Waterbase – Groundwater
Version 8

Quality control documentation

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 Waterbase, a series of water topic-specific databases and web pages, publicly accessible via the EEA Data Service's web site.

Groundwater dataset include list of groundwater bodies and chemical quality data on nutrients, organic matter and selected pesticides 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 and 3).
The Quality control tests have been performed on the Waterbase - Groundwater database provided in April 2008 by ETC/WTR. This database is included in the EEA data service as version 8, and is publicly available. The database and metadata are available at the following URL:
http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=1040

Waterbase – Groundwater dataset contains three data tables:
- • QUALITY
- • BODIES

Three types of the tests have been performed on the data tables. Basic tests, Logical rules violation test and Outlier detection.

1. Basic tests

1.1 Summary

The groundwater quality data have been provided by countries in disaggregated (1 table) or aggregated form (divided to 5 tables - 1 for each of four nutrients and 1 for pesticides). If it was possible the ETC/Water aggregated the disaggregated data and processed them together with data delivered aggregated into the working database. The summary presents number of records subdivided by country for originally disaggregated data and originally aggregated data (total numbers for all 5 tables):

a) which were delivered by the country in the last delivery (very late deliveries are not included)
b) which was possible to process (for disaggregated data the reasons why some of the records was not possible to process and number of these records are also present; for aggregated data the reasons could be various and difficult to summarize)
c) total in the working database (for the disaggregated data the number of the records after the data were aggregated by the ETC is also shown)
d) which are present in the Waterbase - Groundwater v8 (for the disaggregated data the number of the records after the data were aggregated by the ETC is also shown)

1.1.1   Waterbase – Groundwater (disaggregated data)

| Country Code | Numbers of records | | | | | | | | |
| | in the latest delivery | | in the latest delivery - not processed - reasons* | | | total in the working database | | in the Waterbase | |
	Total	Processed	1	2	3	disaggregated	after aggregation	disaggregated	after aggregation
AL									
AT	25080	24818	262			382045	1312	382045	1312
BA	760	0		760					
BE	21810	21810				53515	946	53515	946
BG	5797	1838	31	9	3919	6853	676	6853	676
CH	8000	8000				8000	432	8000	432
CY	387	287	100			1561	27	1561	27
CZ	24435	24435				96533	1757	96533	1757
DE	14040	14040				56588	656	56588	656
DK						38396	293	38396	293
EE	775	775				49432	689	49432	689
ES						1297	361	1297	361
FI									
FR	162197	162179	18			247831	14571	247831	14571
GB						24015	138	24015	138
GR						1186	645	1186	645
HR									
HU									
IE									
IS	72	0				23	3	23	3

IT	14727	0		14727		1459	108	1459	108
LI									
LT	242	242				317	39	317	39
LU	30	30				60	0	60	0
LV	499	499				499	56	499	56
ME									
MK									
MT						844	18	844	18
NL									
NO	398	398				762	6	762	6
PL	264	264				2126	27	2126	27
PT						4827	168	4827	168
RO									
RS	1694	1694				2378	153	2378	153
SE									
SI	7360	7360				20574	434	20574	434
SK	2075	2066	7		2	23087	369	23087	369
TR									
Total	290642	270735	418	15496	3921	1024208	23884	1024208	23884

*

1 – value <= 0 or non-numeric value
2 – monitoring stations without identifiers and/or without relation to groundwater bodies
3 – data for the period 2002 - 2005 already in the database

1.1.2 Waterbase – Groundwater (aggregated data)

Country Code	Number of records			
	in the latest delivery		total in the working database	in the Waterbase
	Total	Processed		
AL			35	35
AT				
BA				
BE	169	169	822	822
BG			2918	2918
CH				
CY	6	6	1	1
CZ				
DE				
DK				
EE	20	0		
ES			47	47
FI	100	100	1910	1910
FR			0	0
GB			679	679
GR			748	748
HR			0	0

HU	54	54	898	898
IE			104	104
IS				0
IT				0
LI	20	20	31	31
LT			390	390
LU	15	15	60	60
LV	16	0	135	135
ME	4	0	0	0
MK			0	0
MT	6	6	12	12
NL	18	18	316	316
NO			33	33
PL	9	0	63	63
PT	176	176	176	176
RO				
RS				
SE	9	9	227	227
SI			5	5
SK			538	538
TR				
Total	622	573	10148	10148

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.

List of data tables primary keys:
BODIES: CountryCode, WaterbaseID
QUALITY: CountryCode, WaterbaseID, Determinand, Year, AggregationPeriod

Result:
No primary key error has been detected.

1.3 Table relations tests

The unique Waterbase identifier (WaterbaseID) is contained in both 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 bodies without any data in the "QUALITY" table by country

Country code	No. of stations	Percentage of total no. of stations
BA	1	100
BG	58	37.91
CZ	4	9.09
ES	5	3.14

Country code	No. of stations	Percentage of total no. of stations
FR	1716	68.31
GR	51	14.09
IT	1	2.27
LV	2	10
MK	7	100
PL	170	98.27
PT	11	11.96
RO	8	100
Total	2034	52.76

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

Quality records missing connection in the Bodies table were been removed from the Waterbase. They will be included after the country will provide such information.

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
Rule	**Combined validation rules**
14	IF NumberOfSamples = 1 THEN (Mean = Minimum = Maximum = Median)
16	IF NumberOfSamples = 0 THEN (AllValueType Is Null)
Rule	**Negative value validation rule**
17	All Values >= 0

A special QA field (QA_LRviolations) has been added to the data table. 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 (25 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 (noted in the Remarks field).

If the record was detected as potential outlier (25 records) the flag (1) have been entered to a special QA field (QA_outlier) added to data table. It is recommended not to use these records in the further analysis until the issue is solved by the data suppliers.