



# **Waterbase – Rivers**

## **Version 11**

---

**Quality control documentation**

**11 May 2011**

## Waterbase – Rivers

Data on quality of water in rivers 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 on nutrients, organic matter, hazardous substances and other chemical determinands in water, proxy pressure data on the upstream catchments and physical characteristics of the WISE-SoE river monitoring stations.

## QA/QC activities

This document briefly presents the ETC-ICM (former ETC Water) and the EEA activities focused on quality of the Waterbase – Rivers dataset and the results of these activities. The Quality control tests have been performed on the Waterbase – Rivers database provided in 16 March 2011 by ETC-ICM. This database is included in the EEA data service as version 11, and is publicly available. The database and metadata are available at the following URL: <http://www.eea.europa.eu/data-and-maps/data/waterbase-rivers-7>

Waterbase – Emissions to water dataset contains five data tables:

- STATIONS
- PRESSURES
- NUTRIENTS
- HAZSUBS
- SUPPDET

Following main 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, Outlier detection tests, Chemical rules violation tests, Stations tests and Data definition compliance test.

## Summary

Abbreviations used:

NSD – non-standard determinands

DelEx – record deleted by expert

DUPL – duplicated records in 2010 data delivery

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

### Waterbase - Rivers: Stations

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		total			QA issue	number		reason	number
		total	QA issue	total	QA issue								
AL	46	3	0	43	0	0		55	0	55	0		
AT	71	0	0	71	0	0		290	0	290	0		
BA	46	4	0	42	0	0		60	0	60	0		
BE	63	1	0	62	0	0		67	0	67	0		
BG	94	0	0	94	0	0		111	0	111	0		
CH	25	4	0	21	0	0		26	0	26	0		
CY	0	0	0	0	0	0		33	0	33	0		
CZ	0	0	0	0	0	0		73	0	73	0		
DE	259	1	0	258	0	0		268	1	268	0		
DK	0	0	0	0	0	0		42	0	42	0		
EE	61	0	0	61	0	0		61	0	61	0		
ES	1403	411	0	992	0	0		3240	33	3240	0		
FI	138	0	0	138	0	0		230	0	230	0		
FR	1564	1	0	1563	0	0		1948	0	1948	0		
GB	727	478	0	249	0	0		727	0	727	0		
GR	0	0	0	0	0	0		94	5	94	0		
HR	45	2	0	43	0	0		52	0	52	0		
HU	0	0	0	0	0	0		152	0	152	0		
IE	180	0	0	180	0	0		209	0	209	0		
IS	3	1	0	2	0	0		3	0	3	0		
IT	122	55	7	67	7	0		1490	51	1490	0		
LI	1	0	0	1	0	0		1	0	1	0		
LT	53	0	0	53	0	0		102	0	102	0		
LU	4	0	0	4	0	0		4	0	4	0		
LV	32	32	0	0	0	0		150	0	150	0		
ME	36	36	31	0	31	0		36	31	36	0		
MK	20	0	0	20	0	0		20	1	20	0		
NL	14	0	0	14	0	0		31	0	31	0		
NO	46	0	0	46	0	0		46	0	46	0		
PL	113	38	0	75	0	0		174	0	174	0		
PT	66	7	0	59	0	0		66	4	66	0		
RO	118	0	0	118	0	0		126	0	126	0		
RS	76	0	0	76	0	0		77	0	77	0		
SE	122	0	0	122	0	0		127	0	127	0		
SI	0	0	0	0	0	0		30	0	30	0		
SK	43	1	0	42	0	0		132	3	132	0		
TR	5	0	0	5	0	0		5	0	5	0		
<b>Total</b>	<b>5596</b>	<b>1075</b>	<b>38</b>	<b>4521</b>	<b>38</b>	<b>0</b>		<b>10358</b>	<b>129</b>	<b>10358</b>	<b>0</b>		

