

## AirBase version 2 data products on EEA data service

A short technical description. Hermann Peifer, EEA, April 2008

### 1. Introduction

AirBase is a public air quality database containing air quality monitoring information for more than 35 countries throughout Europe. The list of countries included in AirBase is available as Annex A. The complete list of available components is given as Annex B. AirBase version 2 data products are available through the EEA data service at: <http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=1029>

General AirBase meta-information and links to further information on legal background (EoI decision) and annual data collection process are available at the indicated URL. The objective of the present document is not to repeat information available elsewhere but to describe the most relevant technical details of AirBase data products in order to facilitate their usage.

This document focuses on the data sets which are available for download. However, there are more AirBase data services available: pivot applications for on-line browsing through the data and AirBase viewer, a map-based user interface to the data. Both data services can be reached via the above-mentioned URL.

In case of any questions related to AirBase data products, please contact: [hermann.peifer@eea.europa.eu](mailto:hermann.peifer@eea.europa.eu)

### 2. European data set

#### a) Data available in CSV format

##### AirBase\_v2\_stations.csv

Table with meta-information on AirBase stations (6363 records). Station identifier and primary key for this table is: `station_european_code`. The information covers, among others: `station_name`, `station_local_code`, `station_altitude`, `type_of_station`, `station_type_of_area` and station co-ordinates.

##### AirBase\_v2\_statistics.csv

Table with annual statistics calculated from AirBase data (1356010 records). Relevant fields with statistical information are: `statistic_name`; `statistic_value`; `statistics_percentage_valid` and `statistics_number_valid`.

The primary key for this table is the combination of `station_european_code`, `component_code`, `measurement_european_code`, `statistics_year`, `statistics_name` and `statistics_average_group`. The latter field describes the data type, on which the statistics are based. The majority of statistics (>95%) is based on hourly or daily data, or on daily maximum (dymax) values. A smaller number of statistics is available for the following `statistics_average` groups: 2week, 3hour, 3month, 4week, month, var (i.e. variable), week and year.

A 1-to-many relationship between station and statistics tables can be established by using the station identifier: `station_european_code`. The relationship between statistics and measurement\_configuration tables is possible by using the fields `station_european_code`, `component_code` and `measurement_european_code`.

##### AirBase\_v2\_measurement\_configurations.csv

Table with measurement configurations (28365 records), describing the conditions under which a component is measured at the station. The information covers, among others: `measurement_technique_principle`, `measurement_equipment`, `measurement_start_date`, `measurement_automatic`, `sampling_time_unit`, `integration_time_unit`, `calibration_frequency` and `calibration_method`.

Identifier for a measurement configuration is the combination of fields: `station_european_code`, `component_code` and `measurement_european_code`. These fields are also used in the statistics table (see above) and in the names of raw data files (see section 3).

#### Technical details of the CSV format

The above tables are available UTF-8 encoded text files. Field delimiter is a tab character (U+0009), records end with carriage return + line feed characters (U+000D, U+000A). A header row with field names is always included. Numerical values use the dot as decimal separator. There is no qualifier around string values.

b) Data available in XML format

AirBase\_v2\_xmldata.zip:

This zip archive contains 1 XML file per country with meta-information on networks, stations and measurement configurations. The XML files also include all statistical values. A description of the document structure in AirBase XML files including a list of legal elements and attributes is available in form of a document type definition (DTD), see: <http://air-climate.eionet.europa.eu/airbase.dtd>

All tables mentioned in section 2a) have been directly generated from the XML files. Data extraction has been done by using XSLT (Extensible Stylesheet Language Transformations). The XSLT stylesheet that has been used for extracting statistical values (i.e. for generating AirBase\_v2\_statistics.csv) is given in Annex C. Another option for extracting data from the XML files is to use the Excel macros as available on [ETC/ACC website](#). More information on how to extract data from the XML files can be made available on demand.

### 3. Data by country (including raw data)

This section on EEA data service contains one zip archive per country with data in CSV and XML format, as explained in section 2. Country data files have been generated by splitting the European data sets into "slices" for individual countries. There is no other difference between country data and European data sets.

