



Waterbase – Emissions to water Version 2

Quality control documentation

28 March 2011

Waterbase – Emissions to water

Data on emissions to water 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.

Dataset contains data selected from reporting of member and collaborating countries on emissions of nutrients and hazardous substances to water, aggregated within River Basin Districts.

QA/QC activities

This document briefly presents the ETC-ICM (former ETC Water) and the EEA activities focused on quality of the Waterbase – Emissions to water dataset and the results of these activities.

The Quality control tests have been performed on the Waterbase – Emissions to water database provided in 19 March 2011 by ETC-ICM. This database is included in the EEA data service as version 2, and is publicly available. The database and metadata are available at the following URL:

<http://www.eea.europa.eu/data-and-maps/data/waterbase-emissions-1>

Waterbase – Emissions to water dataset contains five data tables:

- SPATIAL_UNITS
- NUTRIENTS_POINTS
- NUTRIENTS_DIFFUSE
- HAZSUBS_POINTS
- HAZSUBS_DIFFUSE

Five types of the tests have been performed on the data tables. Mandatory value and Measurement value tests, Primary key/Duplicate tests, Logical rules violation test, Stations tests and Data definition compliance test.

Summary

Abbreviations used:

ADB – delivered record already in database (= redelivery without changes)

DelEx – record deleted by expert

DUPL – duplicated records in 2010 data delivery

Pref – only preferred SoE Haz. Subst. were exported to EU dataset

Waterbase - Emissions: Spatial units

Country	Number of records												
	total	from the last delivery					excluded		in the ETC working database		exported to the EU dataset	excluded from the EU dataset	
		accepted		redelivered			number	reason	total	QA issue		number	reason
		total	QA issue	total	QA issue	total							
AL	0	0	0	0	0	0		0	0	0	0		
AT	0	0	0	0	0	0		8	0	8	0		
BA	0	0	0	0	0	0		0	0	0	0		
BE	33	0	0	2	0	31	DUPL	17	0	17	0		
BG	0	121	0	0	0	0		169	0	169	0		
CH	10	3	0	7	0	0		12	0	12	0		
CY	0	0	0	0	0	0		0	0	0	0		
CZ	3	0	0	3	0	0		3	0	3	0		
DE	0	0	0	0	0	0		0	0	0	0		
DK	0	0	0	0	0	0		0	0	0	0		
EE	8	0	0	8	0	0		8	0	8	0		
ES	0	0	0	0	0	0		0	0	0	0		
FI	0	0	0	0	0	0		8	0	8	0		
FR	33	0	0	33	0	0		33	0	33	0		
GB	0	0	0	0	0	0		0	0	0	0		
GR	0	0	0	0	0	0		0	0	0	0		
HR	0	0	0	0	0	0		0	0	0	0		
HU	0	0	0	0	0	0		0	0	0	0		
IE	0	0	0	0	0	0		0	0	0	0		
IS	1	2	0	1	0	0		3	0	3	0		
IT	0	0	0	0	0	0		0	0	0	0		
LI	0	0	0	0	0	0		0	0	0	0		
LT	0	0	0	0	0	0		5	0	5	0		
LU	0	0	0	0	0	0		0	0	0	0		
LV	0	0	0	0	0	0		4	0	4	0		
ME	0	0	0	0	0	0		0	0	0	0		
MK	1	1	0	0	0	0		1	0	1	0		
MT	0	0	0	0	0	0		0	0	0	0		
NL	0	0	0	0	0	0		0	0	0	0		
NO	0	0	0	0	0	0		0	0	0	0		
PL	0	0	0	0	0	0		0	0	0	0		
PT	0	0	0	0	0	0		0	0	0	0		
RO	1	0	0	1	0	0		1	0	1	0		
RS	0	0	0	0	0	0		0	0	0	0		
SE	5	0	0	5	0	0		5	0	5	0		
SI	158	158	0	0	0	0		158	0	158	0		
SK	2	2	0	0	0	0		2	0	2	0		
TR	0	0	0	0	0	0		0	0	0	0		
Total	255	287	0	60	0	31	0	437	0	437	0		

