations_1">
                <om:phenomenonTime>
                        <gml:TimePeriod gml:id="ObserrvationTimePeriod_3">
                                <gml:beginPosition>2010-01-01T01:00:00+01:00</gml:beginPosition>
                                <gml:endPosition>2010-01-01T24:00:00+01:00</gml:endPosition>
                        </gml:TimePeriod>
                </om:phenomenonTime>
                <om:resultTime>
                        <gml:TimeInstant gml:id="ObservationResultInstant_3">
                                <gml:timePosition>2010-02-T14:00:00+01:00</gml:timePosition>
                        </gml:TimeInstant>
                </om:resultTime>
                <om:procedure xlink:href="xlink:href="http://environment.data.gov.uk/air-
quality/so/GB_SamplingPointProcess_2774"/>
                <om:parameter>
                        <om:NamedValue>
                                <om:name
xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/processParameter/AssessmentType”/>
                                <om:value>>"Fixed measurement from UK compliance assessment network selected by
UTD reporting”</om:value>
                        </om:NamedValue>
                </om:parameter>
                <om:parameter>
                        <om:NamedValue>
                                <om:name xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/processparameter/SP”/>
                                <om:value xlink:href="http://environment.data.gov.uk/air-quality/so/GB_SamplingPoint_6
                        </om:NamedValue>
                </om:parameter>
                <om:observedProperty xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/pollutant/8”/>
                <om:featureOfInterest xlink:href=" http://environment.data.gov.uk/air-quality/so/GB_SamplingFeature_2774"/:
                <om:result xsi:type="swe:DataArrayType">
                        <swe:elementCount>
                                <swe:Count>
                                        <swe:value>24</swe:value>
                                </swe:Count>
                        </swe:elementCount>
                        <swe:elementType name="FixedObservations">
                                <swe:DataRecord>
                                  <swe:field name="StartTime">
                                        <swe:Time definition="http://www.opengis.net/def/property/OGC/0/SamplingTi
                                        <swe:uom xlink:href="http://www.opengis.net/def/uom/ISO-8601/0/Gregorian"/
                                        </swe:Time>
                                  </swe:field>
                                  <swe:field name="EndTime">
                                        <swe:Time definition="http://www.opengis.net/def/property/OGC/0/SamplingTi
                                        <swe:uom xlink:href="http://www.opengis.net/def/uom/ISO-8601/0/Gregorian"/
                                                </swe:Time>
                                  </swe:field>
                                  <swe:field name="Verification">
                                        <swe:Category
definition="http://dd.eionet.europa.eu/vocabularies/aq/observationverification”/>
                                  </swe:field>
                                  <swe:field name="Validity">
                                        <swe:Category
definition="http://dd.eionet.europa.eu/vocabularies/aq/observationvalidity”/>
```

```
                                        </swe:field>
                                        <swe:field name="Value">
                                                <swe:Quantity definition="
http://dd.eionet.europa.eu/vocabulary/aq/primaryObservation/hour">
                                                <swe:uom code=" http://dd.eionet.europa.eu/vocabulary/uom/concentration/ug.n
                                                </swe:Quantity>
                                        </swe:field>
                                      </swe:DataRecord>
                                </swe:elementType>
                                <swe:values>2010-01-01T01:00:00+01:00,2010-01-01T02:00:00+01:00,1,1,15@@2010-01-
01T02:00:00+01:00,2010-01-01T03:00:00+01:00,1,1,15@@2010-01-01T03:00:00+01:00,2010-01-01T04:00:00+01:00,1,1,15@@
01-01T04:00:00+01:00,2010-01-01T05:00:00+01:00,1,1,15@@2010-01-01T05:00:00+01:00,2010-01-
01T06:00:00+01:00,1,1,15@@2010-01-01T06:00:00+01:00,2010-01-01T07:00:00+01:00,1,1,15@@2010-01-01T07:00:00+01:0
01-01T08:00:00+01:00,1,1,15@@2010-01-01T08:00:00+01:00,2010-01-01T09:00:00+01:00,1,1,15@@2010-01-
01T09:00:00+01:00,2010-01-01T10:00:00+01:00,1,1,15@@2010-01-01T10:00:00+01:00,2010-01-11T11:00:00+01:00,1,1,15@@
01-01T11:00:00+01:00,2010-01-01T12:00:00+01:00,1,1,15@@2010-01-01T12:00:00+01:00,2010-01-
01T13:00:00+01:00,1,1,15@@2010-01-01T13:00:00+01:00,2010-01-01T14:00:00+01:00,1,1,15@@</</swe:values>
                        </om:result>
                </om:OM_Observation>
</gml:featureMember>
```

Figure 3 An example XML encoding of a multiday data delivery from a sampler based observation (measured BaP)

```
<gml:featureMember>
        <om:OM_Observation gml:id="PrimaryUTDobservations_1">
                <om:phenomenonTime>
                        <gml:TimePeriod gml:id="ObserrvationTimePeriod_4">
                                <gml:beginPosition>2010-01-01T01:00:00+01:00</gml:beginPosition>
                                <gml:endPosition>2010-12-31T24:00:00+01:00</gml:endPosition>
                        </gml:TimePeriod>
                </om:phenomenonTime>
                <om:resultTime>
                        <gml:TimeInstant gml:id="ObservationResultInstant_4">
                                <gml:timePosition>2011-02-T14:00:00+01:00</gml:timePosition>
                        </gml:TimeInstant>
                </om:resultTime>

                <om:procedure xlink:href="xlink:href="http://environment.data.gov.uk/air-
quality/so/GB_SamplingPointProcess_2003"/>
                <om:parameter>
                        <om:NamedValue>
                                <om:name
xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/processParameter/AssessmentType”/>
                                <om:value>>"Fixed measurement from UK compliance assessment network selected by
UTD reporting”</om:value>
                        </om:NamedValue>
                </om:parameter>
                <om:parameter>
                        <om:NamedValue>
                                <om:name xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/processparameter/SP”/>
                                <om:value xlink:href="http://environment.data.gov.uk/air-quality/so/ GB_SamplingPoint_6
                        </om:NamedValue>
                </om:parameter>
                <om:observedProperty xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/pollutant/5029”/>
                <om:featureOfInterest xlink:href=" http://environment.data.gov.uk/air-quality/so/GB_SamplingFeature_2003"/:
                <om:result xsi:type="swe:DataArrayType">
                        <swe:elementCount>
                                <swe:Count>
                                        <swe:value>12</swe:value>
                                </swe:Count>
                        </swe:elementCount>
                        <swe:elementType name="FixedObservations">
                                <swe:DataRecord>
                                  <swe:field name="StartTime">
                                        <swe:Time definition="http://www.opengis.net/def/property/OGC/0/SamplingTi
                                        <swe:uom xlink:href="http://www.opengis.net/def/uom/ISO-8601/0/Gregorian"/
                                                </swe:Time>
                                 </swe:field>
                                 <swe:field name="EndTime">
                                        <swe:Time definition="http://www.opengis.net/def/property/OGC/0/SamplingTi
                                        <swe:uom xlink:href="http://www.opengis.net/def/uom/ISO-8601/0/Gregorian"/
```

```
                                                    </swe:Time>
                                                </swe:field>
                                                <swe:field name="Verification">
                                                    <swe:Category
definition="http://dd.eionet.europa.eu/vocabularies/aq/observationverification”/>
                                                </swe:field>
                                                <swe:field name="Validity">
                                                    <swe:Category
definition="http://dd.eionet.europa.eu/vocabularies/aq/observationvalidity”/>
                                                </swe:field>
                                                <swe:field name="Value">
                                                    <swe:Quantity definition="
http://dd.eionet.europa.eu/vocabulary/aq/primaryObservation/month">
                                                    <swe:uom code=" http://dd.eionet.europa.eu/vocabulary/uom/concentration/ug.n
                                                    </swe:Quantity>
                                                </swe:field>
                                                <swe:field name="DataCapture">
                                                    <swe:Quantity definition="
http://dd.eionet.europa.eu/vocabulary/aq/primaryObservation/dc">
                                                    <swe:uom code=" http://dd.eionet.europa.eu/vocabulary/uom/statistics/percentag
                                                    </swe:Quantity>
                                                </swe:field>
                                            </swe:DataRecord>
                                        </swe:elementType>
                                        <swe:values>2010-01-01T01:00:00+01:00,2010-01-31T24:00:00+01:00,1,1,1.58,100@@2010-02-
01T01:00:00+01:00,2010-02-28T24:00:00+01:00,1,1,1.324,100@@2010-03-01T01:00:00+01:00,2010-03-
31T24:00:00+01:00,1,1,1.261,70.97@@2010-04-01T01:00:00+01:00,2010-04-30T24:00:00+01:00,1,1,0.736,40@@2010-05-
01T01:00:00+01:00,2010-05-31T24:00:00+01:00,1,1,0.548,100@@2010-06-01T01:00:00+01:00,2010-06-
30T24:00:00+01:00,1,1,0.512,100@@2010-07-01T01:00:00+01:00,2010-07-31T24:00:00+01:00,1,1,0.356,100@@2010-08-
01T01:00:00+01:00,2010-08-31T24:00:00+01:00,1,1,0.462,100@@2010-09-01T01:00:00+01:00,2010-09-
30T01:00:00+01:00,1,1,0.894,93.33@@2010-10-01T01:00:00+01:00,2010-10-31T24:00:00+01:00,1,1,1.092,64.52@@2010-11-
01T01:00:00+01:00,2010-11-30T24:00:00+01:00,1,1,0.99,100@@2010-12-01T01:00:00+01:00,2010-12-
31T24:00:00+01:00,1,1,1.358,100@@</</swe:values>
                    </om:result>
            </om:OM_Observation>
</gml:featureMember>
```

3 Processing of raw data by EEA system

Primary measurements data reported (delivered) in XML by countries are processed by EEA in order to calculate statistics described in this document.

Primary data can be delivered via:

- Primary validated data (E1a) via CDR for official reporting
- Up to date primary data (E2a) via FTP or SOS services

3.1 Storage of primary data & key attributes on records

In order to keep full traceability of data delivered and correctly process it, key attributes are recorded by EEA for all records. These are essential in order to process incoming raw data and to calculate statistics.

For each record reported EEA stores the following:

- Key meta-information (SamplingPoint, SamplingPointProcess, Sample…)
- Pollutant & unit
- Start and end time of the record
- om:resultTime (time when record was generated by provider)
- Validity and verification flag
- Origen of the value (CDR official repository vs FTP/SOS)

3.1.1 Key meta-information

To be completed

3.1.2 Pollutant & unit

All records are stored with the corresponding pollutant code and unit.

3.1.3 Start and end time of the record

All records are stored with the corresponding start and end time recorded. All primary hourly data are stored as UTC+1.

3.1.4 om:resultTime (time when record was generated by provider)

All records are stored with resultTime provided by data provider. This is essential to prevent processing older data over newer datasets.

3.1.5 Validity and verification flag

According to section 2, all primary data must be provided with a validity flag. This validity flag is essential in order to correctly calculate any statistics and accompanying information like data capture/time coverage/data coverage. The validity flag is essential.

Table 3 Validity and verification status flags

EEA's general category	Validity Code	Code number
validdata	**1** **2** **3**	**Valid data will used to calculate any statistic.** **Valid data will also be used to calculate accompanying information like** data capture/time coverage/data coverage
notvalid_maint	**-99**	**Ignored when calculating any statistic.** **Used to calculate accompanying information like** data capture
notvalid_other	**-1**	**Ignored when calculating any statistic.** **Used to calculate accompanying information like** data capture
notvalid_null	0 missing	**Ignored when calculating any statistic.**

3.1.6 Origin of the value (CDR official repository vs FTP/SOS)

In order to keep full traceability of data processed, each value processed by EEA system is stored with a flag according to the origin of the data. This way data, data that has been provided as up-to-date data can be differentiated from CDR official data.

3.2 Database storage of primary data

Primary measurements provided will be processed in 3 different databases depending on the origin of the data and the aggregation purpose:

- E1a database which only includes data from CDR
- E2a database which only includes data from FTP/SOS
- E database which contains all data & aggregations

Key aspects of EEA's data processing:

- Essential links between D & primary data (both E1a & E2a)
- Processing of E1a data in E1a db
- Processing of E2a data in E2a db
- Processing of E1a & E2a in E db
- Aggregation in E db

The **verification status** flag is important for all calculations and provides 3 different options depending on the level of checking that has been conducted on the measurements by data provider. The verification status flag of the data must be considered when aggregating data.