**Waterbase - Rivers: Pressures**

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							
AL	44	1	0	43	0	0		52	0	52	0	
AT	0	0	0	0	0	0		141	2	141	0	
BA	0	0	0	0	0	0		0	0	0	0	
BE	0	0	0	0	0	0		59	0	59	0	
BG	0	0	0	0	0	0		0	0	0	0	
CH	25	4	0	21	0	0		26	0	26	0	
CY	0	0	0	0	0	0		32	0	32	0	
CZ	0	0	0	0	0	0		72	0	72	0	
DE	0	0	0	0	0	0		243	1	243	0	
DK	0	0	0	0	0	0		42	0	42	0	
EE	61	0	0	61	0	0		61	0	61	0	
ES	1403	411	0	992	0	0		3240	33	3240	0	
FI	0	0	0	0	0	0		1	0	1	0	
FR	1564	10	0	1554	0	0		1886	0	1886	0	
GB	0	0	0	0	0	0		190	0	190	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		151	0	151	0	
IE	180	0	0	180	0	0		180	0	180	0	
IS	3	1	0	2	0	0		3	0	3	0	
IT	0	0	0	0	0	0		57	0	57	0	
LI	1	0	0	1	0	0		1	0	1	0	
LT	53	0	0	53	0	0		102	0	102	0	
LU	4	1	0	3	0	0		4	0	4	0	
LV	32	32	0	0	0	0		141	0	141	0	
ME	0	0	0	0	0	0		0	0	0	0	
MK	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		46	0	46	0	
PL	113	38	0	75	0	0		174	0	174	0	
PT	66	7	0	59	0	0		66	4	66	0	
RO	0	0	0	0	0	0		1	0	1	0	
RS	0	0	0	0	0	0		0	0	0	0	
SE	122	0	0	122	0	0		127	0	127	0	
SI	0	0	0	0	0	0		24	0	24	0	
SK	43	2	0	41	0	0		131	3	131	0	
TR	0	0	0	0	0	0		0	0	0	0	
<b>Total</b>	<b>3714</b>	<b>507</b>	<b>0</b>	<b>3207</b>	<b>0</b>	<b>0</b>		<b>7253</b>	<b>43</b>	<b>7253</b>	<b>0</b>	

## Waterbase - Rivers: Nutrients

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	517	517	45	0	0	0		3234	601	3234	0	
AT	929	929	2	0	0	0		26980	1270	26980	0	
BA	640	640	4	0	0	0		4153	241	4153	0	
BE	897	598	65	299	17	0		9013	326	9013	0	
BG	1065	1065	101	0	0	0		13356	673	13356	0	
CH	45	45	0	0	0	0		1897	0	1897	0	
CY	481	481	9	0	0	0		1016	93	1016	0	
CZ	0	0	0	0	0	0		13750	67	13750	0	
DE	2889	2889	30	0	0	0		31525	3483	31184	341	NSD
DK	204	204	0	0	0	0		13057	27	13057	0	
EE	895	895	0	0	0	0		13445	6	13445	0	
ES	17727	17727	1300	0	0	0		105175	17198	104020	1155	NSD
FI	5335	5335	0	0	0	0		167459	610	167061	398	NSD
FR	26843	26843	1106	0	0	0		185390	5523	185390	0	
GB	16606	1759	119	14847	326	0		51356	1601	51356	0	
GR	0	0	0	0	0	0		2093	221	2093	0	
HR	767	767	0	0	0	0		7606	212	7606	0	
HU	0	0	0	0	0	0		94587	10227	94587	0	
IE	716	716	0	0	0	0		7325	336	6549	776	NSD
IS	15	15	2	0	0	0		136	11	89	47	NSD
IT	5776	1304	568	4472	201	0		34224	4769	34219	5	NSD
LI	12	12	0	0	0	0		20	0	20	0	
LT	961	961	0	0	0	0		22802	163	22749	53	NSD
LU	27	27	0	0	0	0		381	0	381	0	
LV	479	479	0	0	0	0		11714	39	11714	0	
ME	468	468	382	0	0	0		468	382	432	36	NSD
MK	339	137	7	202	10	0		2558	412	2558	0	
NL	214	214	0	0	0	0		4915	87	4915	0	
NO	230	230	0	0	0	0		3015	16	3015	0	
PL	1363	1363	32	0	0	0		28982	1950	28982	0	
PT	1185	857	10	328	23	0		1910	134	1888	22	NSD
RO	1357	1357	13	0	0	0		8988	69	8978	10	NSD
RS	1265	1265	9	0	0	0		7392	367	7392	0	
SE	1264	1264	40	0	0	0		41076	208	41076	0	
SI	105	105	0	0	0	0		4959	371	4959	0	
SK	859	859	69	0	0	0		10201	582	10178	23	NSD
TR	40	40	3	0	0	0		160	17	160	0	
<b>Total</b>	<b>92515</b>	<b>72367</b>	<b>3916</b>	<b>20148</b>	<b>577</b>	<b>0</b>		<b>936318</b>	<b>52292</b>	<b>933452</b>	<b>2866</b>	