Waterbase - Emissions: Nutrients - points

Country	Number of records												
	total	from the last delivery						in the ETC working database		exported to the EU dataset	excluded from the EU dataset		
		accepted				excluded		total	QA issue		number	reason	
		new		redelivered		number	reason						
	total	QA issue	total	QA issue	number	reason	total	QA issue	number	reason			
AL	0	0	0	0	0			0	0	0		0	
AT	0	0	0	0	0			180	24	180		0	
BA	0	0	0	0	0			0	0	0		0	
BE	701	25	0	133	0	543	ADB	721	1	721		0	
BG	152	152	0	0	0	0		214	0	214		0	
CH	40	34	28	0	0	6	DUPL	68	28	68		0	
CY	0	0	0	0	0	0		0	0	0		0	
CZ	9	9	0	0	0	0		63	0	63		0	
DE	0	0	0	0	0	0		0	0	0		0	
DK	0	0	0	0	0	0		0	0	0		0	
EE	136	133	14	0	0	0		588	44	588		0	
ES	0	0	0	0	0	0		0	0	0		0	
FI	0	0	0	0	0	0		159	0	159		0	
FR	594	594	0	0	0	0		4355	0	4355		0	
GB	0	0	0	0	0	0		0	0	0		0	
GR	0	0	0	0	0	0		0	0	0		0	
HR	0	0	0	0	0	0		0	0	0		0	
HU	0	0	0	0	0	0		0	0	0		0	
IE	0	0	0	0	0	0		0	0	0		0	
IS	8	8	0	0	0	0		12	0	12		0	
IT	0	0	0	0	0	0		0	0	0		0	
LI	0	0	0	0	0	0		0	0	0		0	
LT	178	178	0	0	0	0		524	0	524		0	
LU	0	0	0	0	0	0		0	0	0		0	
LV	490	133	0	59	0	298	ADB	497	0	497		0	
ME	0	0	0	0	0	0		0	0	0		0	
MK	3	3	0	0	0	0		3	0	3		0	
MT	0	0	0	0	0	0		0	0	0		0	
NL	0	0	0	0	0	0		0	0	0		0	
NO	0	0	0	0	0	0		0	0	0		0	
PL	0	0	0	0	0	0		0	0	0		0	
PT	0	0	0	0	0	0		0	0	0		0	
RO	119	119	0	0	0	0		238	0	238		0	
RS	0	0	0	0	0	0		0	0	0		0	
SE	106	105	0	0	0	0		165	0	165		0	
SI	4619	3868	0	0	0	750	DelEx	4231	0	4231		0	
SK	14	14	0	0	0	0		14	0	14		0	
TR	0	0	0	0	0	0		0	0	0		0	
Total	255	287	0	60	0	31	0	437	0	437		0	

Waterbase - Emissions: Nutrients - diffuse

Country	Number of records											
	total	from the last delivery						in the ETC working database		exported to the EU dataset	excluded from the EU dataset	
		accepted				excluded		total	QA issue		number	reason
		new		redelivered		total	reason					
total	QA issue	total	QA issue	number	reason	total	QA issue	number	reason			
AL	0	0	0	0	0	0		0	0	0	0	
AT	0	0	0	0	0	0		64	0	64	0	
BA	0	0	0	0	0	0		0	0	0	0	
BE	403	28	0	4	0	371	ADB	403	1	403	0	
BG	0	0	0	0	0	0		0	0	0	0	
CH	52	40	0	2	0	10	ADB	94	0	94	0	
CY	0	0	0	0	0	0		0	0	0	0	
CZ	0	0	0	0	0	0		15	0	15	0	
DE	0	0	0	0	0	0		0	0	0	0	
DK	0	0	0	0	0	0		0	0	0	0	
EE	0	0	0	0	0	0		0	0	0	0	
ES	0	0	0	0	0	0		0	0	0	0	
FI	0	0	0	0	0	0		80	0	80	0	
FR	0	0	0	0	0	0		66	0	66	0	
GB	0	0	0	0	0	0		0	0	0	0	
GR	0	0	0	0	0	0		0	0	0	0	
HR	0	0	0	0	0	0		0	0	0	0	
HU	0	0	0	0	0	0		0	0	0	0	
IE	0	0	0	0	0	0		0	0	0	0	
IS	4	4	0	0	0	0		8	0	8	0	
IT	0	0	0	0	0	0		0	0	0	0	
LI	0	0	0	0	0	0		0	0	0	0	
LT	20	20	0	0	0	0		20	0	20	0	
LU	0	0	0	0	0	0		0	0	0	0	
LV	0	0	0	0	0	0		0	0	0	0	
ME	0	0	0	0	0	0		0	0	0	0	
MK	0	0	0	0	0	0		0	0	0	0	
MT	0	0	0	0	0	0		0	0	0	0	
NL	0	0	0	0	0	0		0	0	0	0	
NO	0	0	0	0	0	0		0	0	0	0	
PL	0	0	0	0	0	0		0	0	0	0	
PT	0	0	0	0	0	0		0	0	0	0	
RO	0	0	0	0	0	0		8	0	8	0	
RS	0	0	0	0	0	0		0	0	0	0	
SE	210	0	0	30	0	180	ADB	210	0	210	0	
SI	0	0	0	0	0	0		0	0	0	0	
SK	0	0	0	0	0	0		0	0	0	0	
TR	0	0	0	0	0	0		0	0	0	0	
Total	689	92	0	36	0	561		968	1	968	0	