1. Aggregations for regulatory compliance assessments must only contain valid and verified data records

2. Aggregations for other non-regulatory compliance assessments (e.g. UTD) may (will) contain unverified data too (codes 1, 2 and/or 3)

3. Sometimes it will also be useful to perform aggregations using all verification status types (codes 1-3) but provide % breakdown of the different status types used in the aggregation

4 Calculation of base aggregations for hourly and daily values

Base aggregations on raw hourly and daily data is to be carried out as detailed below, taking into account the proportion of valid data for each base aggregation. The data will be provided using codelist (within swe:field name="Value">) together with the unit (…vocabulary/uom/….):

- Raw hourly - **http://dd.eionet.europa.eu/vocabulary/aq/primaryObservation/hour**
- Raw daily - **http://dd.eionet.europa.eu/vocabulary/aq/primaryObservation/day**

KEY NOTE: *Section 5 specifies what base aggregation must be calculated for which pollutants. For optimisation of processes, only those necessary base aggregations are to be calculated.*

As an overarching rule for **all base aggregating periods**, if the proportion of valid data is below a 75% threshold, the aggregated value will be flagged as "Not valid because of insufficient data capture". This rule applies to hourly[1], daily and 8-hour based statistical aggregations. Invalid base aggregations although calculated and stored are not been taken into account in any the statistical calculations and must be flagged as "Not valid because of insufficient data capture". See Table 4.

Table 4 Aggregation flags for flagging the validity status of base aggregations

Code label	Code id	Concept URI
Valid	1	http://dd.eionet.europa.eu/vocabulary/aq/observationvalidity/1
Not valid because of insufficient data capture	-1	http://dd.eionet.europa.eu/vocabulary/aq/observationvalidity/-1

The calculation rules for the daily averages, 8-hour running means and daily maximum 8-hour running means are described below.

4.1 Daily average (24-hour mean)

Definition: The daily average or daily mean is the average of all valid hourly values for a day.
Minimum required proportion of valid data: a daily or 24-hour average is calculated if at least 18 valid hourly values are available in the 24-hour period starting at the hour ending 0100 hours.
If the **minimum required proportion of valid data is not available**, the daily or 24-hour average is not calculated for that particular day and an invalid aggregation flag shall be return (-1).
The result of this base aggregation is to be encoded as: http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1D
The raw primary data will be encoded like:

[1] Even though the shortest averaging time that will be reported for raw data is one hour, where the 1-hour average is made up from multiple integrated measurements the 75% rule applies; so, for example, if a reported hourly average we composed of 60 1-minute observations , at 45 valid minute observations are required. These checks are the responsibility of the data provider.

http://dd.eionet.europa.eu/vocabulary/aq/primaryObservation/hour

The daily average shall be calculated for pollutants with environmental objectives based on daily values (SO_2, PM_{10} and $PM_{2.5}$) and for which hourly values have been delivered as part of the raw data delivery.

The inclusion of calculation of daily averages for all pollutants with hourly data is recommended as part the end user data interface requirements. These recommendations and interface requirements are a separate deliverable under ETC/ACM Task 1.0.1.3.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

4.2 8-hour running mean

Definition: The 8-hour running mean value for each hour is calculated as the average of the valid hourly values for that hour and the 7 previous hours (therefore cumulatively and 8-hour averaging period). Hence, the averaging period of $hour_1$ of day_n is $hour_{17}$ of day_{n-1} until $hour_1$ of day_n (inclusive). The averaging period of $hour_{24}$ of day_n is $hour_{16}$ of day_n until $hour_{24}$ of day_n (inclusive).
Minimum required proportion of valid data: an 8-hour running average concentration is calculated if at least 6 valid hourly values are available over the 8-hour averaging period.
If the **minimum required proportion of valid data is not available**, the 8-hour running average is not calculated for that particular hour and an invalid aggregation flag shall return.
The result of this base aggregation is to be encoded as: http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P8H-run
The raw primary data will be encoded like: **http://dd.eionet.europa.eu/vocabulary/aq/primaryObservation/hour**

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

4.3 Daily maximum 8-hour running mean (daymax)

Definition: The daily maximum 8-hour running mean is the maximum of the valid 8-hour running means for that day. Calculation of all the 8-hour running means (above) for a given day is a pre-requisite. A day runs from $hour_1$ to $hour_{24}$ as defined above.
Minimum required proportion of valid data: a daily maximum 8-hour mean is calculated if at least 18

valid 8-hour running means are available for that particular day.
If the **minimum required proportion of valid data is not available**, the daily maximum 8-hour running mean is not calculated for that particular day and an invalid aggregation should return.
The result of this base aggregation is to be encoded as: http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P8H-dmax
The raw primary data will be encoded like: http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P8H-run

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

The 8-hour running mean and daily maximum 8-hour running mean shall be calculated for O_3 and CO. Twenty-four 8-hour running means and one daily maximum 8-hour running mean shall be calculated per each day.

5 Calculation of data capture, time coverage & data coverage

KEY NOTE: *All statistics, calculated from the primary data or from the base aggregations, described in section 5 shall be accompanied with information on data capture, time coverage and data coverage.*

These will be provided for all aggregations in order for user to select data appropriately for their needs. If aggregation is used for compliance purposes, the user needs to use those percentages required under the Directive and latest IPR Guidance from DG Environment.

5.1 Definition

The definitions of these 3 statistics are key in order to understand the aggregation results. In order to automate the calculation across all pollutants, reported primary data and statistics, the 3 accompanying statistics are define as below.

5.1.1 Data capture (Dcap)

The data capture is the percentage of valid measurements over the total number of reported measurements over the averaging period.

For continuous measurement monitoring, EEA expects full delivery across all measuring period with appropriate validity flag. See section 2.1.1 in order to understand how to report data. Please note that if full delivery are not provided for continuous measurements (hourly/daily), the data capture calculated by EEA might differ from reporters calculations. If full data delivery is not provided, please use data coverage to obtain the % of valid measures over the averaging period.

5.1.2 Time coverage (TC)

The time coverage is the percentage of measurement time in a given averaging period.

5.1.3 Data coverage (DC)

The data coverage is the percentage of valid measurements over the given averaging period.

When assessing the appropriateness of any statistics, data coverage is considered to be most appropriate statistic to understand the representativeness of a statistic over an averaging period.

For all aggregations, the following is “counted”:

- count_validdata (validity flag 1, 2, 3)
- count_notvalid_maint (validity flag -99)
- count_notvalid_other (validity flag -1)
- count_notvalid_null (validity flag 0 or missing)

For aggregation, Dcap DC & TC are calcuated following these algorithms (**PLEASE NOTE THIS UPDATES PREVIOUS GUIDANCE**):

DataCapture = [count_validdata] / ([count_validdata]+[count_notvalid_maint]+[count_notvalid_other])

DataCoverage = [count_validdata] / [interval]

TimeCoverage = [count_validdata]+[count_notvalid_maint]+[count_notvalid_other] / [interval]

[ie. Data capture x Time Coverage = Data Coverage]

5.2 Data capture rates

5.2.1 Data capture calculations for hourly and daily data types

The data capture for hourly and daily measurements is the percentage of valid measurement values in a given data set. The data capture in a given averaging period is defined as follows:

$$\textit{Data capture} = N_{valid}/N_{totalReported} * 100\ \%$$

where N_{valid} is the number of valid hourly/daily values (validity flag > 0[2]) and $N_{totalReported}$ is the number of hours/days reported in that measurement period.

Example [*tc1*] when calculating an annual mean from NO_2 primary hourly values
N_{valid} = 7724 h
$N_{totalReported}$ = 8760 h (full dataset reported for averaging period)
Time Coverage = 100%
Data Capture = 88,17 %

Example [*tc3*] when calculating an annual mean from PM10 primary daily values
N_{valid} = 309 d
$N_{totalReported}$ = 365 d (full dataset reported for averaging period)
Time Coverage = 100%
Data Capture = 84,66 %

Note: Where a measurement method delivers hourly values, N_{valid} and $N_{totalReported}$ shall be calculated using number of hours. Where a measurement method delivers daily values, N_{valid} and $N_{totalReported}$ shall be calculated using number of days.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

If data capture needs to be encoded separately to a statistic, the aggregation type can use the following codes (Internal EEA system codes for data capture):

[2] *Valid primary data or valid base aggregation data will be flagged using 1, 2 or 3 according to codelist* ***http://dd.eionet.europa.eu/vocabulary/aq/observationvalidity/1***

- Data capture of hourly values in 1 day - http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1D-dc

- Data capture of hourly/daily/other values in 1 year - http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y-dc

5.2.2 Data capture calculations from base aggregation data

The data capture for base aggregation data is calculated following same logic as shown in section 5.2.1. The data capture is the percentage of valid aggregated values in **a given averaging period**. The data capture in a given averaging period is defined as follows:

$$Data\ capture = N_{valid}/N_{averaging_period} * 100\ \%$$

where N_{valid} is the number of valid hourly/daily values (validity flag > 0[3]) and N_{total} is the number of hours/days in that measurement period.

Example: A year O_3 daily maximum values
N_{valid} = 315 d
N_{total} = 365 d
Time Coverage = 100%
Data Capture = 86,3 %

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

If data capture needs to be encoded separately to a statistic, the aggregation type can use the following codes (Internal EEA system codes for data capture):

- Data capture of hourly/daily/other values in 1 year - http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y-dc

- Data capture of daily maximum values in 1 year – http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y-dx-dc

5.2.3 Missing data fractions arising from maintenance and calibration

The percentage of values missing due to station maintenance or calibration in a given averaging period is calculated as follows:

$$Perc_{main_cal} = N_{main_cal}/N_{averaging_period} * 100\ \%$$

[3] *Valid primary data or valid base aggregation data will be flagged using 1, 2 or 3 according to codelist* ***http://dd.eionet.europa.eu/vocabulary/aq/observationvalidity/1***

where N_{main_cal} is the number of invalid hourly or daily values with a validity flag = -99 (Not valid due to station maintenance or calibration) and $N_{averaging_period}$ is the number of hours/days in the averaging period.

Example: A year NO_2 hourly values
N_{main_cal} = 300 h
$N_{averaging_period}$ = 8760 h
$Perc_{main_cal}$ = 3,42 %

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

5.3 Data capture for multi day sampler based data types integrated over an averaging period

For handling multi day sampler based data types, we are proposing to change the format and content of the data delivery to include information on the % of valid data within the sampling period. The calculation of these parameters for aggregations may be performed as follows.[4]

5.3.1 Data capture for multi day sampler based data types

The data capture for multi day sampler based data types integrated over an averaging period is calculated as the proportion of valid measurement time (at a minimum hours, alternatively days) relative **total measured time in the averaging period**, expressed as an percentage.

KEY NOTE: *The data capture for multi day sampler based data types is to be provided via the extended version of the swe:array*

The data capture for a given averaging period shall be calculated as follows:

$$Data\ capture = \Sigma_i DC_i * N_{valid,i} / \Sigma_i N_i * 100\ \%$$

Where

DC_i = the data capture (in %) of the valid sample I, that is, the effective sampling time in a measuring period[5]

$N_{valid,i}$ = the number of hours in the valid sample i (validity flag sample > 0)

N_i = number of hours in sample i

[4] This represents a departure from the way in which AirBase has traditionally performed these calculations which is presents for information and context in Annex D.
[5] For example, to avoid overflow of the filter during a measuring period of 8 days, active sampling is done during 4x24 hours. In this case the data capture for this period is 50%.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

5.3.2 Time coverage for multi day sampler based data types

The time coverage is the percentage of measurement time in a given averaging period. Also the time coverage is calculated on base of the number of hours in the sample periods. The time coverage for a given averaging period is calculated as follows (weighted average):

$$\textit{Time coverage} = \Sigma_i TC_i * N_i / N_{averaging_period} * 100\ \%$$

Where

TC_i = the time coverage of sample i

N_i = number of hours in sample i

$N_{averaging_period}$ = the total number of hours in the given averaging period (e.g. a year)

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

6 Calculation of statistics

This section provides a description of **compliance assessment and general statistics** that are required by the e-Reporting system. The statistical are parameters are calculated from the raw and/or base aggregations presented in section 3. The statistics to be calculated on a pollutant and SamplingPoint basis and the base parameters used are presented in the array. All statistics shall be accompanied by an associate % data capture rate as specified in section 5.