The country zip archives also include raw data files. These are organised in text files, each one representing 1 time series of data. All necessary meta-information about the time series is included in the file name, e.g.:

FR0871A0000700100day.1-1-1999.31-12-2006

Position 1 – 7 station\_european\_code (here: FR0871A)  
Position 8 – 12 component\_code (here: 00007, i.e. Ozone)  
Position 13 – 17 measurement\_european\_code (here: 00100)  
Position 18 until first dot: data type (here: day)  
followed by start date and end date of the time series

Data types in raw data file names correspond to the statistics\_average\_group values as used in the statistics tables, see section 2a).

Raw data files are tab-delimited text files. The general format of hourly and daily raw data files is as follows:

Date Value1 Flag1 Value2 Flag2 Value3 Flag3 Value4 Flag4 ..... ValueN FlagN

Hourly data (data type: hour and hour8)

The date in column 1 identifies the calendar day to which the record relates. Measurement values 1...24 and quality flags 1...24 in columns 2-49 correspond to the hours of that date. Value 1 is understood to represent the measurements during the first hour of the day (i.e.: 00:00-00:59), and so on.

Raw data files with data type hour8 contain running 8-hour means calculated from the corresponding hourly data. Their format is identical to hourly data, with 24 values and 24 flags per day. However, any hour8 value represents the mean of hourly values in the preceding 8-hour period. E.g. hour8 value 10 is based on hourly values 3...10 of the same day, the averaging period for hour8 value 1 starts on 17:00H of the preceding day.

Daily data (data type: day and dymax)

The date in column 1 identifies the calendar month to which the record relates. Measurement values 1...31 and quality flags 1...31 in columns 2-63 correspond to the days of that month. All records have 31 values and 31 flags. For months with less than 31 days, the records are filled with invalid data.

Raw data files with data type dymax contain daily maximum values, based on running 8-hour means (see above). Both hour8 and dymax data files are only available for ozone and CO.

Quality flags in raw data

Flag values indicate the quality of the preceding measurement value. A quality flag value > 0 indicates valid measurement data. A quality flag <= 0 indicates invalid or missing data.

## Annex A: Country codes, names and share in AirBase data

code	country_name	% data	# components
AT	Austria	9.8	31
BA	Bosnia and Herzegovina	0.1	8
BE	Belgium	3.0	19
BG	Bulgaria	0.7	21
CH	Switzerland	2.3	13
CY	Cyprus	0.0	6
CZ	Czech Republic	3.9	15
DE	Germany	24.9	29
DK	Denmark	0.5	19
EE	Estonia	0.3	7
ES	Spain	12.6	17
FI	Finland	1.1	11
FR	France	11.6	12
GB	United Kingdom	8.3	82
GR	Greece	1.0	11
HR	Croatia	0.0	8
HU	Hungary	0.4	10
IE	Ireland	0.3	17
IS	Iceland	0.1	15
IT	Italy	8.2	31
LI	Liechtenstein	0.0	4
LT	Lithuania	0.3	12
LU	Luxembourg	0.0	2
LV	Latvia	0.2	17
MK	Macedonia, FYR of	0.3	9
MT	Malta	0.1	18
NL	Netherlands	3.4	65
NO	Norway	0.4	12
PL	Poland	1.9	15
PT	Portugal	2.0	13
RO	Romania	0.4	20
RS	Serbia	0.1	9
SE	Sweden	0.4	13
SI	Slovenia	0.3	6
SK	Slovakia	1.0	14

### N.B.

code            country\_iso\_code as used in AirBase  
 % data        relative share of country data in AirBase data, 0.0 means less than 0.05%  
 # components   number of components available in AirBase data

AD and TR     Andorra and Turkey are missing in the above table, as no measurement data are available. However, a small number of air quality stations is registered in AirBase for these two countries (Andorra: 2, Turkey: 3).