Waterbase - Emissions: HazSubs - points

Country	Number of records											
	total	from the last delivery						in the ETC working database		exported to the EU dataset	excluded from the EU dataset	
		accepted			excluded			total	QA issue		number	reason
		total	QA issue	redelivered	total	QA issue	number					
AL	0	0	0	0	0	0		0	0	0	0	
AT	0	0	0	0	0	0		674	440	578	96	Pref
BA	0	0	0	0	0	0		0	0	0	0	
BE	960	333	0	89	0	538	ADB	1028	0	873	155	Pref
BG	229	229	0	0	0	0		270	0	270	0	
CH	81	76	76	0	0	5	DUPL	168	76	139	29	Pref
CY	0	0	0	0	0	0		0	0	0	0	
CZ	84	84	0	0	0	0		207	26	204	3	Pref
DE	0	0	0	0	0	0		0	0	0	0	
DK	0	0	0	0	0	0		0	0	0	0	
EE	671	649	25	0	0	0		741	28	741	0	
ES	0	0	0	0	0	0		0	0	0	0	
FI	0	0	0	0	0	0		85	0	85	0	
FR	3399	3399	2144	0	0	0		9950	2144	9190	760	Pref
GB	0	0	0	0	0	0		0	0	0	0	
GR	0	0	0	0	0	0		0	0	0	0	
HR	0	0	0	0	0	0		0	0	0	0	
HU	0	0	0	0	0	0		0	0	0	0	
IE	0	0	0	0	0	0		0	0	0	0	
IS	10	10	0	0	0	0		10	0	9	1	Pref
IT	0	0	0	0	0	0		0	0	0	0	
LI	0	0	0	0	0	0		0	0	0	0	
LT	85	85	0	0	0	0		144	0	103	41	Pref
LU	0	0	0	0	0	0		0	0	0	0	
LV	202	155	1	0	0	47	ADB	290	1	290	0	
ME	0	0	0	0	0	0		0	0	0	0	
MK	0	0	0	0	0	0		0	0	0	0	
MT	0	0	0	0	0	0		0	0	0	0	
NL	0	0	0	0	0	0		0	0	0	0	
NO	0	0	0	0	0	0		0	0	0	0	
PL	0	0	0	0	0	0		0	0	0	0	
PT	0	0	0	0	0	0		0	0	0	0	
RO	119	119	0	0	0	0		238	0	238	0	
RS	0	0	0	0	0	0		0	0	0	0	
SE	396	396	0	0	0	0		456	0	373	83	Pref
SI	1204	1204	0	0	0	0		1204	0	1194	10	Pref
SK	25	25	0	0	0	0		25	0	24	1	Pref
TR	0	0	0	0	0	0		0	0	0	0	
Total	7465	6764	2246	89	0	590		15490	2715	14311	1179	

Waterbase - Emissions: HazSubs - diffuse

Country	Number of records											
	total	from the last delivery						in the ETC working database		exported to the EU dataset	excluded from the EU dataset	
		accepted				excluded		total	QA issue		number	reason
		new		redelivered		number	reason					
	total	QA issue	total	QA issue	number	reason	total	QA issue		number	reason	
AL	0	0	0	0	0		0	0	0		0	
AT	0	0	0	0	0		0	0	0		0	
BA	0	0	0	0	0		0	0	0		0	
BE	520	256	0	12	0	ADB	252	0	520	0	454	Pref
BG	0	0	0	0	0		0	0	0		0	
CH	48	48	0	0	0		0	0	144	0	55	Pref
CY	0	0	0	0	0		0	0	0		0	
CZ	0	0	0	0	0		0	0	0		0	
DE	0	0	0	0	0		0	0	0		0	
DK	0	0	0	0	0		0	0	0		0	
EE	0	0	0	0	0		0	0	0		0	
ES	0	0	0	0	0		0	0	0		0	
FI	0	0	0	0	0		0	0	0		0	
FR	0	0	0	0	0		0	0	0		0	
GB	0	0	0	0	0		0	0	0		0	
GR	0	0	0	0	0		0	0	0		0	
HR	0	0	0	0	0		0	0	0		0	
HU	0	0	0	0	0		0	0	0		0	
IE	0	0	0	0	0		0	0	0		0	
IS	0	0	0	0	0		0	0	0		0	
IT	0	0	0	0	0		0	0	0		0	
LI	0	0	0	0	0		0	0	0		0	
LT	0	0	0	0	0		0	0	0		0	
LU	0	0	0	0	0		0	0	0		0	
LV	0	0	0	0	0		0	0	0		0	
ME	0	0	0	0	0		0	0	0		0	
MK	0	0	0	0	0		0	0	0		0	
MT	0	0	0	0	0		0	0	0		0	
NL	0	0	0	0	0		0	0	0		0	
NO	0	0	0	0	0		0	0	0		0	
PL	0	0	0	0	0		0	0	0		0	
PT	0	0	0	0	0		0	0	0		0	
RO	0	0	0	0	0		0	0	0		0	
RS	0	0	0	0	0		0	0	0		0	
SE	180	180	0	0	0		0	0	260	0	260	
SI	0	0	0	0	0		0	0	0		0	
SK	0	0	0	0	0		0	0	0		0	
TR	0	0	0	0	0		0	0	0		0	
Total	748	484	0	12	0		252	0	924	0	769	116