KEY NOTE: *The table details all the KEY aggregation (necessary for checking compliance with the different AQ Directives and the IPR Decision to be calculated at SamplingPoint (i.e station) level. The EXTRA statistics might be useful for Air Quality Assessement.*

KEY ANNUAL (or multi-annual) STATISTICS FOR DATA PROVIDED HOURLY/DAILY

Component ****		Aggregation based on: Hourly values (…/aq/primaryObservation/hour)		Daily values (…/aq/primaryObservation/day) (…/aq/aggregationprocess/P1D)		Daily maximum 8-hour running mean (aggregated from hourly) (…/aq/aggregationprocess/P8H-dmax)		Non hourly/daily (sample) data (…/aq/primaryObservation/week) (…/aq/primaryObservation/fortnight) (…/aq/primaryObservation/month) (…/aq/primaryObservation/quarter) (…/aq/primaryObservation/var)	
Sulphur dioxide (SO_2) CODE: 1	KEY STATS	• annual mean*	…/P1Y	• *days with c > 125 µg/m3*	P1Y-daysAbove125			• *annual mean weighted average*	**…/P1Y-WA-avg**
		• hours with c > 350 µg/m^{3} *	…/P1Y-hrsAbove350	• annual mean**	P1Y			• *winter mean weighted average*	**…/winter-WA-avg**
		• winter mean*	…/winter-avg	• *winter mean* **	winter-avg				
		• periods of 3 consecutive hours c > 500 µg/m3*	…/P1Y-3hAbove500						
	EXTRA	• 25th highest hourly value*	…/P1Y-hr-max25	• *4th highest daily value*	P1Y-day-max4				
		• 99.73 percentile*	P1Y-hr-max-per99.73	• *99.18 percentile*	P1Y-day-max-per99.18				
		• 50 percentile*	…/P1Y-hr-per50	• *daily maximum*	P1Y-day-max				
		• hourly maximum*	…/P1Y-hr-max	• *daily minimum*	P1Y-day-min				
		• hourly minimum*	…/P1Y-hr-min						
Nitrogen dioxide (NO_2) CODE: 8	KEY	• annual mean*	…/P1Y	• annual mean**	P1Y			• *annual mean weighted average*	**…/P1Y-WA-avg**
		• hours with c > 200 µg/m3*	…/P1Y-hrsAbove200						
		• periods 3 consecutive hours c > 400 µg/m3*	…/P1Y-3hAbove400						
	EXTRA	• 19th highest hourly value*	…/P1Y-hr-max19						
		• 99.79 percentile (max19)*	…/P1Y-hr-per99.79						
		• hours c > 400 µg/m3*	…/P1Y-hrsAbove400						
		• 50 percentile*	…/P1Y- hr-per50						
		• hourly maximum*	…/P1Y-hr-max						
		• hourly minimum*	…/P1Y-hr-min						
Nitrogen oxides (NO_x) CODE: 9	KEY	• annual mean*	…/P1Y					• *annual mean weighted average*	**…/P1Y-WA-avg**
	EXTRA	• 50 percentile*	…/P1Y- hr-per50						
		• hourly maximum*	…/P1Y-hr-max						
		• hourly minimum*	…/P1Y-hr-min						
Ozone (O_3) CODE: 7	KEY	• hours with c > 180µg/m^{3}*	…/P1Y-hrsAbove180			• *days with c >120 µg/m3*	**…/P1Y-dmaxAbove120**		
		• hours with c > 240µg/m^{3}*	…/P1Y-hrsAbove240			• *days with c >120 µg/m3 averaged over 3 years*	**…/P3Y-dmaxAbove120**		
		• AOT40*	…/AOT40c						
		• AOT40 averaged over 5 y*	…/AOT40c-P5Y						
	EXTRA	• AOT40 forest protection*	…/AOT40f			• *Maximum daily 8-hour max in a year*	**…/P1Y-dx-max**		
		• 50 percentile*	…/P1Y- hr-per50			• *26th highest 8h day max*	**…/P1Y-dx-max26**		
		• annual mean*	…/P1Y			• *93.15 percentile (max26)*	**…/P1Y-dmax-per93.15**		
		• hourly maximum*	…/P1Y-hr-max			• *day max average over 1y*	**…/P1Y-dx-avg**		
		• hourly minimum*	…/P1Y-hr-min						
		• summer average*	…/summer-avg						
		• SOMO35*	…/P1Y-SOMO35						

KEY ANNUAL (or multi-annual) STATISTICS FOR DATA PROVIDED HOURLY/DAILY										
	Aggregation based on									
Component ****	**Hourly values** (…/aq/primaryObservation/hour)		**Daily values** (…/aq/primaryObservation/day) (…/aq/aggregationprocess/P1D)		**Daily maximum 8-hour running mean** (aggregated from hourly) (…/aq/aggregationprocess/P8H-dmax)		**Non hourly/daily** (sample) data (…/aq/primaryObservation/week) (…/aq/primaryObservation/fortnight) (…/aq/primaryObservation/month) (…/aq/primaryObservation/quarter) (…/aq/primaryObservation/var)			
Carbon monoxide (CO) CODE: 10					• *days with c >10 mg/m3*	**…/P1Y-8hdmxAbove10**				
	EXTRA	• annual mean* • 50 percentile* • hourly maximum* • hourly minimum*	…/P1Y …/P1Y- hr-per50 **…/P1Y-hr-max** **…/P1Y-hr-min**			• *day max average over 1y* • *Maximum daily 8-hour max over 1y*	**…/P1Y-dx-avg** **…/P1Y-dx-max**			
Particulate matter (PM_{10}) CODE: 5	KEY	• annual mean*	**…/P1Y**	• *annual mean*** • *days with c > 50 µg/m³* • *90.4 percentile*	**…/P1Y** **…/P1Y-daysAbove50** **…/P1Y-P1D-per90.4**			• *annual mean weighted average*	**…/P1Y-WA-avg**	
	EXTRA	• daily mean* • 50 percentile* • hourly maximum* • hourly minimum*	**…/P1D** …/P1Y- hr-per50 …/P1Y-hr-max **…/P1Y-hr-min**	• *90.41 percentile* • *99 percentile* • *36th highest daily value* • *50th percentile* • *daily maximum* • *daily minimum*	**…/P1Y-P1D-per90.41** **…/P1Y-P1D-per99** **…/P1Y-day-max36** **…/P1Y-day-per50** **…/P1Y-day-max** **…/P1Y-day-min**					
Particulate matter ($PM_{2.5}$) CODE: 6001	KEY	• annual mean* • interim yearly AEI* • AEI*	**…/P1Y** **…/AEI-P1Y** **…/AEI**	• *annual mean*** • interim yearly AEI • *AEI*	…/P1Y …/AEI-P1Y …/AEI			• *annual mean weighted average*	**…/P1Y-WA-avg**	
	EXT.	• daily mean* • 50 percentile • hourly maximum* • hourly minimum*	**…/P1D** …/P1Y- hr-per50 …/P1Y-hr-max **…/P1Y-hr-min**	• 99 percentile • 50 percentile • daily maximum • *daily minimum*	**…/P1Y-P1D-per99** …/P1Y-day-per50 …/P1Y-day-max …/P1Y-day-min					
Benzene (C6H6) CODE: 20		• annual mean* • 50 percentile* • hourly maximum* • hourly minimum*	**…/P1Y** **…/P1Y-** hr-**per50** **…/P1Y-hr-max** **…/P1Y-hr-min**							
Arsenic (As)PM_{10} CODE: 5018 Cadmium (Cd) PM_{10} CODE: 5014 Nickel (Ni) PM_{10} CODE: 5015 Lead (Pb) PM_{10} CODE: 5012 Benzo(a)pyrene in PM_{10}	KEY	•		*annual mean***	**…/P1Y**			• *annual mean weighted average*	**…/P1Y-WA-avg**	
	EXTRA			• 50 percentile • daily maximum • daily minimum	**…/P1Y-day-per50** **…/P1Y-day-max** **…/P1Y-day-min**					

KEY ANNUAL (or multi-annual) STATISTICS FOR DATA PROVIDED HOURLY/DAILY							
	Aggregation based on						
Component ****	**Hourly values** (…/aq/primaryObservation/hour)		**Daily values** (…/aq/primaryObservation/day) (…/aq/aggregationprocess/P1D)		**Daily maximum 8-hour running mean** (aggregated from hourly) (…/aq/aggregationprocess/P8H-dmax)	**Non hourly/daily** (sample) data (…/aq/primaryObservation/week) (…/aq/primaryObservation/fortnight) (…/aq/primaryObservation/month) (…/aq/primaryObservation/quarter) (…/aq/primaryObservation/var)	
CODE: 5029							
Pollutants with Monitoring Objective **** (see annex E)	KEY • annual mean*	**…/P1Y**	• *annual mean***	**…/P1Y**			
	EXTRA • daily mean*	**…/1D**	• 50 percentile	**…/P1Y-per50**			
	• 50 percentile*	**…/P1Y-hr-per50**	• daily maximum	**…/P1Y-day-max**			
	• hourly maximum*	**…/P1Y-hr-max**	• daily minimum	**…/P1Y-day-min**			
	• hourly minimum*	**…/P1Y-hr-min**					
All other pollutants	EXTRA • annual mean*	**…/P1Y**	• *annual mean***	**…/P1Y**			
	• daily mean*	**…/P1D**	• 50 percentile	**…/P1Y-day-per50**			
	• 50 percentile*	**…/P1Y-hr-per50**	• daily maximum	**…/P1Y-day-max**			
	• hourly maximum*	**…/P1Y-hr-max**	• daily minimum	**…/P1Y-day-min**			
	• hourly minimum*	**…/P1Y-hr-min**					
NOTES:	* If hourly values reported		*** If daily values reported*			*** if values are not hourly or daily*	

***** Component codelist is http://dd.eionet.europa.eu/vocabulary/aq/pollutant/[code]*

Editors note: Guidance from AQUILA on calculations is pending

The following sections mathematically describe the statistics to be calculated as per table above.

6.1 Averages and means

The annual averages (means) shall be calculated using the shortest reported averaging time available (see Directive 2008/50/EC (EU, 2008) Annex XI). Under this premise, the preference for calculation of annual means can be ranked as follows; hourly, daily, weekly, monthly etc. (sampler data may have irregularly resolved averaging periods, the highest resolution averaging times available shall be used on a case by case basis).

6.1.1 For hourly and daily data types

The mean for any given averaging period shall be calculated as follows:

$$Mean = \Sigma_i C_i / N_{valid}$$

where C_i is the valid hourly/daily/daymax concentration and the summation is over all valid hourly/daily/daymax values measured in that averaging period. N_{valid} is the total number of valid hourly/daily/daymax values in the period.

As noted above, the mean shall be calculated from the data series with the highest time resolution, for example, if both hourly and daily values are available the mean shall be calculated from the hourly data.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

The resultant aggregation should be encoded as (see XML example in 6.4):

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1D

  (see benchmarking example: tc2 from hourly)

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y

  (see benchmarking example: tc1 from hourly & tc3 from daily)

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y-dx-avg

  (see benchmarking example: tc2_dymax from daymax)

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/summer-avg

  (see benchmarking example: tc1 from hourly)

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/winter-avg

  (see benchmarking example: tc8)

6.1.2 For multi-day sampler based data types

The mean for a multi-day sampler based data types shall be evaluated in a similar way to the data capture (see section 5.3), although in this instance is calculated as the time weighted mean of valid measurements over the sample period.

$$Weighted\ mean = \Sigma_i C_{valid,i} * N_{averaging_period} * DC_i / \Sigma_i N_{averaging_period} * DC_i \quad i=1.... n$$

Where

$C_{valid,i}$ = concentration for valid sampling period i

$N_{averaging_period}$ is the number of seconds in the averaging period.

DC_i = the data capture of sample i as a % (default = 100% if not provided) (i.e % of valid data within the individual averaging period)

When the sampling time falls across two calendar years, the sampling time refers to time in the current reporting year only

KEY NOTE: The data capture for multi day sampler based data types is to be provided via the extended version of the swe:array

Start of sampling period	End of sampling period	Measured concentration ng/m3 (C)	Validity flag ($_{valid}$)	Valid data in period (%) (DCi)	Seconds in period ()	Weighted conc. (taking into account DL flag, ng/m3) (X * Y * Z)	Weighted sampled time?? (Y * Z)
01/01/2011 00:00	01/02/2011 00:00	3.90532	1	100	31	12106.49	3100
01/02/2011 00:00	01/03/2011 00:00	1.40378	1	100	28	3930.58	2800
01/03/2011 00:00	01/04/2011 00:00	0.976502	1	70.97	31	2148.37	2200.07
01/04/2011 00:00	01/05/2011 00:00	0.129041	1	40	30	154.85	1200
01/05/2011 00:00	01/06/2011 00:00	0.078314	1	100	31	242.77	3100
01/06/2011 00:00	01/07/2011 00:00	0.088956	1	100	30	266.87	3000
01/07/2011 00:00	01/08/2011 00:00	0.077084	1	100	31	238.96	3100
01/08/2011 00:00	01/09/2011 00:00	0.140637	1	100	31	435.97	3100
01/09/2011 00:00	01/10/2011 00:00	0.070045	1	93.33	30	196.12	2799.9
01/10/2011 00:00	01/11/2011 00:00	0.338542	1	64.52	31	677.12	2000.12
01/11/2011 00:00	01/12/2011 00:00	1.41943	1	100	30	4258.29	3000
01/12/2011 00:00	01/01/2012 00:00	1.09143	1	100	31	3383.43	3100
Sum					365	28039.84	32500.09
Annual Mean						0.8627620	

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

The resultant aggregation should be encoded as (see XML example in 6.4):

- **http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y-WA-avg**

(see benchmarking examples tc4, tc5, tc6 & tc7)

6.1.3 Winter mean

The winter averages shall be calculated using the shortest reported averaging time available like the annual means.

The mean for shall be calculated as follows:

$$Mean = \Sigma_i CW_i / NW_{valid}$$

where CW_i is the valid hourly/daily/var concentration and the summation is over all valid hourly/daily/var values measured over winter period (1st October Year X-1 to 31st March Year X). NW_{valid} is the total number of valid hourly/daily/var values in the winter period.

