Waterbase – Groundwater
Version 10

Quality control documentation

Waterbase – Groundwater

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.

Goundwater dataset include physical characteristics and proxy pressures information of the groundwater bodies, chemical quality data on selected nutrients and hazardous substances in groundwater, characteristics of sampling sites and data on saltwater intrusions.

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 2010 by ETC/WTR. This database is included in the EEA data service as version 10, and is publicly available. The database and metadata are available at the following URL:
http://www.eea.europa.eu/data-and-maps/data/waterbase-groundwater-6

Waterbase – Groundwater dataset contains five data tables:
- BODIES
- BODIES_MAPPING
- QUALITY
- SALTWATER_INTRUSIONS
- STATIONS

Five types of the tests have been performed on the data tables. Basic tests, Logical rules violation test, Outlier detection, Stations tests, and Valid data type and codes tests.

1. Basic tests

1.1 Summary
1.1.1 Waterbase - Groundwater: BODIES

Country	Number of records								
	from the last delivery					in the ETC working database		Waterbase	
	total	inserted into working database							
		new		redelivered old					
		total	QA issue	total	QA issue	total	QA issue	**total**	QA issue
AL						10		10	
AT	14	0				28		28	
BA	11	8				12		12	
BE	110	0				110		110	
BG	138	0				270		270	
CH	27	5				43		43	
CY	19	16				19		19	
CZ						265		265	
DE						438		438	
DK	385	385				388		388	
EE	5	0				5		5	
ES	334	59				493		493	
FI	62	21				62		62	
FR	1118	0				3073		3073	74
GB						41		41	
GR						362		362	
HR						0		0	
HU						54		54	36
IE	152	2				155		155	3
IS	2	1				3		3	
IT						45		45	
LI	1	0				1		1	
LT						7		7	
LU	10	0		10		11		11	
LV	16	0				20		20	14
ME						10		10	
MK	7	0				7		7	
MT						2		2	2
NL						9		9	
NO						1		1	
PL						173		173	3
PT	83	0				93		93	
RO	25	0				25		25	
RS	10	0				11		11	
SE	42	0				45		45	
SI						11		11	
SK	101	0				111		111	71
TR	8	4				8		8	
Total	2680	501	0	10	0	6421	0	6421	203

1.1.2   Waterbase - Groundwater: BODIES_MAPPING

Country	Number of records
BE	55
BG	69
FI	21
FR	559
HU	18
LU	5
SE	21
Total	748

1.1.3   Waterbase - Groundwater: QUALITY

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

Country	Number of records									
	from the last delivery					in the ETC working database				Waterbase
		inserted into working database				both aggregated and disaggregated		after the final aggregation		
		new		redelivered old						
	total	total	QA issue	total	QA issue	total	QA issue	total	QA issue	total	QA issue
AL						34	1	34	1	34	1
AT	9199	9199	7			419621	96	2982	21	2982	21
BA	1193	1193	0			2366	0	53	0	53	
BE	37667	37667	1			121151	22	2419	22	2419	22
BG	9581	9316	0			23918	17	4918	17	4918	17
CH	21935	10141	0	605	0	26488	0	1052	0	1052	
CY	4962	4962	0			7072	1	135	1	135	1
CZ	41046	41046	0			180372	0	7704	0	7704	
DE	23185	22605	0			88624	0	12142	24	12142	24
DK	494899	476907	0			515138	0	4092	58	4092	58
EE	1323	1303	0			51516	26	724	12	724	12
ES	2856	1191	0	1665	0	4410	0	3837	0	3837	
FI	112	112	2			2127	23	2127	23	2127	23
FR	360333	360331	740			876267	755	24056	0	24056	
GB						24694	1	817	1	817	1
GR						1934	0	1393	0	1393	
HR						0	0	0	0	0	
HU						1006	8	1006	8	1006	8
IE	26633	21480	0	3001	0	29867	0	4690	24	4690	24
IS	70	70	0			185	0	18	0	18	
IT						15912	16	111	0	111	
LI	1269	1227	1			1246	1	61	0	61	
LT	236	236	0			1565	5	489	6	489	6
LU	12	12	0			177	0	87	0	87	
LV	2335	2319	0			3460	1	300	1	300	1
ME						68	8	68	8	68	8
MK						0	0	0	0	0	
MT						862	0	36	0	36	
NL						334	2	334	2	334	2
NO	116	116	0			1311	1	45	0	45	
PL						2453	0	99	0	99	

Country	Number of records											
	from the last delivery					in the ETC working database				Waterbase		
	total	inserted into working database				both aggregated and disaggregated		after the final aggregation				
		new		redelivered old								
		total	QA issue	total	QA issue	total	QA issue	total	QA issue	total	QA issue
PT	1373	1373	0			8202	0	965	0	965	
RO	91	91	0			262	0	262	0	262	
RS	2571	2571	0			6293	0	270	0	270	
SE	900	848	0			1176	1	328	1	328	1
SI	4615	4615	0			36778	1	744	5	744	5
SK	24407	24407	0			60457	5	2172	31	2172	31
TR	447	447	0			467	0	44	0	44	
Total	1073366	1035785	751	5271	0	2517813	991	80614	266	80614	266

1.1.4 Waterbase - Groundwater: SALTWATER_INTRUSIONS

Country	Number of records				
	from the last delivery			in the ETC working database	Waterbase
	total	inserted into working database			
		new	redelivered old		
BG				1	1
CY	7	7		9	9
LV	1	0		1	1
MT				2	2
Total	8	7	0	13	13