## Annex B: AirBase component codes, names, units and relative share in raw data

code	component_name	measurement_unit	FWD	% data	# countries
1	Sulphur dioxide (air)	ug/m3	yes	13.9	33
3	Strong acidity (air)	ug SO2/m3	no	0.1	7
4	Total suspended particulates (aerosol)	ug/m3	yes	3.1	18
5	Particulate matter < 10 µm (aerosol)	ug/m3	yes	5.0	34
6	Black smoke (air)	ug/m3	yes	0.2	15
7	Ozone (air)	ug/m3	yes	27.5	34
8	Nitrogen dioxide (air)	ug/m3	yes	16.8	34
9	Nitrogen oxides (air)	ug NO2/m3	yes	4.2	27
10	Carbon monoxide (air)	mg/m3	yes	15.2	33
11	Hydrogen sulphide (air)	ug/m3	no	0.1	6
12	Lead (aerosol)	ug/m3	yes	0.1	19
13	Mercury (aerosol)	ng/m3	yes	0.0	1
14	Cadmium (aerosol)	ng/m3	yes	0.0	14
15	Nickel (aerosol)	ng/m3	yes	0.0	12
16	Chromium (aerosol)	ng/m3	no	0.0	2
17	Manganese (aerosol)	ng/m3	no	0.0	2
18	Arsenic (aerosol)	ng/m3	yes	0.0	11
19	Carbon disulphide (air)	ug/m3	no	0.0	1
20	Benzene (air)	ug/m3	yes	0.9	28
21	Toluene (air)	ug/m3	no	0.4	12
22	Styrene (air)	ug/m3	no	0.0	2
24	1.3 Butadiene (air)	ug/m3	no	0.1	1
25	Formaldehyde (air)	ug/m3	no	0.0	1
26	Trichloroethylene (air)	ug/m3	no	0.0	1
27	Tetrachloroethylene (air)	ug/m3	no	0.0	1
29	Benzo(a)pyrene (precip)	ng/m3	no	0.0	2
30	Polyaromatic hydrocarbons (air+aerosol)	ng/m3	no	0.0	1
32	Total non-methane hydrocarbons (air)	ug C/m3	no	0.1	5
33	Total volatile organic compounds (air)	ug/m3	no	0.1	4
34	Peroxyacetyl nitrate (air)	ug/m3	no	0.0	1
35	Ammonia (air)	ug/m3	no	0.1	5
36	Wet nitrogen deposition (flux)	mg N/(m2*m)	no	0.0	1
37	Wet sulphur deposition (flux)	mg S/(m2*m)	no	0.0	2
38	Nitrogen monoxide (air)	ug/m3	no	9.7	26
39	Hydrogen chloride (air)	ug/m3	no	0.0	2
40	Hydrogen fluoride (air)	ug/m3	no	0.0	1
41	Methane (air)	ug/m3	no	0.1	4
45	Particulate ammonium (aerosol)	ug/m3	no	0.0	2
46	Particulate nitrate (aerosol)	ug/m3	no	0.0	2
47	Particulate sulphate (aerosol)	ug/m3	no	0.0	8
51	HC C2-C6(excl. AROM. & CHLH) (air+aerosol)	ug/m3	no	0.0	1
63	Zinc (aerosol)	ng/m3	no	0.0	2
65	Iron (aerosol)	ng/m3	no	0.0	1
67	Total nitrate (air+aerosol)	ug N/m3	no	0.0	2
68	Total ammonium (air+aerosol)	ug N/m3	no	0.0	2
69	Radioactivity ( )		no	0.0	1
73	Copper (aerosol)	ng/m3	no	0.0	1
76	acidity (aerosol)	ne H/m3	no	0.0	2

N.B.:

code component\_code as used in AirBase

FWD = yes mandatory pollutant

% data relative share of component in AirBase data, 0.0 means less than 0.05%

# countries number of countries reporting this component

The component names in this table include both, the actual pollutant name and the matrix in which the concentrations are measured.