As noted above, the mean shall be calculated from the data series with the highest time resolution, for example, if both hourly and daily values are available the mean shall be calculated from the hourly data.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

The resultant aggregation should be encoded as (see XML example in 6.4):

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/winter-avg

Clarification :

The winter mean corresponding to year X is calculated over the winter months (i.e. 1 October year x-1 to 31March year x)

Example:

Winter mean corresponding to year 2014 corresponds to the mean over the period with a begin date 2013-10-01T01:00:00+01:00 and an end date 2014-03-31T24:00:00+01:00.

6.2 Other statistics for hourly and daily data

6.2.1 Percentiles

The y^{th} percentile should be selected from the valid values (for hourly/daily/daymax concentrations). All the values shall be ranked in increasing order:

$$X_1 \leq X_2 \leq X_3 \leq \ldots \leq X_k \leq \ldots \leq X_{N-1} \leq X_N$$

The y^{th} percentile is the concentration X_k, where the value of k is calculated as follows:

$$k = (q.N)$$

Where q is equal to $y/100$ and N the number of valid values.

The value of k $(q.N)$ shall be rounded off to the nearest whole number (values < 0.499999… are rounded to 0, values = 0.5 are rounded to 1).

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

The relationship of important percentiles for the AQ Directives to corresponding k^{th} highest values are presented below

Calculation based on the combination of Pollutant & specific values:	***k*th highest values**	**AQD percentiles**	**Aggregation code**
…/aq/pollutant/1 …/aq/aggregationprocess/P1D	4th highest value (max4)	**99.18 percentile**	P1Y-day-max-per99.18
…/aq/pollutant/1 …/aq/primaryObservation/hour	25th highest value (max25)	**99.73 percentile**	P1Y-hr-max-per99.73
…/aq/pollutant/9 …/aq/primaryObservation/hour	19th highest value (max19)	**99.79 percentile**	P1Y-hr-max-per99.79
…/aq/pollutant/5 …/aq/primaryObservation/day or …/aq/aggregationprocess/P1D	AQ Directive	**90.4 percentile**	P1Y-P1D-per90.4
…/aq/pollutant/5 …/aq/primaryObservation/day or …/aq/aggregationprocess/P1D	36th highest value (max36)	**90.41 percentile**	P1Y-P1D-per90.41
…/aq/pollutant/5 or /6001 …/aq/primaryObservation/day or …/aq/aggregationprocess/P1D	WHO	**99 percentile**	P1Y-P1D-per99
…/aq/pollutant/7 …/aq/aggregationprocess/P8H-dmax	26ht highest value (max26)	**93.15 percentile**	P1Y-dmax-per93.15
…/aq/primaryObservation/hour or …/aq/primaryObservation/day	Median	**50 percentile**	P1Y–per50

6.2.2 Maxima

The (annual) maximum shall be calculated as follows:

$$Maximum = max\ (C_i)$$

Where C_i is the valid hourly/daily/daymax concentrations measured in the calendar year.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

The resultant aggregation should be encoded as:

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y-hr-max
- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y-day-max
- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y-dx-max

6.2.3 k^{th} highest value (maxk)

Calculation of the k^{th} highest value shall be based on ranked valid measurement values

$$X_1 \geq X_2 \geq X_3 \geq \ldots \geq X_k \geq \ldots \geq X_{N1} \geq X_N$$

The k^{th} highest value is the concentration X_k evaluated in descending order from the maximum (1st highest). The 4th, 25th, 19th, 26th and 36th highest values (kmax) are important to air quality regulations, see the array at the top of this section.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

Calculation based on the combination of Pollutant & specific values:	**k^{th} highest values**	**Aggregation code**
…/aq/pollutant/1 …/aq/aggregationprocess/P1D	4th highest value (max4)	P1Y-day-max4
…/aq/pollutant/1 …/aq/primaryObservation/hour	25th highest value (max25)	P1Y-hr-max25
…/aq/pollutant/9 …/aq/primaryObservation/hour	19th highest value (max19)	P1Y-hr-max19
…/aq/pollutant/5 …/aq/primaryObservation/day or …/aq/aggregationprocess/P1D	AQ Directive	P1Y-day-max36
…/aq/pollutant/7 …/aq/aggregationprocess/P8H-dmax	26ht highest value (max26)	P1Y-dx-max26

6.2.4 AOT40 (only O_3)

AOT40 is the cumulative concentration observed above 80 µg/m^3 (= 40 parts per billion) based on hourly measurements. There are currently 2 different AOT40 to be calculated:

- AOT40 crops (vegetation protection) – 1 May to 31 July – AOT40c
- AOT40 forest (forest protection) – 1 April to 30 September – AOT40f

In order to calculate AOT40, some base aggregation might be necessary. These are described below in more detail in section 6.2.4.1. Two statistics shall be calculated in order to obtain a final AOT40 for compliance:

- AOT40 measured
- AOT40 estimate (= final)

$$AOT40_{measured} = \Sigma_i \, max(0,(C_i - 80))$$

where C_i is the hourly mean ozone concentration in µg/m^3 and the summation is to the period 08.00 – 20.00 Central European Time each day in a certain time period[6]. For the AOT40 vegetation (AOT40c) the time period is the 3 month growing season crops from 1 May to 31 July each year. For the AOT40f, the time period is 1 April to 30 September.

AOT40 has a dimension of (µg/m^3)·hours and is sensitive to missing values. The required proportion of valid data is 90% of the one hour values over the time period defined for calculating the AOT40 value. Therefore, $AOT40_{measured}$ shall be routinely corrected to full time coverage to derive $AOT40_{estimate}$ as follows;

$$AOT40_{estimate} = (AOT40_{measured} \cdot N_{period}) / N_{valid}$$

where N_{valid} is the number of valid hourly values and N_{period} is the number of hours in the period.

6.2.4.1 Base aggregation for AOT40

In order to process AOT40, some internal base aggregation might be necessary which has been included in the codelist:

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/**AOT40h**

AOT40h: max (0, C_i – 80)

Calculate the difference between hourly value in µg/m^{-3} and 80 for hours with StartTime => 07:00 UTC and EndTime =< 19:00 and period YYYY-04-01T01:00:00+00:00 and YYYY-09-31T24:00:00.

[6] The AQD prescribes the use of CET as time reference in the calculation of the AOT40; this is in contrast to the calculations for all other aggregates and averages which are based on the time reference of the network i.e. the time references declared in the swe:DataArray of the data set E delivery..

In addition, the number of valid hourly values will need to be calculated. This will depend on whether the AOT40 is calculated for the protection of vegetation and the forest protection. The codelists are:

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/**AOT40c-nv**
- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/**AOT40f-nv**

AOT40c-nv

Number of valid hourly values (flag 1, 2, 3) between StartTime => 07:00 UTC and EndTime =< 19:00 AND YYYY-05-01T01:00:00+00:00 and YYYY-07-31T24:00:00+00:00.

AOT40f-nv

Number of valid hourly values (flag 1, 2, 3) between StartTime => 07:00 UTC and EndTime =< 19:00 AND YYYY-04-01T01:00:00+00:00 and YYYY-09-31T24:00:00+00:00).

6.2.4.2 AOT40 for vegetation protection

For the AOT40 vegetation (AOT40c) the time period is the 3 month growing season for crops from 1 May to 31 July each year.

AOT40cmeasured

SUM of all AOT40-h between YYYY-05-01T01:00:00+00:00 and YYYY-07-31T24:00:00+00:00.

AOT40c

AOT40c = AOT40c-measured * 1104 / AOT40c-nv

6.2.4.3 AOT40 for forest protection

For the AOT40 forest (AOT40f) the time period from 1 April to 30 September each year.

AOT40fmeasured

SUM of all AOT40-h between YYYY-04-01T01:00:00+00:00 and YYYY-09-30T24:00:00+00:00.

AOT40f

AOT40f = AOT40f-measured * 2196 / AOT40f-nv

6.2.4.4 AOT40 for vegetation protection averaged over 5 years

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/**AOT40c-P5y**

A year shall be included in the 5-year average if at least 86 % of data coverage achieved for the hours between 0800 and 2000 CET in this year (949 hours).

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

6.2.5 SOMO35 (only O_3)

For quantification of the health impacts the World Health Organisation recommends the use of the SOMO35 indicator. SOMO35 stands for means the sum of the differences between maximum daily 8-hour running mean concentrations greater than 70 µg/m^3 (= 35 parts per billion) and 70 µg/m^3.

- http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/**SOMO35**

$$SOMO35_{measured} = \Sigma_i \, max(0, (C_i - 70))$$

where C_i is the maximum daily 8-hour running mean ozone concentration in µg/m^3 and the summation is over all days per calendar year.

SOMO35 has a dimension of (µg/m^3)·days. SOMO35 is sensitive to missing values and a correction to full time coverage has been applied:

$$SOMO35_{estimate} = (SOMO35_{measured} \cdot N_{period}) / N_{valid}$$

where N_{valid} is the number of valid daily values and N_{period} is the number of days per year.

6.2.6 AEI (only $PM_{2.5}$) (Average Exposure Indicator)

Part A of Annex XIV of Directive 2008/50/EC defines that the AEI shall be assessed as a three-calendar year running annual mean concentration averaged over all sampling points established pursuant to Section B of Annex V of the Directive.

KEY NOTE FOR EEA DEVELOPERS: AEI statistics are calculated at National level. For this reason, the calculations at SamplingPoint level for PM2.5 will go as far as calculating the Annual average (P1Y) and the corresponding data capture.

To ensure transparent and unambiguous calculation for the reported AEI, the following shall be observed:

- calculate the annual average PM2.5 concentration at each AEI monitoring station, applying the criteria laid down in Annex XI.A of Dir. 2008/50/EC for each year
- calculate the average over all AEI monitoring stations within the MS for each year
- calculate the average over three years (the actual reporting year, and the two years before).

These aggregation steps have to be applied on data with numeric accuracy available in the monitoring network. Rounding has to be done at the end of the aggregation procedure.

Data on stations and measurement configurations are considered to be stable. Any modification of the set of AEI monitoring stations is strongly discouraged. The selection of the sampling points has to be documented in dataset D.

The AEI shall be reported annually as a three year running mean (covering the actual reporting year, and the two years before).

In the case when the data capture of ≥ 90% was not achieved then it is recommended to follow the procedure developed by AQUILA. [see AQUILA_AEI guidance as a separate document].

6.2.7 National Exposure Reduction Target (NERT)

It is acknowledged that the table given in Annex XIV, B of Directive 2008/50/EC is not consistent as regards the numerical accuracy of the initial concentration threshold for the AEI. It is recommended that the initial AEI concentration threshold should be presented to one decimal place i.e. 8.5, 13.0, 18.0 and 22.0 µg/m3. In order to be consistent, also the numerical accuracy of the AEI obtained in practice in a MS should be rounded to 1 decimal place.

Example: the average exposure indicator of one MS is equal to 16.4 µg/m3. In order to establish the reduction target, this value has to be compared with the initial concentration threshold of 18.0 µg/m3.

6.3 Calculation of basic exceedance information

Some basic exceedance information is required to support the assessment of attainment environmental objectives and evaluation of UTD data flows.

To determine exceedance situations, concentrations are compared with a threshold (known variously as limit values, target values, thresholds and margins above thresholds in the Directives). In this comparison numerical rounding shall be the last step of any calculation, i.e. immediately before comparing the result with any threshold of any type and shall be done only once. For information on the how validity and verification flags shall be applied as part of exceedance calculation see section 2.1.1. In general, the verification status (flag) for validated assessment datasets (Data flows E1a – measurements and E1b - modelling) can only be 3 (verified). The verification status for UTD datasets (Data flows E2a) may be of any type.

The rounding off rules shall be based on the principle presented in the table below;

Value x	**Number of decimals**
$x \geq 10$	integer
$1 \leq x < 10$	1 decimal
$0.1 \leq x < 1$	2 decimals
$0.01 \leq x < 0.1$	3 decimals
Etc…	

Data Example 1

PM_{10} daily value of 50.49 µg/m^3 would be rounded to 50 µg/m^3 (when comparing this value to the daily limit value of 50 µg/m^3 the result would be a no exceedance on that day.

Data Example 2

PM_{10} daily value of 50.5 µg/m^3 would be rounded to 51 µg/m^3 (when comparing this value to the daily limit value of 50 µg/m^3 the result would be an exceedance.

Data Example 3

Ozone (O_3) hourly value of 180.49 µg/m^3 would be rounded to 180 µg/m^3 (when comparing this value to the hourly information threshold of 180 µg/m^3 the result would be a no exceedance of the information threshold).

Data Example 4

Ozone (O_3) hourly value of 180.50 µg/m^3 would be rounded to 181 µg/m^3 (when comparing this value to the hourly information threshold of 180 µg/m^3 the result would be an exceedance of the information threshold).

Data Example 5

Ozone (O_3) hourly value of 180.90 µg/m^3 would be rounded to 181 µg/m^3 (when comparing this value to the hourly information threshold of 180 µg/m^3 the result would be an exceedance of the information threshold).

Data Example 6

Benzene annual mean of 4.36 µg/m^3 would be rounded to 4.4 µg/m^3 (when comparing this value to the calendar year limit value of 5 µg/m^3 the result would return no exceedance of the limit value).

