



Waterbase – Groundwater

Version 9

Quality control documentation

04 May 2009

Waterbase – Groundwater

In the context of the implementation of the Water Framework Directive (WFD), the European Environment Agency (EEA) EIONET-Water annual data flow for waters is in the process of being transferred into the WISE 'State of the Environment' (SoE) voluntary data flow. With this it remains one of the EIONET Priority Data Flows, but gains full integration into the reporting under WISE as the single entry point of water information in Europe and complementarily with data collected under the WFD. Most information that is used for European level 'state of environment' assessments needs to be provided by member countries and there it usually comes from monitoring networks that are to meet several assessment purposes, SOE, as well as different legal requirements..

Data on groundwater are collected annually through the WISE-SoE data collection process. Data and information obtained through the WISE-SoE data collection process are primarily used to compile indicator factsheets, associated with the EEA's Core Set Indicators, upon which EEA assessment reports are based. Data collected through the WISE-SoE data collection process are also published in WISE map viewer, Waterbase, a series of water topic-specific databases and web pages, publicly accessible via the EEA Data Service's web site.

Groundwater dataset include physical characteristics and proxy pressures information of the groundwater bodies and chemical quality data on selected nutrients and hazardous substances in groundwater.

QA/QC activities

This document briefly presents the ETC/Water and the EEA activities focused on quality of the Waterbase - Groundwater dataset and the results of these activities. In addition a warning is given on the use of certain records for analytical purposes (see section 2, 3 and 4).

The Quality control tests have been performed on the Waterbase - Groundwater database provided in April 2009 by ETC/WTR. This database is included in the EEA data service as version 9, and is publicly available. The database and metadata are available at the following URL:

<http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=1090>

Waterbase – Groundwater dataset contains four data tables:

- BODIES
- QUALITY

Five types of the tests have been performed on the data tables. Basic tests, Logical rules violation test, Outlier detection, Groundwater bodies data tests and Data type tests.

1. Basic tests

1.1 Summary

1.1.1 Waterbase - Groundwater: Bodies

Country Code	Number of records								
	latest (2008) country delivery						ETC working database	Waterbase	
	total	processed				note*		Total	with quality issue/s detected
		total		with quality issue/s detected by the ETC					
new	redelivered	new	redelivered						
AL	0	0	0	0	0		10	10	
AT	14	14	0	0	0		28	28	
BA	3	2	1	2	0	1	4	4	
BE	55	55	0	0	0		110	110	2
BG	138	5	133	0	0		270	270	28
CH	22	22	0	0	0		38	38	
CY	3	0	3	0	0		3	3	
CZ	221	221	0	0	0		265	265	
DE	428	428	0	0	0		438	438	
DK	0	0	0	0	0		3	3	
EE	5	0	5	0	0		5	5	
ES	275	275	0	0	0		434	434	
FI	41	0	41	0	0		41	41	
FR	1118	561	557	0	0		3073	3073	
GB	0	0	0	0	0		41	41	
GR	0	0	0	0	0		362	362	
HR	0	0	0	0	0		0	0	
HU	36	36	0	0	0		54	54	
IE	150	150	0	0	0		153	153	3
IS	1	1	0	0	0		2	2	
IT	1	1	0	1	0	2	45	45	
LI	0	0	0	0	0		1	1	
LT	0	0	0	0	0		7	7	
LU	10	6	4	0	0		11	11	
LV	16	0	16	0	0		20	20	
ME	7	0	7	0	0		10	10	
MK	7	0	7	0	0		7	7	
MT	2	0	2	0	0		2	2	
NL	0	0	0	0	0		9	9	
NO	0	0	0	0	0		1	1	
PL	3	0	3	0	0		173	173	3
PT	83	1	82	0	0		93	93	
RO	25	25	0	0	0		25	25	
RS	10	0	10	0	0		11	11	
SE	42	23	19	0	0		45	45	
SI	0	0	0	0	0		11	11	
SK	101	101	0	0	0		111	111	
TR	4	4	0	0	0		4	4	
Total	2709	1826	890	3	0		5920	5920	36