## Waterbase - Rivers: HazSubs

Country	Number of records											
	from the last delivery – aggregated / disaggregated						in the ETC working database (aggregated)		exported to the EU dataset	excluded from the EU dataset		
	total	accepted		excluded		total	QA issue	number		reason		
		new	redelivered	total	QA issue				number		reason	
AL	276 / 0	0 / 0	0 / 0	0 / 0	0 / 0	276 / 0	NUD	0	0	0	0	
AT	0 / 1308	0 / 1308	0 / 0	0 / 0	0 / 0	0 / 0		16163	3105	16126	37	NSD
BA	0 / 2307	0 / 2307	14 / 0	0 / 0	0 / 0	0 / 0		2211	29	1941	270	NSD
BE	0 / 31290	0 / 31290	0 / 0	0 / 0	0 / 0	0 / 0		12303	432	11335	968	NSD
BG	326 / 1751	326 / 1751	5 / 0	0 / 0	0 / 0	0 / 0		1878	239	1878	0	
CH	0 / 10	0 / 10	0 / 0	0 / 0	0 / 0	0 / 0		232	0	80	152	NSD
CY	1187 / 4889	1187 / 4889	0 / 0	0 / 0	0 / 0	0 / 0		2995	0	1635	1360	NSD
CZ	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		7931	1003	7050	881	NSD
DE	1151 / 0	1151 / 0	0 / 0	0 / 0	0 / 0	0 / 0		77237	1370	41383	35854	NSD
DK	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
EE	40 / 142	40 / 142	0 / 0	0 / 0	0 / 0	0 / 0		1972	0	81	1891	NSD
ES	7324 / 0	7324 / 0	2566 / 0	0 / 0	0 / 0	0 / 0		7324	2566	6964	360	NSD
FI	0 / 18192	0 / 18192	0 / 0	0 / 0	0 / 0	0 / 0		28418	193	25896	2522	NSD
FR	0 / 591475	0 / 591475	0 / 0	0 / 0	0 / 0	0 / 0		906544	12961	266898	639646	NSD
GB	1019 / 141952	1019 / 53840	0 / 0	0 / 88112	0 / 0	0 / 0		65840	21548	44282	21558	NSD
GR	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		2668	266	1987	681	NSD
HR	0 / 15750	0 / 15750	0 / 0	0 / 0	0 / 0	0 / 0		4332	148	3685	647	NSD
HU	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		17159	573	16825	334	NSD
IE	0 / 74126	0 / 74126	0 / 0	0 / 0	0 / 0	0 / 0		16141	10	10654	5487	NSD
IS	0 / 84	0 / 84	0 / 0	0 / 0	0 / 0	0 / 0		77	0	25	52	NSD
IT	2910 / 36103	2910 / 36103	0 / 0	0 / 0	0 / 0	0 / 0		65061	5000	41735	23326	NSD
LI	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
LT	0 / 4373	0 / 4373	0 / 0	0 / 0	0 / 0	0 / 0		7891	384	6398	1493	NSD
LU	54 / 1460	54 / 1460	0 / 0	0 / 0	0 / 0	0 / 0		787	0	327	460	NSD
LV	373 / 1334	373 / 1334	0 / 0	0 / 0	0 / 0	0 / 0		3174	169	2889	285	NSD
ME	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
MK	140 / 1520	140 / 1520	5 / 0	0 / 0	0 / 0	0 / 0		1370	62	1310	60	NSD
NL	0 / 34356	0 / 34356	21 / 0	0 / 0	0 / 0	0 / 0		8502	21	4401	4101	NSD
NO	230 / 1360	230 / 1360	35 / 0	0 / 0	0 / 0	0 / 0		2589	35	2192	397	NSD
PL	1708 / 9308	1708 / 9308	35 / 0	0 / 0	0 / 0	0 / 0		6123	968	3604	2519	NSD
PT	1031 / 0	1009 / 0	0 / 0	22 / 0	0 / 0	0 / 0		1661	116	1512	149	NSD
RO	832 / 0	832 / 0	5 / 0	0 / 0	0 / 0	0 / 0		2042	17	1428	614	NSD
RS	0 / 10442	0 / 10442	0 / 0	0 / 0	0 / 0	0 / 0		8510	0	5298	3212	NSD
SE	0 / 12709	0 / 12709	0 / 0	0 / 0	0 / 0	0 / 0		8588	62	7714	874	NSD
SI	0 / 949	0 / 949	0 / 0	0 / 0	0 / 0	0 / 0		10354	594	7925	2429	NSD
SK	3731 / 0	3731 / 0	711 / 0	0 / 0	0 / 0	0 / 0		12162	2010	6417	5745	NSD
TR	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
<b>Total</b>	<b>22332 / 997190</b>	<b>22034 / 909078</b>	<b>3397 / 0</b>	<b>22 / 88112</b>	<b>0 / 0</b>	<b>276 / 0</b>		<b>1310239</b>	<b>53881</b>	<b>551875</b>	<b>758364</b>	