Data Example 7

Benzene annual mean of 5.12 µg/m^3 would be rounded to 5.1 µg/m^3 (when comparing this value to the calendar year limit value of 5 µg/m^3 the result would return an exceedance of the limit value).

6.3.1 Exceedance of thresholds based on calendar year, winter mean and maximum daily eight hour mean aggregated statistics

An exceedance of the calendar year (annual mean), winter mean (1 October to 31 March) and maximum daily eight hour mean environmental thresholds when a rounded average from a time series of valid measurement values is > y µg/m^3 (with y = limit, threshold or target value).

6.3.2 Number of hours or days with concentration > y µg/m^3

The number of hours or days with rounded concentration > y µg/m^3 (with y = limit or threshold value) shall be calculated from the valid measurement values:

$$Z_1 , Z_2 , Z_3 , \dots , Z_k , \dots , Z_{N-1} , Z_N$$

N is the number of Z_k-values for which $Z_k > y$ µg/m^3 where Z_k is the concentration rounded according to the rules given before.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

6.3.3 Three consecutive hours with concentration > y µg/m^3

Three consecutive hours are in exceedance when a consecutive measurement triple (Z_{k-1} , Z_k , Z_{k+1}) from a time series of valid measurement values is > y µg/m^3 (with y = limit or threshold value).

$$Z_1 , Z_2 , Z_3 , \dots , Z_k , \dots , Z_{N-1} , Z_N$$

N is the number of triples in exceedance.

Scripts written in C: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/c-sources

Awk scripts: http://acm.eionet.europa.eu/schemas/airbase/calc_stats_aggregs/AWK-scripts-aggregation-and-statistical-routines_EEA-checks-ver20120131.zip

6.4 XML encoding of aggregated statistics

Aggregated statistics generated using the rules presented above will be stored and made available via an aggregation and download service for all measurement data. The results of aggregations will form part of Data flow F - Information on generated aggregated data and will be encoded in XML.

Figure 4 presents a proposed example XML encoding of aggregated statics (Data flow F). The example is configured similarly to the Data flow E deliveries presented in section 2, using a combination of GML, O&M and SWE data standards all of which are imported by the AQD.xsd. In contrast to the Dataflow E deliveries which utilised an swe:DataBlock to encode measurements over the observational period, Dataflow F will use a simple swe:DataRecord element encoding for the aggregated statistics derived from input measurements. The input datasets (Data flow E) used to derive aggregated data (Data flow F) may be cited via xlink href within the data flow F within the om:relatedObservation element.

Figure 4 Example XML instance file of the proposed XML encoding of Data flow F –aggregated statistics

```
<gml:featureMember>
        <om:OM_Observation gml:id="AggregatedObservation_1">
                <om:phenomenonTime>
                        <gml:TimePeriod gml:id="ObserrvationTimePeriod_3">
                                <gml:beginPosition>2010-01-01T01:00:00+01:00</gml:beginPosition>
                                <gml:endPosition>2010-12-31T24:00:00+01:00</gml:endPosition>
                        </gml:TimePeriod>
                </om:phenomenonTime>
                <om:resultTime>
                        <gml:TimeInstant gml:id="ObservationResultInstant_3">
                                <gml:timePosition>2011-07-29T14:00:00+01:00</gml:timePosition>
                        </gml:TimeInstant>
                </om:resultTime>
                <om:procedure xlink:href="http://environment.data.gov.uk/air-
quality/so/GB_SamplingpointProcess_2774"/>
                <om:parameter>
                        <om:NamedValue>
                                <om:name xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/processparameter/EO”/>
                                <om:value>
                                        <aqd:environmentalObjective>
                                                <aqd:EnvironmentalObjective>
                                                        <aqd:objectiveType
xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/objectivetype/LV”/>
                                                        <aqd:reportingMetric
xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/reportingmetric/hrsAbove”/>
                                                        <aqd:protectionTarget
xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/protectiontarget/H”/>
                                                </aqd:EnvironmentalObjective>
                                        </aqd:environmentalObjective>
                                </om:value>
                        </om:NamedValue>
                </om:parameter>
                <om:parameter>
                        <om:NamedValue>
                                <om:name xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/processparameter/AT”/>
                                <om:value>"Fixed measurement"</om:value>
                        </om:NamedValue>
                </om:parameter>
                <om:observedProperty xlink:href="http://dd.eionet.europa.eu/vocabulary/aq/pollutant/8”/>
                <om:featureOfInterest xlink:href="http://data.umweltbundesamt.at/inspire/AT.UBA.AQD/SamplingFeature_1"/>
```

```
		<om:resultQuality>
			<gmd:DQ_DomainConsistency>
				<gmd:result>
					<gmd:DQ_ConformanceResult>
						<gmd:specification>
							<gmd:CI_Citation>
								<gmd:title>

	<gco:CharacterString>EC/50/2008</gco:CharacterString>
								</gmd:title>
								<gmd:date>
									<gmd:CI_Date>
										<gmd:date>

	<gco:Date>2008</gco:Date>
										</gmd:date>
										<gmd:dateType>
											<gmd:CI_DateTypeCode
codeListValue="publication" codeList="eng">publication</gmd:CI_DateTypeCode>
										</gmd:dateType>
									</gmd:CI_Date>
								</gmd:date>
							</gmd:CI_Citation>
						</gmd:specification>
						<gmd:explanation>
							<gco:CharacterString>Time Coverage</gco:CharacterString>
						</gmd:explanation>
						<gmd:pass>
							<gco:Boolean>true</gco:Boolean>
						</gmd:pass>
					</gmd:DQ_ConformanceResult>
				</gmd:result>
			</gmd:DQ_DomainConsistency>
		</om:resultQuality>
		<om:resultQuality>
			<gmd:DQ_DomainConsistency>
				<gmd:result>
					<gmd:DQ_ConformanceResult>
						<gmd:specification>
							<gmd:CI_Citation>
								<gmd:title>

	<gco:CharacterString>EC/50/2008</gco:CharacterString>
								</gmd:title>
								<gmd:date>
									<gmd:CI_Date>
										<gmd:date>

	<gco:Date>2008</gco:Date>
										</gmd:date>
										<gmd:dateType>
											<gmd:CI_DateTypeCode
codeListValue="publication" codeList="eng">publication</gmd:CI_DateTypeCode>
										</gmd:dateType>
									</gmd:CI_Date>
								</gmd:date>
							</gmd:CI_Citation>
						</gmd:specification>
						<gmd:explanation>
							<gco:CharacterString>Data Capture</gco:CharacterString>
						</gmd:explanation>
						<gmd:pass>
							<gco:Boolean>true</gco:Boolean>
						</gmd:pass>
					</gmd:DQ_ConformanceResult>
				</gmd:result>
			</gmd:DQ_DomainConsistency>
		</om:resultQuality>

		<om:resultQuality>
			<gmd:DQ_DomainConsistency>
				<gmd:result>
					<gmd:DQ_QuantitativeResult>
					<!--	<gmd:valueUnit
xlink:href="http://dd.eionet.europa.eu/vocabulary/uom/statistics/percentage"/> -->
```

```
                                                <gmd:valueUnit>
                                                        <gml:BaseUnit gml:id="PercentageUnit_3">
                                                                <gml:identifier
codeSpace="http://dd.eionet.europa.eu/vocabularies/aq/resultquality/uncertaintyestimation/"/>
                                                                <gml:catalogSymbol codeSpace="
http://dd.eionet.europa.eu/vocabulary/uom/statistics/">http://dd.eionet.europa.eu/vocabulary/uom/statistics/percentage</gml:cat
alogSymbol>
                                                                <gml:unitsSystem
xlink:href="http://www.opengis.net/def/uom/UCUM/"/>
                                                        </gml:BaseUnit>
                                                </gmd:valueUnit>
                                                <gmd:value>
                                                        <gco:Record>
                                                                5
                                                        </gco:Record>
                                                </gmd:value>
                                        </gmd:DQ_QuantitativeResult>
                                </gmd:result>
                        </gmd:DQ_DomainConsistency>
                </om:resultQuality>
                <om:result xsi:type="swe:DataArrayType">
                        <swe:elementCount>
                                <swe:Count>
                                        <swe:value>1</swe:value>
                                </swe:Count>
                        </swe:elementCount>
                        <swe:elementType name=" FixedAggregatedObservations">
                                <swe:DataRecord>
                                        <swe:field name="StartTime">
                                                <swe:Time
definition="http://www.opengis.net/def/property/OGC/0/SamplingTime">
                                                        <swe:uom xlink:href="http://www.opengis.net/def/uom/ISO-
8601/0/Gregorian"/>
                                                </swe:Time>
                                        </swe:field>
                                        <swe:field name="EndTime">
                                                <swe:Time
definition="http://www.opengis.net/def/property/OGC/0/SamplingTime">
                                                        <swe:uom xlink:href="http://www.opengis.net/def/uom/ISO-
8601/0/Gregorian"/>
                                                </swe:Time>
                                        </swe:field>
                                        <swe:field name="Verification">
                                                <swe:Category
definition="http://dd.eionet.europa.eu/vocabularies/aq/observationverification”/>
                                        </swe:field>
                                        <swe:field name="Validity">
                                                <swe:Category
definition="http://dd.eionet.europa.eu/vocabularies/aq/observationvalidity”/>
                                        </swe:field>
                                        <swe:field name="Value">
                                                <swe:Quantity definition="
http://dd.eionet.europa.eu/vocabulary/aq/aggregationprocess/P1Y-hrsAbove200">
                                                        <swe:uom code="
http://dd.eionet.europa.eu/vocabulary/uom/statistics/count"/>
                                                </swe:Quantity>
                                        </swe:field>
                                </swe:DataRecord>
                        </swe:elementType>
                        <swe:values>2010-01-01T01:00:00+01:00,2010-12-31T24:00:00+01:00,1,1,20@@</swe:values>
                </om:result>
        </om:OM_Observation>
</gml:featureMember>
```

7 Up-To-Date and Summer Ozone Reporting calculations

7.1 Data sources

For Up-To-Date (UTD) data, countries will provide the primary data in a timely manner in order to submit the data as early as possible. Ideally data will be provided on an hourly basis as in the current operational Near Real Time (NRT) system at the EEA.

The UTD primary data will be reported with a timestamp and validity and a verification status flag as previously discussed and presented in the example XML in Figure 1.

Section 6, describes all the statistics to be carried out for all primary data. This section describes all specific statistics to be carried out on UTD only data. **All the statistics in Section 6 applicable to hourly data shall be carried out using UTD data**

7.2 Rounding off rules

See section 6.3.

7.3 Aggregations

See the 75% rule in section 3.

7.4 Aggregating valid data

For any aggregation, only those values with a validity flag of "Valid" (1), "Valid below detection limit and the measurement value is given" (2) or "Valid below detection limit & number replaced by 0.5*detection limit" (3) are to be used. For UTD data, most valid data will be flag as 1.

All data, regardless of their validity status, should be stored in the UTD database. This will allow EEA, ETC/ACM and the data provider to assess the data capture. This will allow the system to distinguish between data of all validity data types. Provisionally we propose another internally generated flag which will be used to identify parts of UTD time series not delivered by data provider as 0. This is analogous with validity flag used in EOI reporting which when a complete time series is delivered (for all hours in a year) uses a validity flag=0 to indicate no measurement is available for the time period specified. Likewise under the same system, if there is a gap in the time series the missing time series is flagged with validity flag= 0.