## Annex B: AirBase component codes, names, units and relative share in raw data (continued)

code	component_name	measurement_unit	FWD	% data	# countries
316	i-Hexane (air)	ug/m3	no	0.0	2
351	acenaphtene (air+aerosol)	ng/m3	no	0.0	1
380	Benzo(b+j+k)fluoranthenes (air+aerosol)	ng/m3	no	0.0	1
394	n-Butane (air)	ug/m3	no	0.1	2
412	conductivity (precip)	uS/cm	no	0.0	1
416	cyclohexane (air)	pptv	no	0.0	1
428	Ethane (air)	ug/m3	no	0.1	2
430	Ethene (Ethylene) (air)	ug/m3	no	0.1	2
431	Ethyl benzene (air)	ug/m3	no	0.1	7
432	Ethyne (Acetylene) (air)	ug/m3	no	0.1	2
435	fluorene (air+aerosol)	ng/m3	no	0.0	1
441	n-Heptane (air)	ug/m3	no	0.1	3
443	n-Hexane (air)	ug/m3	no	0.1	3
447	i-Butane (air)	ug/m3	no	0.1	1
449	i-Octane (air)	ug/m3	no	0.0	2
450	i-Pentane (air)	ug/m3	no	0.1	2
451	Isoprene (air)	ug/m3	no	0.1	1
464	m,p-Xylene (air)	ug/m3	no	0.1	8
475	n-Octane (air)	ug/m3	no	0.0	3
482	o-Xylene (air)	ug/m3	no	0.1	8
486	n-Pentane (air)	ug/m3	no	0.1	3
503	Propane (air)	ug/m3	no	0.1	2
505	Propene (air)	ug/m3	no	0.1	2
520	sum_sulph_diox_sulphate (air+aerosol)	ug S/m3	no	0.0	1
606	anthracene (air+aerosol)	ng/m3	no	0.0	1
609	Benzo(a)anthracene (air+aerosol)	ng/m3	no	0.0	1
616	Benzo(b)fluoranthene (air+aerosol)	ng/m3	no	0.0	1
622	Benzo(ghi)perylene (air+aerosol)	ng/m3	no	0.0	1
625	Benzo(k)fluoranthene (air+aerosol)	ng/m3	no	0.0	1
629	calcium (aerosol)	ug/m3	no	0.0	1
630	calcium (precip)	mg/l	no	0.0	1
631	chloride (aerosol)	ug/m3	no	0.0	1
632	chloride (precip)	mg/l	no	0.0	1
643	fluoranthene (air+aerosol)	ng/m3	no	0.0	1
648	acidity (precip)	ue H/l	no	0.0	1
654	indeno_123cd_pyrene (air+aerosol)	ng/m3	no	0.0	1
658	potassium (precip)	mg/l	no	0.0	1
660	magnesium (precip)	mg/l	no	0.0	1
664	ammonium (precip)	mg N/l	no	0.0	1
666	nitrate (precip)	mg N/l	no	0.0	1
669	sodium (precip)	mg/l	no	0.0	1
673	PCB_114 (air+aerosol)	pg/m3	no	0.0	1
674	PCB_118 (air+aerosol)	pg/m3	no	0.0	1
677	PCB_138 (air+aerosol)	pg/m3	no	0.0	1
679	PCB_141 (air+aerosol)	pg/m3	no	0.0	1
680	PCB_153 (air+aerosol)	pg/m3	no	0.0	1
683	PCB_157 (air+aerosol)	pg/m3	no	0.0	1
684	PCB_167 (air+aerosol)	pg/m3	no	0.0	1
685	PCB_170 (air+aerosol)	pg/m3	no	0.0	1
686	PCB_180 (air+aerosol)	pg/m3	no	0.0	1
689	PCB_183 (air+aerosol)	pg/m3	no	0.0	1
690	PCB_187 (air+aerosol)	pg/m3	no	0.0	1
691	PCB_189 (air+aerosol)	pg/m3	no	0.0	1
692	PCB_194 (air+aerosol)	pg/m3	no	0.0	1
695	PCB_28 (air+aerosol)	pg/m3	no	0.0	1