*

1 - three municipalities used instead of groundwater bodies

2 - virtual groundwater body "IT_ITALY_whole-country"

1.1.2 Waterbase - Groundwater: Quality

- latest delivery and ETC working database - both disaggregated and aggregated data together
 - Waterbase - only aggregated data

Country Code	Number of records									
	latest (2008) country delivery							ETC working database	Waterbase	
	total	processed by the ETC				excluded			Total	with quality issue/s detected
		total		with quality issue/s detected by the ETC		number	reason*			
	new	redelivered	new	redelivered						
AL	0	0	0	0	0	0		34	34	1
AT	106595	106595	0	0	0	0		488640	3633	
BA	495	495	0	0	0	0		1173	12	
BE	29387	29387	0	0	0	0		83741	3456	
BG	1613	1386	0	0	0	227	1	14602	5593	
CH	15676	8347	7329	0	0	0		16347	748	
CY	559	549	0	0	0	10	1	2110	38	1
CZ	42812	42793	0	0	0	19	2	139326	6192	
DE	9961	9925	0	0	0	36	3	66019	10544	51
DK	0	0	0	0	0	0		37961	293	
EE	796	781	0	0	0	15	1	50213	704	
ES	1875	1875	0	0	0	0		3219	2646	
FI	130	112	18	0	0	0		2015	2015	
FR	235825	212179	0	0	0	23646	4	511987	19148	
GB	0	0	0	0	0	0		24694	817	1
GR	0	0	0	0	0	0		1934	1393	
HR	0	0	0	0	0	0		0	0	
HU	216	216	0	0	0	0		1114	1114	2
IE	8272	8272	0	0	0	0		8387	2683	
IS	52	48	0	0	0	4	2	115	12	
IT	14979	14453	0	0	0	526	5	15912	111	
LI	0	0	0	0	0	0		31	31	
LT	626	622	4	0	0	0		1329	465	
LU	45	45	0	0	0	0		165	75	48
LV	523	507	0	0	0	16	1	1141	246	
ME	28	28	0	2	0	0		68	68	8
MK	0	0	0	0	0	0		0	0	
MT	6	6	0	0	0	0		862	36	
NL	18	18	0	0	0	0		334	334	
NO	401	401	0	0	0	0		1196	42	
PL	273	264	0	0	0	9	1	2453	99	
PT	2146	1826	0	0	0	320	1	6829	771	
RO	171	171	0	0	0	0		171	171	
RS	1344	1344	0	0	0	0		3722	210	
SE	151	101	0	0	0	50	1	328	325	
SI	7488	7488	0	0	0	0		32163	696	
SK	12425	12425	0	0	0	0		36050	1775	
TR	20	20	0	0	0	0		20	20	
Total	494908	462679	7351	2	0	24878		1556405	66550	112

*

- 1 - aggregated data replaced by disaggregated
- 2 - data of insufficient quality removed
- 3 - duplicates removed
- 4 - duplicates and records from unknown stations removed
- 5 - measurements taken later than in 2007 and records from unknown stations removed

1.2 Primary key tests

Primary key is a field or combination of fields with values which have to be unique in the data table. If primary key is duplicated it is an error which has to be solved.

List of data tables primary keys:

BODIES: CountryCode, WaterbaseID

QUALITY: CountryCode, WaterbaseID, Determinand, Year, AggregationPeriod

Result:

903 QUALITY records from AT and 307 QUALITY records from ES are duplicated. The duplicated records were either aggregated differently (AT) or aggregated by different contributors (ES). The duplications will be tried to be solved in the near future.

1.3 Table relations tests

The unique Waterbase identifier (WaterbaseID) is contained in each of the data tables. It can be used to link data from one table to another. The table relations tests detect identifiers which are not present in some of the tables.