7.5 Verification of data.

See section 2.1.1

7.6 Percentage of valid data

Having taken into account the 75% criteria of minimum required see section 3 to calculate 8 hourly means and daily maximum 8 hourly mean, it is important that for any given aggregation, the system also calculates the percentage of valid data that the aggregation has used. In particular, the following are important:

a) % of valid data including 1 to 3

b) % of not valid due to station maintenance or calibration (-99)

7.7 Calculation statistics on UTD data

The table below gives the Environmental Objectives and Reporting Metrics covered for ozone (O_3), nitrogen dioxide (NO_2) and sulphur dioxide (SO_2) used in UTD

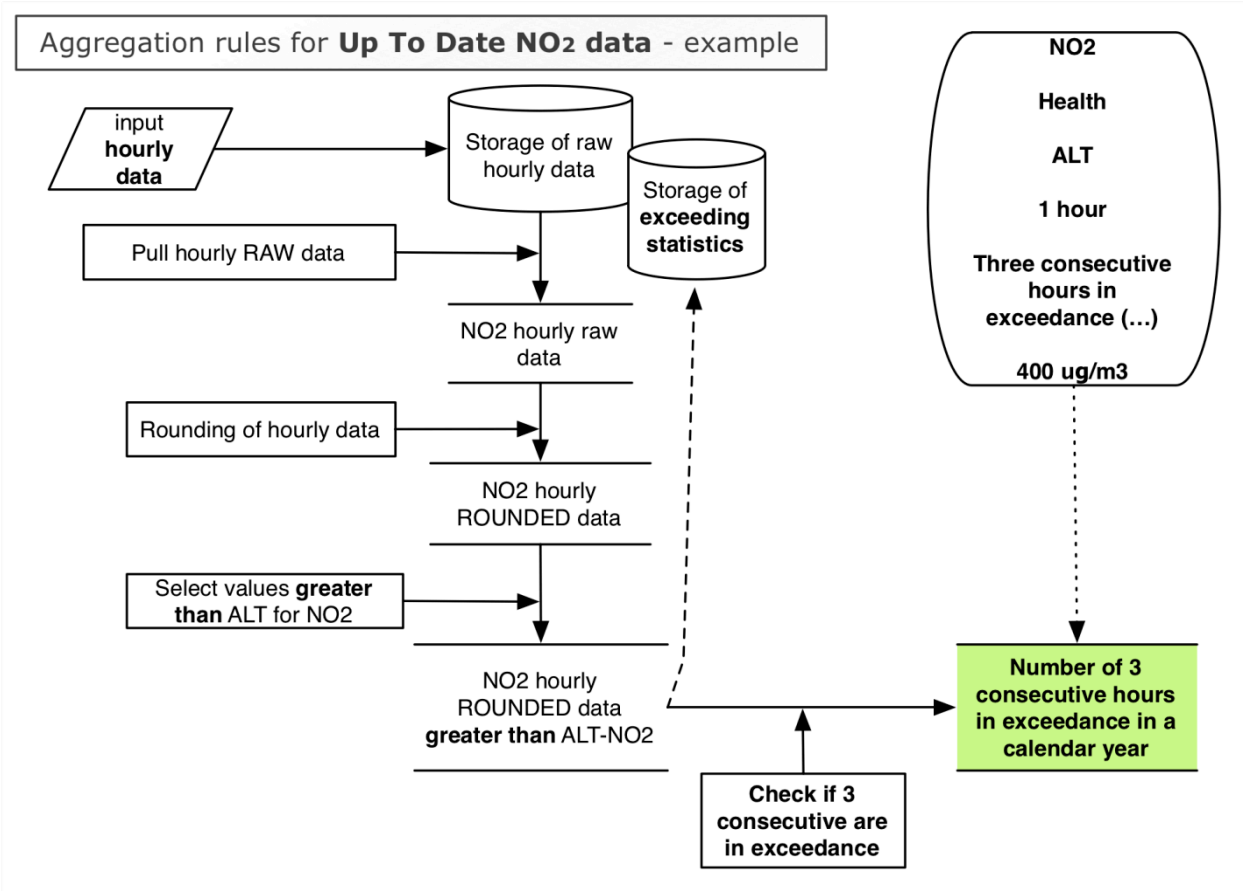
Table 5 Environmental Objectives and Reporting Metrics relevant to UTD

Formula	Protection target	Environ -mental Objective type (Code (1))	Averaging period of assessments	Reporting Metric of environmental objective	Numerical values of the environmental objective (allowed number of exceedances)
NO_2	Health	ALT	1 hour	Three consecutive hours in exceedance (at locations representative of air quality over at least 100 km 2 or an entire zone or agglomeration, whichever is smaller*)	400 µg/m^3
SO_2	Health	ALT	1 hour	Three consecutive hours in exceedance (at locations representative of air quality over at least 100 km 2 or an entire zone or agglomeration, whichever is smaller*)	500 µg/m^3
O_3	Health	LTO	Maximum daily 8-hour mean	Days when maximum daily eight- hour mean exceeded the long term objective in 1 calendar year	120 µg/m^3
		INT	One hour	Hours in exceedance in a calendar year	180 µg/m^3
		ALT	One hour	Hours in exceedance in a calendar year	240 µg/m^3

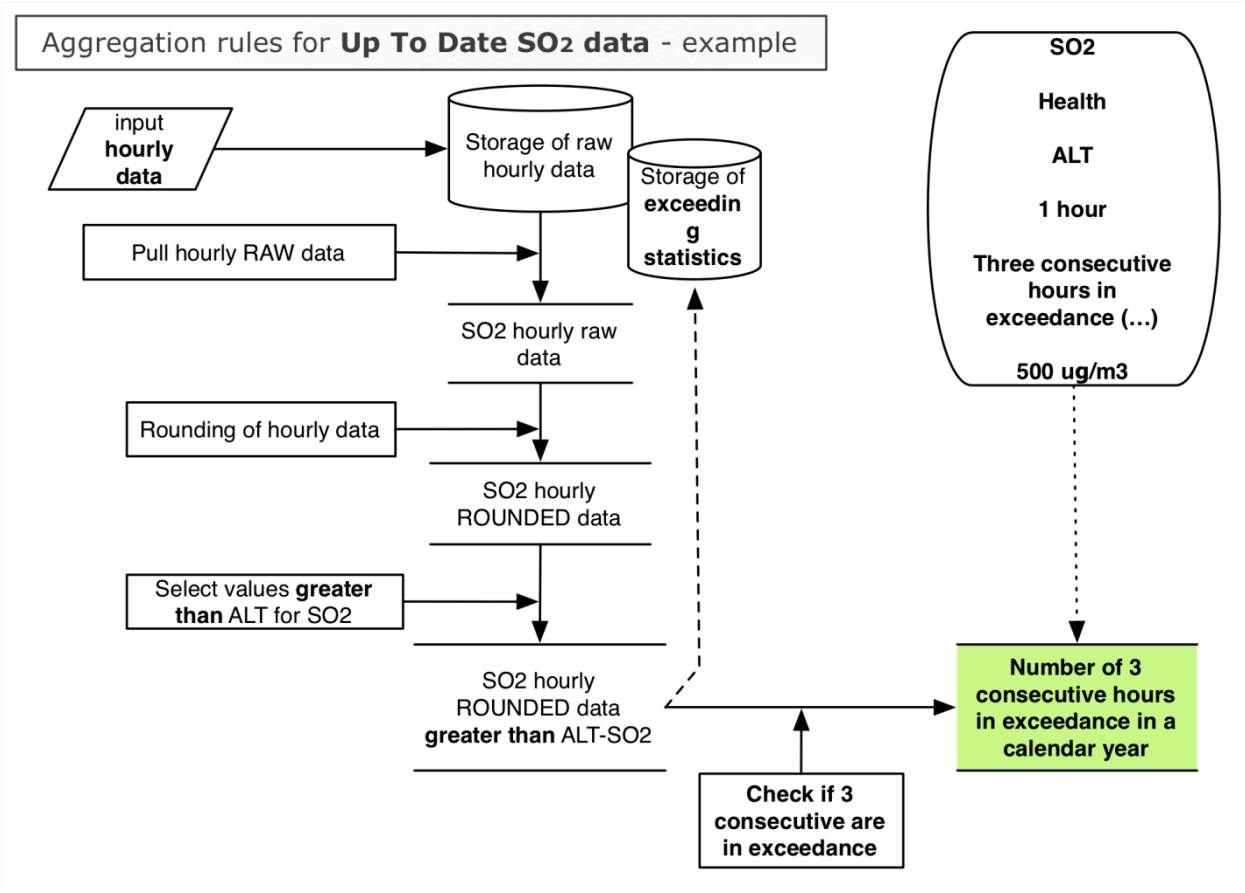