## Waterbase - Rivers: SuppDet

Country	Number of records											
	from the last delivery – aggregated / disaggregated						in the ETC working database (aggregated)		exported to the EU dataset	excluded from the EU dataset		
	total	accepted		redelivered		excluded		total		QA issue	number	reason
		total	QA issue	total	QA issue	number	reason					
AL	129 / 0	129 / 0	0 / 0	0 / 0	0 / 0	0 / 0		129	0	129	0	
AT	0 / 9590	0 / 9590	0 / 0	0 / 0	0 / 0	0 / 0		726	0	726	0	
BA	0 / 607	0 / 607	0 / 0	0 / 0	0 / 0	0 / 0		76	0	76	0	
BE	0 / 2368	0 / 2368	0 / 0	0 / 0	0 / 0	0 / 0		217	0	217	0	
BG	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
CH	0 / 14	0 / 14	0 / 0	0 / 0	0 / 0	0 / 0		7	0	7	0	
CY	0 / 1063	0 / 1063	0 / 0	0 / 0	0 / 0	0 / 0		224	0	224	0	
CZ	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
DE	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
DK	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
EE	0 / 916	0 / 916	0 / 0	0 / 0	0 / 0	0 / 0		86	0	86	0	
ES	0 / 8861	0 / 8861	0 / 0	0 / 0	0 / 0	0 / 0		1555	0	1555	0	
FI	0 / 6029	0 / 6029	0 / 0	0 / 0	0 / 0	0 / 0		672	0	672	0	
FR	0 / 63882	0 / 63882	0 / 0	0 / 0	0 / 0	0 / 0		11320	1	11320	0	
GB	0 / 5560	0 / 5560	0 / 0	0 / 0	0 / 0	0 / 0		486	2	486	0	
GR	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
HR	0 / 3674	0 / 3674	0 / 0	0 / 0	0 / 0	0 / 0		339	0	339	0	
HU	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
IE	0 / 14396	0 / 14396	0 / 0	0 / 0	0 / 0	0 / 0		1091	0	1091	0	
IS	0 / 96	0 / 96	0 / 0	0 / 0	0 / 0	0 / 0		24	0	24	0	
IT	0 / 2137	0 / 2137	0 / 0	0 / 0	0 / 0	0 / 0		280	0	280	0	
LI	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
LT	0 / 1333	0 / 1333	0 / 0	0 / 0	0 / 0	0 / 0		112	0	105	7	NSD
LU	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
LV	0 / 304	0 / 304	0 / 0	0 / 0	0 / 0	0 / 0		86	0	86	0	
ME	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
MK	0 / 2198	0 / 2198	0 / 0	0 / 0	0 / 0	0 / 0		200	0	200	0	
NL	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
NO	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
PL	0 / 7594	0 / 7594	0 / 0	0 / 0	0 / 0	0 / 0		778	0	778	0	
PT	0 / 1734	0 / 1734	0 / 0	0 / 0	0 / 0	0 / 0		196	0	196	0	
RO	10 / 9002	0 / 8992	0 / 0	10 / 10	0 / 0	0 / 0		983	0	983	0	
RS	0 / 8550	0 / 8550	0 / 0	0 / 0	0 / 0	0 / 0		834	0	834	0	
SE	0 / 9601	0 / 9601	0 / 0	0 / 0	0 / 0	0 / 0		722	0	722	0	
SI	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
SK	8 / 4207	0 / 4111	0 / 0	8 / 96	0 / 0	0 / 0		354	0	354	0	
TR	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0		0	0	0	0	
<b>Total</b>	<b>147 / 163716</b>	<b>129 / 163610</b>	<b>0 / 0</b>	<b>18 / 106</b>	<b>0 / 0</b>	<b>0 / 0</b>		<b>21497</b>	<b>3</b>	<b>21490</b>	<b>7</b>	