1.1.5 Waterbase - Groundwater: STATIONS

Country	Number of records								
	from the last delivery					in the ETC working database		Waterbase	
	total	inserted into ETC working database							
		new		redelivered old					
		total	QA issue	total	QA issue	total	QA issue	total	QA issue
AL						0	0	0	
AT	756	1	0			783	14	783	14
BA	46	13	0			46	14	46	14
BE	381	18	0			436	48	436	48
BG	112	7	0	14	0	394	12	394	17
CH	34	8	0			35	0	35	
CY	84	57	0	3	0	439	207	439	207
CZ	36	36	0			519	22	519	22
DE	848	0	0			1181	223	1181	229
DK	707	707	0			793	86	793	86
EE	7	7	0			718	3	718	3
ES	696	172	0			723	8	723	8
FI						0	0	0	
FR	1693	49	0	9		2941	221	2941	221
GB						988	237	988	237
GR						442	442	442	442
HR						0	0	0	

Country	Number of records									
	total	from the last delivery				in the ETC working database		Waterbase	
		inserted into ETC working database							
		new		redelivered old					
		total	QA issue	total	QA issue	total	QA issue	**total**	QA issue
HU						0	0	0	
IE	217	7	0			229	0	229	9
IS	5	2	0			5	0	5	
IT						3835	804	3835	804
LI	8	0	0	8	0	8	0	8	
LT	59	0	0			265	1	265	1
LU	5	1	0			7	0	7	
LV	139	12	0	3	0	202	0	202	
ME						0	0	0	
MK	35	0	0			35	0	35	
MT						9	0	9	
NL						0	0	0	
NO	78	0	0			83	2	83	2
PL						45	0	45	
PT	349	0	0	1	0	532	104	532	104
RO						0	0	0	
RS	64	0	0			77	0	77	
SE	24	0	0			26	0	26	
SI	1	1	0			77	0	77	
SK	429	6	0			477	17	477	17
TR	87	79	0			91	0	91	
Total	6900	1183	0	38	0	16443	2465	16441	2485

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, Waterbase_GWB_Code

- BODIES_MAPPING: CountryCode, Waterbase_GWB_Code_EIONET, Waterbase_GWB_Code_WFD

- QUALITY: CountryCode, Waterbase_GWB_Code_EIONET, Waterbase_GWB_Code_WFD, Determinand_code, Year, AggregationPeriod

- STATIONS: CountryCode, WISE-SoE_station_ID

- SALTWATER_INTRUSIONS: CountryCode, SALT_Code

Result:
No primary key error has been detected.

1.3 Table relations tests

The unique Waterbase identifiers (Waterbase_GWB_Code, Waterbase_GWB_Code_EIONET, Waterbase_GWB_Code_WFD) are present in each of the data tables. They 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 GW bodies	Percentage of total no. of GW bodies
BA	1	8.33%
BG	48	17.78%
CH	16	37.21%
CZ	48	18.11%
DE	14	3.20%
DK	336	86.60%
ES	10	2.03%
FR	1697	55.22%
GR	51	14.09%
IE	15	9.68%
IS	1	33.33%
IT	1	2.22%
LV	2	10.00%
MK	7	100.00%
PL	170	98.27%
PT	6	6.45%
RO	1	4.00%
SK	29	26.13%
Total	2453	41.05%

1.3.2   Number of GW bodies without any data in the "STATIONS" table by country

Country Code	No. of GW bodies	Percentage of total no. of GW bodies
AL	10	100.00%
BA	1	8.33%
BE	46	41.82%
BG	40	14.81%
CH	16	37.21%
CZ	48	18.11%
DK	336	86.60%
ES	4	0.81%
FI	62	100.00%
FR	2177	70.84%
GB	1	2.44%
GR	130	35.91%
HU	54	100.00%
IE	15	9.68%
IS	1	33.33%
LT	1	14.29%
LU	2	18.18%
LV	6	30.00%
ME	10	100.00%
NL	9	100.00%
PL	170	98.27%
PT	4	4.30%
RO	25	100.00%
SE	3	6.67%
SK	29	26.13%
Total	3200	54.75%

1.3.3 "QUALITY", "STATIONS" and "SALTWATER_INTRUSIONS" table records where none of the GWB identifiers is present in the "BODIES" table

Table	Country Code	No of records	Percentage of total no of records
QUALITY	Total	0	0.00%
STATIONS	BG	5	1.27%
STATIONS	EE	1	0.14%
STATIONS	FR	21	0.71%
STATIONS	GB	29	2.94%
STATIONS	Total	56	1.11%
SALTWATER_INTRUSIONS	Total	0	0.00%

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

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 or only after a careful consideration. 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, both aggregated and disaggregated table.

Measurement values were statistically 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.

Records where Mean value is not provided are also acknowledged as outliers.

The findings described 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 **(8 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 **(255 Quality record)**

10 – the Mean value = 0 **(9 Quality records)**. Value is not correct and records should not be used.

4.   Stations tests

Positions of all reported monitoring stations have been tested using the coordinates provided as well as stations availability. If the station coordinates fall outside the respective country borders or if coordinates are missing, this information is stored in a special QA field (QA_station_problem). The same field was used for documenting situation when link between Stations and Groundwater bodies is missing. Following QA flags have been used:

1 – monitoring station is located outside the respective country borders – either on the sea or in another country **(104 stations)**

2 – coordinates are missing **(2357 stations)**

10 – link to groundwater body is missing (**56 stations)** – see result 1.3.3

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

5. Data type and codes 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: 203 records
Stations table: 20 records

Some of these "errors" are only formally wrong. The current DD definitions of the respective table fields do not cover situation when more than just one code can be selected to properly describe the GWB or monitoring station. It should be solved in the next DD update.