**Annex B: AirBase component codes, names, units and relative share in raw data (continued)**

code	component_name	measurement_unit	FWD	% data	# countries
701	PCB_52 (air+aerosol)	pg/m3	no	0.0	1
706	PCB_74 (air+aerosol)	pg/m3	no	0.0	1
707	PCB_99 (air+aerosol)	pg/m3	no	0.0	1
709	PCB_123 (air+aerosol)	pg/m3	no	0.0	1
712	phenanthrene (air+aerosol)	ng/m3	no	0.0	1
715	pyrene (air+aerosol)	ng/m3	no	0.0	1
719	sulphate (precip)	mg S/l	no	0.0	1
753	precipitation_amount (precip)	mm	no	0.0	1
754	precipitation_amount_off (precip)	mm	no	0.0	1
2012	Lead (precip)	ug/l	no	0.0	1
2014	Cadmium (precip)	ug/l	no	0.0	1
2063	Zinc (precip)	ug/l	no	0.0	1
2065	Iron (precip)	ug/l	no	0.0	1
2073	Copper (precip)	ug/l	no	0.0	1
2076	acidity (precip)	pH units	no	0.0	1
4013	Mercury (air+aerosol)	ng/m3	no	0.0	3
4330	PCB_105 (air+aerosol)	pg/m3	no	0.0	1
4336	PCB_149 (air+aerosol)	pg/m3	no	0.0	1
4339	PCB_156 (air+aerosol)	pg/m3	no	0.0	1
4341	PCB_18 (air+aerosol)	pg/m3	no	0.0	1
4347	PCB_31 (air+aerosol)	pg/m3	no	0.0	1
4406	chrysene (air+aerosol)	ng/m3	no	0.0	1
5012	Lead in PM10 (aerosol)	ug/m3	no	0.0	2
5014	Cadmium in PM10 (aerosol)	ng/m3	no	0.0	2
5015	Nickel in PM10 (aerosol)	ng/m3	no	0.0	2
5018	Arsenic in PM10 (aerosol)	ng/m3	no	0.0	2
5029	Benzo(a)pyrene in PM10 (aerosol)	ng/m3	no	0.0	2
5609	Benzo(a)anthracene in PM10 (air+aerosol)	ng/m3	no	0.0	1
5616	Benzo(b)fluoranthene in PM10 (air+aerosol)	ng/m3	no	0.0	1
5625	Benzo(k)fluoranthene in PM10 (air+aerosol)	ng/m3	no	0.0	1
5654	indeno_123cd_pyrene in PM10 (air+aerosol)	ng/m3	no	0.0	1
5762	Benzo(j)fluoranthene in PM10 (air+aerosol)	ng/m3	no	0.0	1
5763	Dibenzo(ah)anthracene in PM10 (air+aerosol)	ng/m3	no	0.0	1
6001	Particulate matter < 2.5 µm (aerosol)	ug/m3	yes	0.3	23
6002	Particulate matter < 1 µm (aerosol)	ug/m3	no	0.0	2
6005	1-Butene (air)	ug/m3	no	0.1	2
6006	trans-2-Butene (air)	ug/m3	no	0.1	2
6007	cis-2-Butene (air)	ug/m3	no	0.1	2
6008	1-Pentene (air)	ug/m3	no	0.0	2
6009	2-Pentenes (air)	ug/m3	no	0.1	2
6011	1,2,4-Trimethylbenzene (air)	ug/m3	no	0.0	4
6012	1,2,3-Trimethylbenzene (air)	ug/m3	no	0.0	3
6013	1,3,5-Trimethylbenzene (air)	ug/m3	no	0.0	4
6015	Benzo(a)pyrene (air+aerosol)	ng/m3	no	0.0	6

## Annex C: XSLT Stylesheet for extracting annual statistics from XML files

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
<!-- Stylesheet created by Hermann Peifer, EEA, March 2008 -->

<!-- Variable definition for tab delimited output -->
<xsl:variable name="newline" select="'&#x0A;'" />
<xsl:variable name="tab" select="'&#x09;'" />
<xsl:output method="text" encoding="UTF-8" />

<xsl:template match="/">

  <!-- Write out field names -->
  <xsl:text>station_european_code</xsl:text>          <xsl:value-of select="$tab" />
  <xsl:text>component_code</xsl:text>              <xsl:value-of select="$tab" />
  <xsl:text>measurement_european_code</xsl:text>   <xsl:value-of select="$tab" />
  <xsl:text>measurement_automatic</xsl:text>       <xsl:value-of select="$tab" />
  <xsl:text>measurement_technique_principle</xsl:text> <xsl:value-of select="$tab" />
  <xsl:text>component_name</xsl:text>              <xsl:value-of select="$tab" />
  <xsl:text>component_caption</xsl:text>          <xsl:value-of select="$tab" />
  <xsl:text>component_FWD</xsl:text>              <xsl:value-of select="$tab" />
  <xsl:text>measurement_unit</xsl:text>           <xsl:value-of select="$tab" />
  <xsl:text>statistics_period</xsl:text>          <xsl:value-of select="$tab" />
  <xsl:text>statistics_year</xsl:text>            <xsl:value-of select="$tab" />
  <xsl:text>statistics_average_group</xsl:text>    <xsl:value-of select="$tab" />
  <xsl:text>statistic_shortname</xsl:text>        <xsl:value-of select="$tab" />
  <xsl:text>statistic_name</xsl:text>            <xsl:value-of select="$tab" />
  <xsl:text>statistic_value</xsl:text>           <xsl:value-of select="$tab" />
  <xsl:text>statistics_percentage_valid</xsl:text> <xsl:value-of select="$tab" />
  <xsl:text>statistics_number_valid</xsl:text>     <xsl:value-of select="$tab" />
  <xsl:text>statistics_calculated</xsl:text>      <xsl:value-of select="$newline" />