***Please note that for the NO_2 and SO_2 Health Alert threshold of three consecutive hours in exceedances is always calculated due to currently missing information on the 100 km^2 criteria**

The following sections present data flow diagrams which describe the aggregation, statistical calculations, rounding at storage requirements for different UTD metrics and pollutants.

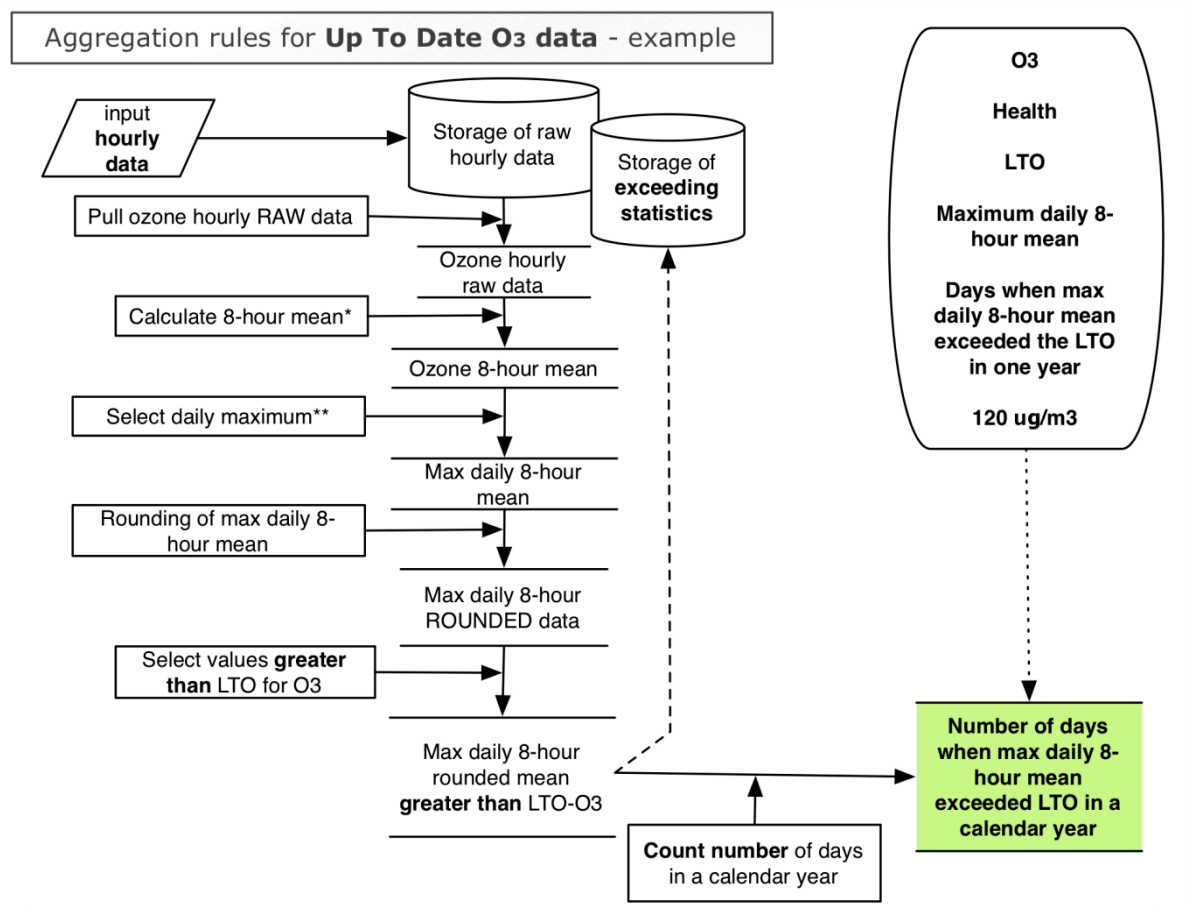
7.7.1 ALT – NO2: Number of times when 3 consecutive hours have been above ALT environmental Objective of 400 µg/m³



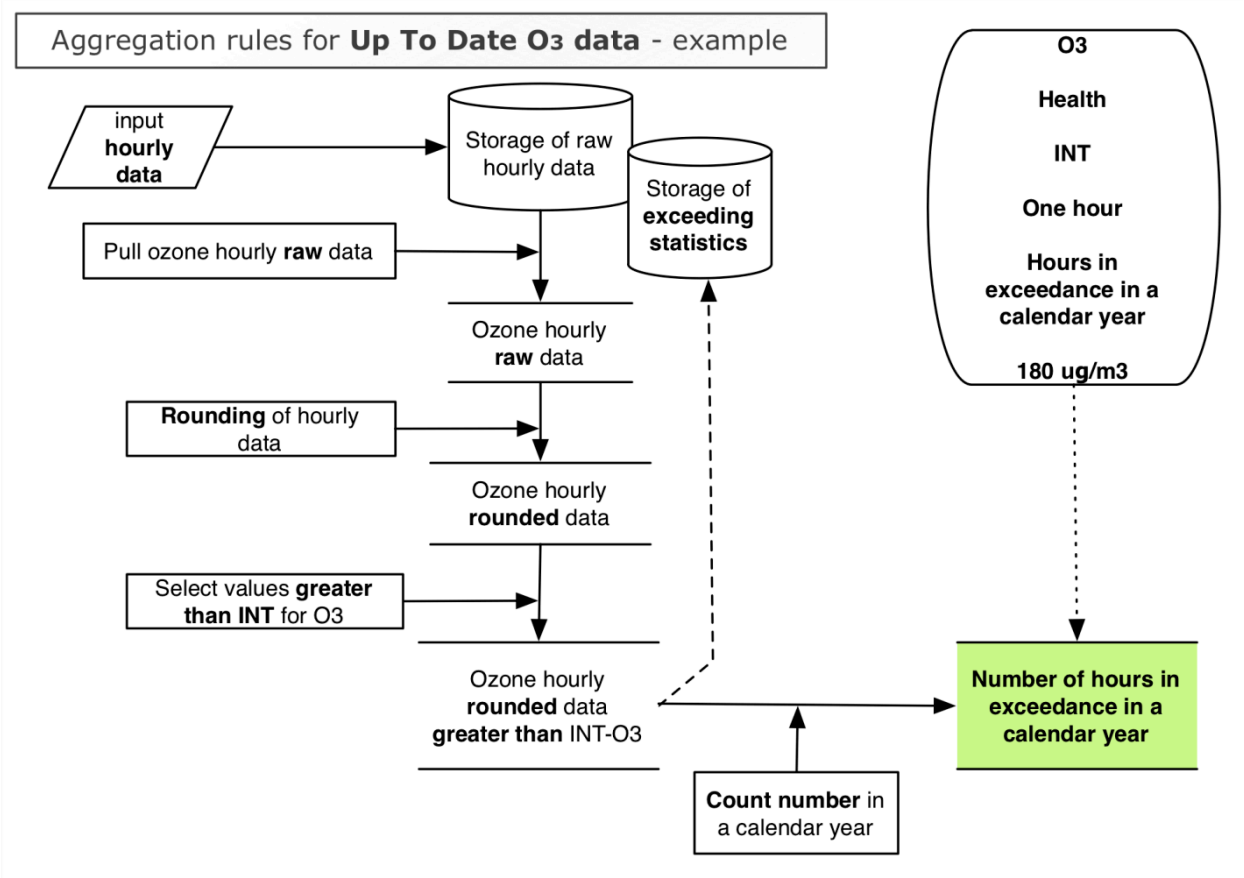
7.7.2 ALT – SO2: Number of times when 3 consecutive hours have been above ALT environmental Objective of 500 µg/m³



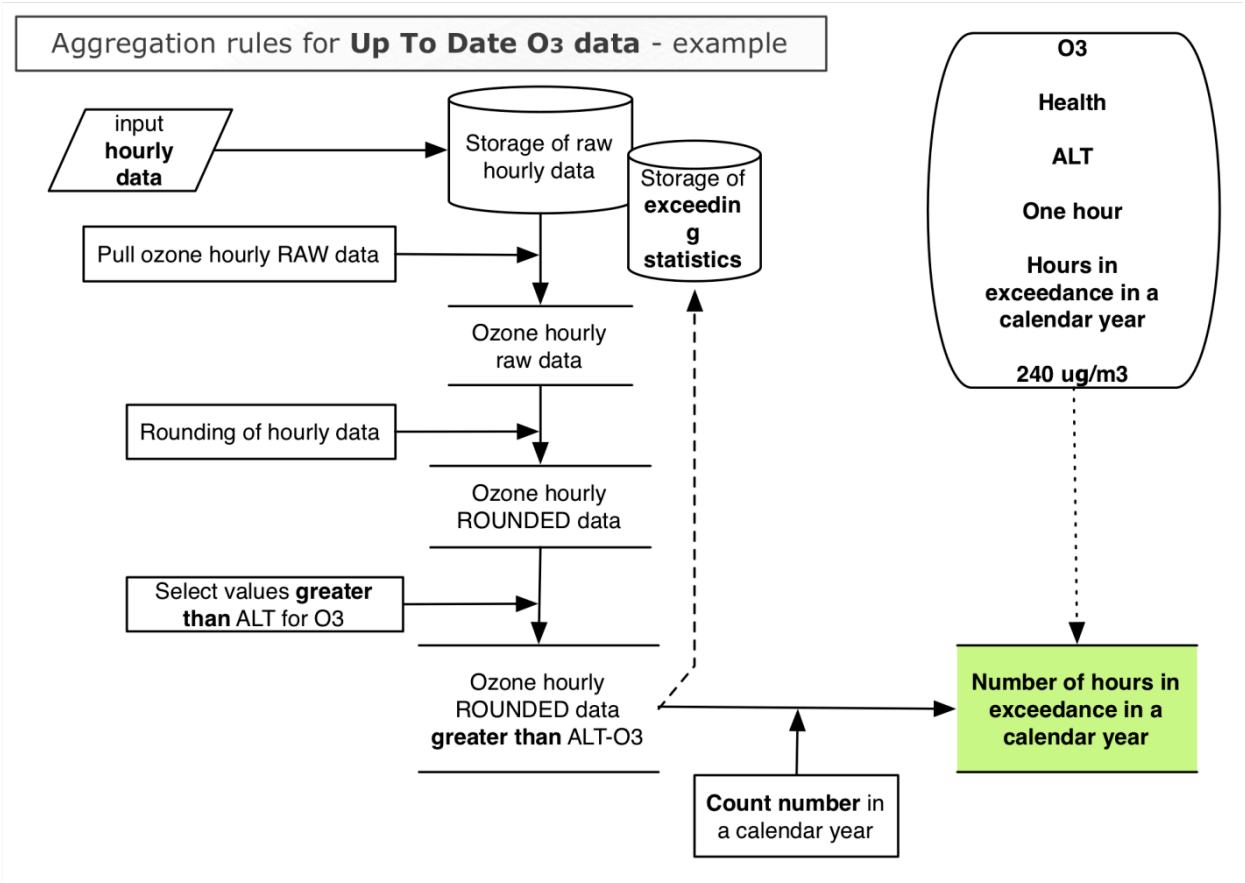
7.7.3 LTO –Max. daily 8-hour mean - Days when max daily 8-hour mean exceeded the LTO in one year - 120 µg/m³



7.7.4 O3 - Health - INT - One hour - Hours in exceedance in a calendar year -180 µg/m³



7.7.5 O3 -> Health -> ALT -> One hour -> Hours in exceedance in a calendar year ->240 µg/m³



7.8 Additional aggregation on Summer Ozone Reporting and output report

In addition to the environmental objectives calculation required above which are required throughout the year the Summer Ozone Reporting (SOR) are required for the periods between April and September http://www.eea.europa.eu/maps/ozone/compare/explorer. SOR requires the following metric to be calculated and evaluated separately over this period

- Health LTO for ozone,
- Health INT for ozone and
- Health ALT for ozone.

The aggregation for SOR will need to be grouped in order to produce the report similar to http://www.eea.europa.eu/maps/ozone/compare/explorer. The data is currently grouped in the as follows:

1. Country per month
2. Country per summer season

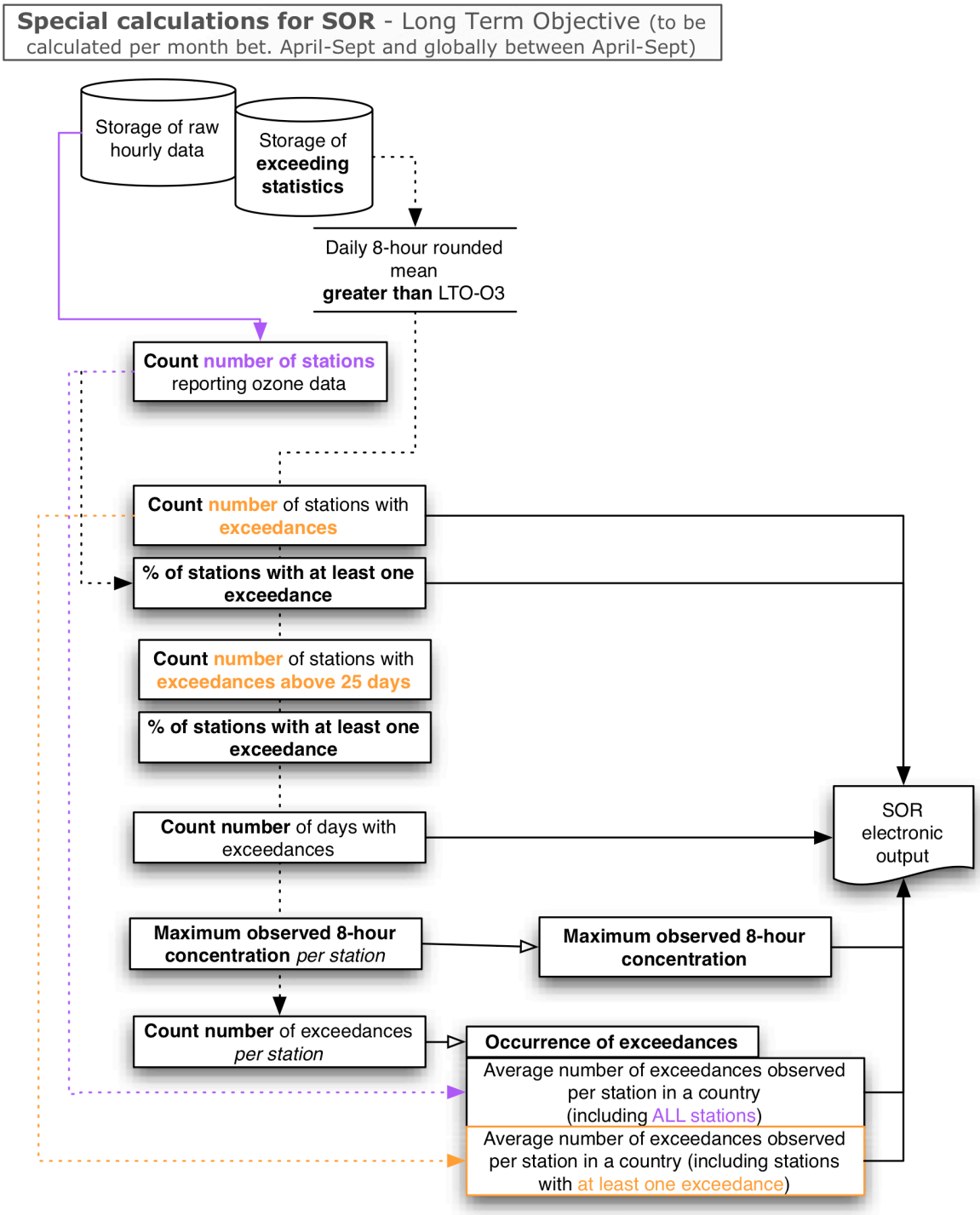
3. Europe per month
4. Europe per summer season

In the future, the grouping might be carried out by zone declared via dataset B of the IPR.

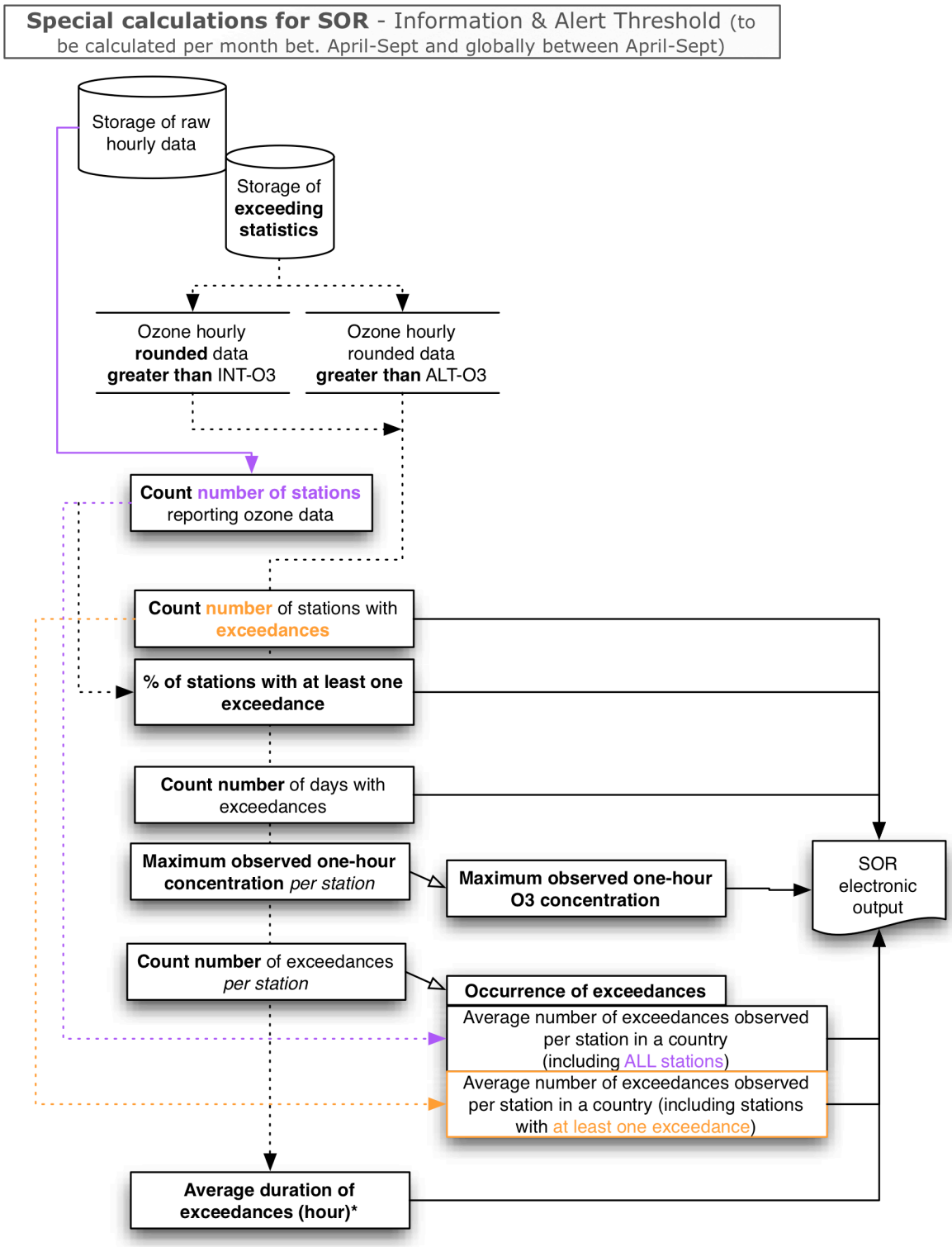
The following diagrams reflect the specific outputs necessary to calculate the SOR aggregation statistics.