1.3.1 Number of GW bodies without any data in the "QUALITY" table by country

Country Code	No. of bodies	Percentage of total no. of bodies
BA	1	25.00
BG	86	31.85
CH	16	42.11
CZ	49	18.49
DE	28	6.39
ES	13	3.00
FR	2193	71.36
GR	51	14.09
IE	18	11.76
IS	1	50.00
IT	1	2.22
LU	5	45.45
LV	2	10
MK	7	100
PL	170	98.27
PT	7	7.53
RO	1	4
SE	2	4.44
SK	32	28.83
Total	2683	45.32

1.3.2 “QUALITY” table records where “WaterbaseID” is not present in the “BODIES” table

Country Code	No. of records	Percentage of total no. of records
LU	48	64
Total	48	0.07

All of these records are marked in the dataset (see section 4 for more details)

2. Logical rule violation tests

Logical rules were tested in the “QUALITY” data table. This table contains several measurement value fields, calculated in the aggregation process. Logical relations can be detected between them and mathematically transformed in a set of rules. Following rules have been detected and tested:

Rule	Basic validation rules
1	Mean >= Minimum
2	Mean <= Maximum
3	Median >= Minimum
4	Median <= Maximum
5	Minimum <= Maximum
6	StandardDeviation < Maximum
Rule	Combined validation rules
13	IF Minimum < Maximum THEN (StandardDeviation > 0)
14	IF NumberOfSamples = 1 THEN (Mean = Minimum = Maximum = Median)
15	IF NumberOfSamples = 1 THEN (StandardDeviation = 0)
16	IF NumberOfSamples = 0 THEN (AllValueType Is Null)
Rule	Negative value validation rule
17	All Values >= 0

The following exceptions and modifications were been applied:

IF Maximum = 0 AND StandardDeviation = 0 THEN rule 6 is not violated

IF Determinand = Temperature the values can be negative (exception of the rule 17)

IF Rule 13 is violated THEN change StandardDeviation to Null

A special QA field (QA_LRviolations) has been added to the data tables. Information of the rules violated in the respective record are kept there as a coma separated list of those rules numbers (the numbers are the same as in the table above). It is recommended that the records where QA_LRviolation field is not empty (**3 Quality records**), should not be used in a further analysis. The detected data quality inconsistencies will be tried to be solved in the near future.

3. Outlier detection

Detection of outliers was performed on the “QUALITY” data table. Following values were analyzed:

Measurement values: mean

Determinands: all

Aggregation periods: all

Years: all

Measurement values were compared with other values from the same time series. If the value was detected as an outlier it was analyzed whether it can be a possible error or whether it was caused by natural conditions.

Some of previously detected errors have been already corrected by countries or were approved as natural high/low values.

Some whole time series where the measurement values are naturally very high (e.g. because of the positioning of the monitoring station close to the source of the pollution) have been also detected. These time series have not been included in the subset used for the WISE update.

Last part of the outlier detection process was detection of records where Mean value is not provided.

All types of the information mentioned above have been stored in a special QA field (QA_outlier) added to data table. Following QA flags have been used:

-1 – record has been confirmed by the respective country as correct (**11 Quality records**)

1 – standard potential outlier - value is either higher/lower than limit value or is suspiciously high/low comparing to the rest of the time series or value change between two consecutive values is suspiciously abrupt or was marked as an potential outlier by a content expert (**1 Quality record**)

10 – the Mean value = 0 (**60 Quality records**). Value is not correct and records should not be used.

4. Groundwater bodies availability tests

If the groundwater body available in the Quality data table is not available in the Bodies table, this information is stored in a special QA field (QA_GWbody_problems). Following QA flags have been used:

99 – GW body is not available in the Bodies table (**48 Quality records**) – see result 1.3.2

These data quality inconsistencies will be tried to be solved in the near future.

5. Data type tests

All Groundwater dataset values have to follow specifications defined in the respective Data dictionary. Values, which are of a different data type as requested (e.g. string instead of numeric) or which are not available in a set of allowable values, have been either removed or, if possible, replaced by a correct value. The original, incorrect value has been stored in a special QA field (QA_datatype_error) in the following format:

Name_of_field: Erroneous_Value; [Name_of_field: Erroneous_Value; ...]

Test result summary:

Bodies table: 36 records