  <xsl:apply-templates select="airbase/country/network/station/measurement_configuration" />

</xsl:template>

<!-- Loop through the measurement configurations -->
<xsl:template match="measurement_configuration">
  <xsl:param name="station_european_code" select="../station_info/station_european_code" />
  <xsl:param name="component_code" select="measurement_info/component_code" />
  <xsl:param name="measurement_european_code" select="measurement_info/measurement_european_code" />
  <xsl:param name="measurement_automatic" select="measurement_info/measurement_automatic" />
  <xsl:param name="measurement_technique_principle" select="measurement_info/measurement_technique_principle" />
  <xsl:param name="component_name" select="measurement_info/component_name" />
  <xsl:param name="component_caption" select="measurement_info/component_caption" />
  <xsl:param name="component_FWD" select="measurement_info/component_FWD" />
  <xsl:param name="measurement_unit" select="measurement_info/measurement_unit" />

  <!-- Loop through the statistic results, exclude statistics_average_groups where value = hour8 -->
  <xsl:for-each select="statistics/statistics_average_group[@value != 'hour8']/statistic_set/statistic_result">

    <xsl:value-of select="$station_european_code" />          <xsl:value-of select="$tab" />
    <xsl:value-of select="$component_code" />              <xsl:value-of select="$tab" />
    <xsl:value-of select="$measurement_european_code" />    <xsl:value-of select="$tab" />
    <xsl:value-of select="$measurement_automatic" />       <xsl:value-of select="$tab" />
    <xsl:value-of select="$measurement_technique_principle" /> <xsl:value-of select="$tab" />
    <xsl:value-of select="$component_name" />              <xsl:value-of select="$tab" />
    <xsl:value-of select="$component_caption" />          <xsl:value-of select="$tab" />
    <xsl:value-of select="$component_FWD" />              <xsl:value-of select="$tab" />
    <xsl:value-of select="$measurement_unit" />           <xsl:value-of select="$tab" />
    <xsl:value-of select="../../../../statistics_period" /> <xsl:value-of select="$tab" />
    <xsl:value-of select="../../../../@Year" />           <xsl:value-of select="$tab" />
    <xsl:value-of select="../../../../@value" />          <xsl:value-of select="$tab" />
    <xsl:value-of select="statistic_shortname" />         <xsl:value-of select="$tab" />
    <xsl:value-of select="statistic_name" />              <xsl:value-of select="$tab" />
    <xsl:value-of select="statistic_value" />             <xsl:value-of select="$tab" />

    <xsl:choose>
      <xsl:when test="../@type = 'Exceedances'"> <!-- Take values from General statistic_set -->
        <xsl:value-of select="../../../../statistic_set[@type = 'General']/statistics_percentage_valid" />
        <xsl:value-of select="$tab" />
        <xsl:value-of select="../../../../statistic_set[@type = 'General']/statistics_number_valid" />
        <xsl:value-of select="$tab" />
        <xsl:value-of select="../../../../statistic_set[@type = 'General']/statistics_calculated" />
        <xsl:value-of select="$newline" />
      </xsl:when>
      <xsl:otherwise> <!-- Take values from parent statistic_set -->
        <xsl:value-of select="../statistics_percentage_valid" /> <xsl:value-of select="$tab" />
        <xsl:value-of select="../statistics_number_valid" /> <xsl:value-of select="$tab" />
        <xsl:value-of select="../statistics_calculated" /> <xsl:value-of select="$newline" />
      </xsl:otherwise>
    </xsl:choose>

  </xsl:for-each>

</xsl:template>
</xsl:stylesheet>
```