7.8.1 LTO –Max. daily 8-hour mean –

I



7.8.2 INT & ALT – hourly mean



8 References

EU (2004): Directive 2004/107/EC of the European Parliament and the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air (fourth Daughter Directive (2004/107/EC)). Official Journal L 23, 26/01/2005, pp. 3–16. See http://ec.europa.eu/environment/air/ambient.htm

EU (2008): Directive 2008/50/EC of the European Parliament and the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Official Journal L 152, 11/06/2008, pp. 1–44. See http://ec.europa.eu/environment/air/quality/legislation/existing_leg.htm

EU (2011): Directive 2011/850/EC of the European Parliament and the Council of 17 December 2011 laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality. Official Journal L 335, 17/12/2011, pp. 86–106. See http://ec.europa.eu/environment/air/quality/legislation/existing_leg.htm

EU (2012): Guidance on the Implementing Provisions for the Directives 2004/107/EC and 2008/50/EC. In development in line with the Implementing Provisions.

Annex A Environmental objectives and reporting metrics

LV: limit value, LVMT: Limit value plus margin of tolerance, TV: target value, LTO: long-term objective, INT: Information threshold, ALT: Alert threshold, CL: Critical level, NAT: Assessment of natural contribution, WSS: Assessment of winter sanding and salting, ERT: Exposure reduction target, ECO: Exposure concentration obligation; * indicates: at locations representative of air quality over at least 100 km^2 or an entire zone or agglomeration, whichever is the smaller

Formula	Protection target	Env. Objective type	Reporting Metric	Objective Value	Averaging period	Alternative Reporting Metric	Target date
NO_2	Health	LV	Hours in exceedance in a calendar year	200 µg/m^3 (18)	calendar year	max19 (99.79 percentile)	1.1.2010
		LVMT					
		LV	Annual average	40 µg/m^3	calendar year		1.1.2010
		LVMT					
		ALT	(Number of) three consecutive hours in exceedance *	400 µg/m^3			
NO_x	Vegetation	CL	Annual average	30 µg/m^3	calendar year		
PM_{10}	Health	LV	Days in exceedance in a calendar year	50 µg/m^3 (35)	calendar year	max36 (90.41 percentile)	
		LV	Annual average	40 µg/m^3	calendar year		
		WSS	Deducted days in exceedance in a calendar year	50 µg/m^3	calendar year		
			Deduction of annual average	40 µg/m^3	calendar year		
		NAT	Deducted days in exceedance in a calendar year	50 µg/m^3	calendar year		
			Deduction of the annual average	40 µg/m^3	calendar year		
$PM_{2.5}$	Health	ECO	Average Exposure Indicator: (calculation see Directive 2008/50/EC)	20 µg/m^3	3 subsequent calendar years		1.1.2015
		ERT	Exposure reduction target	Percentage reduction			1.1.2020

Formula	Protection target	Env. Objective type	Reporting Metric	Objective Value	Averaging period	Alternative Reporting Metric	Target date
		TV	Annual average	25 µg/m^3	calendar year		1.1.2010
		LV		25 µg/m^3 (20 µg/m3)			1.1.2015 (1.1.2020)
		LVMT		27 (2012); 26 (2013,2014); 25 (2015)			
SO_2	Health	LV	Hours in exceedance in a calendar year	350 µg/m^3 (24)	calendar year	max25 (99.73 percentile)	1.1.2005
			Days in exceedance in a calendar year	125 µg/m^3 (3)	calendar year	max4 (99.18 percentile)	1.1.2005
		ALT	(Number of) three consecutive hours in exceedance *	500 µg/m^3			
		NAT	Deducted hours in exceedance in a calendar year	350 µg/m^3	calendar year		
			Deducted days in exceedance in a calendar year	125 µg/m^3	calendar year		
	Vegetation	CL	Annual average	20 µg/m^3	calendar year		
			Average value over the winter months (winter average), i.e. 1 October year x-1 to 31March year x	20 µg/m^3	Winter mean		
O3	Health	TV	Days when maximum daily 8-hour mean exceeded the target value (averaged over three years)	120 µg/m3 (25)	3 years	max26 (93.15 percentile)	1.1.2010
		LTO	(number of) days when maximum daily 8-hour mean exceeded the long term objective in one calendar year	120 µg/m3 (1)	calendar year		
		INT	(Number of) hours in exceedance in a calendar year	180 µg/m3			
		ALT	(Number of) hours in exceedance in a calendar year	240 µg/m3			

Formula	Protection target	Env. Objective type	Reporting Metric	Objective Value	Averaging period	Alternative Reporting Metric	Target date
	Vegetation	TV	AOT40 (calculation see Dir. 2008/50/EC Annex VII) (averaged over five years)	180000 (µg/m3).h	May until July; averaged over 5 years		1.1.2010
		LTO	AOT40 (calculation see Dir. 2008/50/EC Annex VII) (calculation over 1 year)	6000 (µg/m3).h	May until July		
CO	Health	LV	(Number of) days when maximum daily 8-hour mean exceeded the limit value	10 mg/m3 (1)	calendar year		1.1.2005
Benzene	Health	LV	Annual average	5 µg/m3	calendar year		1.1.2010
Lead	Health	LV	Annual average	0.5 µg/m3			1.1.2005
Cadmium	Health	TV	Annual average	5 ng/m3			
Arsenic	Health	TV	Annual average	6 ng/m3			
Nickel	Health	TV	Annual average	20 ng/m3			
B(a)P	Health	TV	Annual average	1 ng/m3			

Annex B Margin of tolerance and upper/lower assessment thresholds

Formula	Protection target	Env. Objective type	Reporting Metric	Objective Value	MOT	LAT	UAT
NO_2	Health	LV	Hours in exceedance in a calendar year	200 µg/m3		100 µg/m3	140 µg/m3
		LVMT					
		LV	Annual average	40 µg/m3		26 µg/m3	32 µg/m3
		LVMT					
NO_x	Vegetation	CL	Annual average	30 µg/m3		19.5 µg/m3	24 µg/m3
PM_{10}	Health	LV	Days in exceedance in a calendar year	50 µg/m3		25 µg/m3	35 µg/m3
		LV	Annual average	40 µg/m3		20 µg/m3	28 µg/m3
		TV	Annual average	25 µg/m3			
		LV		25 µg/m3 (20 µg/m3)	2 (2012); 1 (2013,2014); 0 (2015)	12 µg/m3	17 µg/m3
		LVMT		27 (2012); 26 (2013,2014); 25 (2015)			
SO_2	Health	LV	Hours in exceedance in a calendar year	350 µg/m^3			
			Days in exceedance in a calendar year	125 µg/m^3		50 µg/m3	75 µg/m3
	Vegetation	CL	Annual average	20 µg/m^3			
			Average value over the winter months (winter average), i.e. 1 October year x-1 to 31March year x	20 µg/m^3		8 µg/m3	12 µg/m3
CO	Health	LV	(Number of) days when maximum daily 8-hour mean exceeded the limit value (annual mean?)	10 mg/m^3		.5*10 = 5 µg/m3	.7*10 = 7 µg/m3
Benzene	Health	LV	Annual average	5 µg/m^3		2 µg/m3	3.5 µg/m3
Lead	Health	LV	Annual average	0.5 µg/m^3		.25 µg/m3	0.35 µg/m3
Cadmium	Health	TV	Annual average	5 ng/m^3		2 ng/m^3	3 ng/m^3
Arsenic	Health	TV	Annual average	6 ng/m^3		2.4 ng/m^3	3.6 ng/m^3
Nickel	Health	TV	Annual average	20 ng/m^3		10 ng/m^3	14 ng/m^3
B(a)P	Health	TV	Annual average	1 ng/m^3		0.4 ng/m^3	0.6 ng/m^3

Annex C Definition of the averaging time “var”

The averaging time is the period of the sample (end date/time minus start date/time). If the sample periods of a component differ 25% or more from a constant averaging time, the averaging time has been defined as “var”. Example: if all periods of 4-week samples are within 21 and 35 days, the averaging time is still 4-week. The 100% period for an n-month sample has been defined as the period starting from the start date/time of the sample and ending on the same day number and time n months later. Example: the sample starts at 5 March at 00:00, the 100% 1-month period is until 5 April at 00:00. Other example: the sample starts at 30 January at 00:00, the 100% 1-month period is until “virtual” 30 February, that is actually 2 March at 00:00 (no leap year). So if the end date/time is between 27 March 18:00 and 22 April 18:00 the sample period has still 1month averaging time.

Annex D Historical methods for calculation of data capture and time coverage from multi day sampler based data types integrated over an averaging period

Because no data capture and time coverage per sample/observation are delivered to the existing AirBase systems, it is assumed that the time coverage per sample/observation is 100%. If this sample/observation has a validity flag >0, we also assume that the validity flag of all component values in the sample/observation also have validity flag >0.

For the calculation of the data capture and time coverage we count the hours in the sampling period (end time sample – start time sample).

The **data capture** DC in a given averaging period (e.g. a year) is calculated as follows

$$DC = \Sigma_i N_{valid,i} / \Sigma_i N_i \ * 100\ \%$$

Where

$N_{valid,i}$ = the number of hours in the valid sample i (validity flag sample > 0)
N_i = number of hours in sample i

The **time coverage** TC is the percentage of measurement time in a given averaging period. Also the time coverage is calculated on base of the number of hours in the sample periods. The time coverage for a given averaging period is calculated as follows:

$$TC = \Sigma_i N_i / N_{averaging_period} * 100\ \%$$

Where

N_i = number of hours in sample i
$N_{averaging_period}$ = the total number of hours in the given averaging period (e.g. a year)

The **mean** for a given period is like the data capture calculated on base of the number of hours in the valid sample periods. The mean for a given period are calculated according to the formula:

$$mean = \Sigma_i N_{valid,i} C_{valid,i} / \Sigma_i N_{valid,i} \qquad i=1 \quad n$$

Where

$N_{valid,i}$ = the number of hours of the valid sampling period i
$C_{valid,i}$ = concentration for valid sampling period i

When a sampling period falls over two calendar years, N_i refers to the hours in the considered calendar year only.

Annex E List of pollutants with Monitoring Objective

Pollutant name	Pollutant code
Ammonium in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1045
Nitrate in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1046
sulphate in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1047
Elemental carbon in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1771
Organic carbon in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1772
calcium in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1629
magnesium in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1659
potassium in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1657
sodium in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1668
chloride in PM2.5 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1631
Lead (precip)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/2012
Cadmium (precip)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/2014
Nickel (precip)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/2015
Arsenic (precip)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/2018
Mercury (precip+dry_dep)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/7013
Elemental Gaseous Mercury (air+aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/4013
Total gaseous mercury (air+aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/4813
reactive_mercury (air+aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/653
Mercury in PM10 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/5013
Benzo(a)anthracene in PM10 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/5610
Benzo(b)fluoranthene in PM10 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/5617
Benzo(j)fluoranthene in PM10 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/5759
Benzo(k)fluoranthene in PM10 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/5626
indeno_123cd_pyrene in PM10 (aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/5655
Dibenzo(ah)anthracene in PM10 (air+aerosol)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/5763
Benzo(a)pyrene (precip+dry_dep)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/7029
Benzo(a)anthracene (precip+dry_dep)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/611
Benzo(b)fluoranthene (precip+dry_dep)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/618
Benzo(j)fluoranthene (precip+dry_dep)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/760
Benzo(k)fluoranthene (precip+dry_dep)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/627
indeno_123cd_pyrene (precip+dry_dep)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/656
Dibenzo(ah)anthracene (precip+dry_dep)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/7419
Benzene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/20
Ethane (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/428
Ethene (Ethylene) (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/430
Ethyne (Acetylene) (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/432
Propane (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/503
Propene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/505
n-Butane (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/394
i-Butane (2-methylpropane) (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/447
1-Butene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/6005
trans-2-Butene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/6006
cis-2-Butene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/6007

1.3 Butadiene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/24
n-Pentane (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/486
i-Hexane (2-methylpentane) (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/316
1-Pentene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/6008
2-Pentenes (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/6009
Isoprene (2-methyl-1,3-butadiene) (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/451
n-Hexane (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/443
i-Hexane (2-methylpentane) (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/316
n-Heptane (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/441
n-Octane (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/475
i-Octane (2,2,4-trimethylpentane) (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/449
Toluene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/21
Ethyl benzene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/431
m,p-Xylene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/464
o-Xylene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/482
1,2,4-Trimethylbenzene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/6011
1,2,3-Trimethylbenzene (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/6012
Total non-methane hydrocarbons (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/32
Formaldehyde (air)	http://dd.eionet.europa.eu/vocabulary/aq/pollutant/25