1. Mandatory values tests

Mandatory values have to be present in each of the records. Records where any of these values is missing are excluded from the dataset:

- SPATIAL_UNITS: Country Code, Spatial Unit Code, Spatial Unit Name, Spatial Unit Category, River Basin District Code or River Basin District Name
- NUTRIENTS_POINTS: Country Code, Spatial Unit Code, Spatial Unit Name, Spatial Unit Category, Period - Years, Nutrient Determinand ID, Source of Point Emissions
- NUTRIENTS_DIFFUSE: Country Code, Spatial Unit Code, Spatial Unit Name, Spatial Unit Category, Period - Years, Nutrient Determinand ID, Source of Diffuse Emissions
- HAZSUBS_POINTS: Country Code, Spatial Unit Code, Spatial Unit Name, Spatial Unit Category, Period - Years, Hazardous Substance Determinand ID or Hazardous Substance Determinand Name, Source of Point Emissions
- HAZSUBS_DIFFUSE: Country Code, Spatial Unit Code, Spatial Unit Name, Spatial Unit Category, Period - Years, Hazardous Substance Determinand ID or Hazardous Substance Determinand Name, Source of Diffuse Emissions

1.1 Measurement value tests

Emission values in all four tables are subject of this test. Detected issues are then stored as a code in a special QA field (QA_MVissues) as follows:

101 – the Emission value is missing

102 – the Emission value is negative and negative values are not allowed or possible

103 – the Emission value is equal 0 and 0 values are not allowed or possible

In a case that Emission values have been provided under WFD, UWWTD, E-PRTR or SoE (Riverine Load or Direct Discharges) reporting, the reporters can use value -1, -2, -3 or -5 respectively instead of the real Emission value. If the real Emissions values could not be provided by checking the dataset of the above mentioned reportings, the reason is indicated in the QA_MVissues field as follows:

161 – the real value of Emissions was checked out, but it is not available in relevant dataset

162 – the relevant dataset containing the real value of Emissions does not exist at all or it is not available

163 – the relevant dataset containing the real value of Emissions was not finished or not provided yet => there is a chance to get the real value in the future

164 – the real value of Emissions was not checked out yet

In addition if a) the value of emissions is not known or not available, or b) emissions are not relevant or not significant, then reporters use value (a) -8 or (b) -9 instead of the real Emission values.

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 or the records are excluded from the dataset.

List of data tables primary keys:

- SPATIAL_UNITS: Country Code, Spatial Unit Code
- NUTRIENTS_POINTS: Country Code, Spatial Unit Code, Period - Years, Nutrient Determinand ID, Source of Point Emissions
- NUTRIENTS_DIFFUSE: Country Code, Spatial Unit Code, Period - Years, Nutrient Determinand ID, Source of Diffuse Emissions
- HAZSUBS_POINTS: Country Code, Spatial Unit Code, Period - Years, Hazardous Substance Determinand ID or Hazardous Substance Determinand Name, Source of Point Emissions, EPRTR Facility
- HAZSUBS_DIFFUSE: Country Code, Spatial Unit Code, Period - Years, Hazardous Substance Determinand ID or Hazardous Substance Determinand Name, Source of Diffuse Emissions

3. Logical rule violation tests

The following logical rules were tested in all emissions data tables:

261 – NP_calculated = NP1_reported + NP2_reported + NP3_reported + NP4_reported + NP5_reported + NP6_reported + NP7_reported + NP8_reported

262 – U1_calculated = U11_reported + U12_reported + U13_reported + U14_reported

263 – U2_calculated = U21_reported + U22_reported + U23_reported + U24_reported

264 – U_calculated = U1_reported + U2_reported

265 – I_calculated = I3_reported + I4_reported

266 – O_calculated = O5_reported + O6_reported

267 – PT_calculated = U_reported + I_reported + O_reported

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 codes (the codes are the same as the numbers of the rules above). It is recommended that the records where QA_LRviolation field is not empty 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.

4. Stations tests

Spatial unit code of each of the records in any of the emission data tables should be present in the Spatial_units table. If it is not true the emission record is flagged in a special QA_field (QA_station_issues) witch code 599.

No such issues were detected in this dataset.

5. Data definition compliance tests

All 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_DDviolations) in the following format:

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

No such issues were detected in this dataset.