## 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:

- STATIONS: Country Code, NationalStationID
- PRESSURES: Country Code, NationalStationID
- NUTRIENTS: Country Code, NationalStationID, Year, Aggregation Period, Determinand Nutrients
- HAZSUBS: Country Code, NationalStationID, Year, Determinand HazSubs
- SUPPDET: Country Code, NationalStationID, Year, Determinand Suportive

### 1.1 Measurement value tests

Mean values in all three tables containing the determinand concentrations are subject of this test. Detected issues are then stored as a code in a special QA field (QA\_MVissues) as follows:

101 – the Mean value is missing

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

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

Records flagged with any of these flags either can't be used (101) or it is recommended that they are excluded from further use or analysis (102, 103).

## 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:

- STATIONS: Country Code, NationalStationID
- PRESSURES: Country Code, NationalStationID
- NUTRIENTS: Country Code, NationalStationID, Year, Aggregation Period, Determinand Nutrients
- HAZSUBS: Country Code, NationalStationID, Year, Determinand HazSubs
- SUPPDET: Country Code, NationalStationID, Year, Determinand Suportive

### 3. Logical rules violation tests

The following logical rules were tested in “NUTRIENTS” and “HAZSUBS” data tables:

201 – Mean  $\geq$  Minimum

202 – Mean  $\leq$  Maximum

203 – Median  $\geq$  Minimum

204 – Median  $\leq$  Maximum

205 – Minimum  $\leq$  Maximum

206 – If Minimum  $> 0$  Then StandardDeviation  $<$  Maximum

207 – If Minimum  $<$  Maximum Then StandardDeviation  $> 0$

210 – All measurement values  $\geq 0$  (exceptions: Alkalinity, Temperature)

211 – If NumberOfSamples = 1 Then (Mean = Minimum = Maximum = Median)

212 – If NumberOfSamples = 1 Then StandardDeviation = 0

213 – If NumberOfSamples = 0 Then all measurement values are Null

217 – NumberOfSamplesBelowLOQ  $\leq$  NumberOfSamples

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 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. Chemical rules violation tests

Chemical rules were tested in the “NUTRIENTS” data table. Following chemical rules were defined between Mean concentrations of certain related determinands from the same monitoring station, year and aggregation period:

310 – COD \* 1.01 > BOD5

320 – Total Phosphorus \* 1.01 >= Orthophosphate

330 – Total Nitrogen \* 1.05 >= Kjeldahl Nitrogen + Nitrate + Nitrite

340 – Kjeldahl Nitrogen \* 1.05 >= Organic Nitrogen + Total Ammonium

350 – Total Oxidised Nitrogen \* 1.05 >= Nitrate + Nitrite

The 1.01 and 1.05 tolerance was applied in order to avoid situations when records are detected only due to value differences caused by rounding.

Due to the fact that the concentration values in the tested records are results of aggregation and it is possible that the individual measurements participating in the aggregation could be taken for different determinands in different times and/or by different methods, thus the discrepancies in concentrations are justified because the particular records are not comparable, following additional testing steps were taken to filter out incomparable records.

1 – comparison of Number of Samples

2 – comparison of Aggregation Length and Aggregation Months values

3 – comparison of Limit of Detection and Limit of Quantification values

4 – considering clarifications and approvals of chemical rules violations that were provided by data reporters

The details about the individual tests can be found at <https://taskman.eionet.europa.eu/ETCW/ticket/86#comment:7>.

A special QA field (QA\_CRviolations) has been added to the data table. Information of the rules violated in the respective records (all records of each of the determinands from both sides of formula) are kept there as a coma separated list of those rules numbers (the numbers are the same as in the table above).

Records that were detected by the first step of the test but have been found incomparable in any of the further steps are also flagged with the code of the test but “0” is replaced by the number of the step as used above (e.g 311, 324, 332). Such records are then treated as not violating the chemical rules.

It is recommended that the records flagged in the QA\_CRviolations with the any of the codes ending with “0” are not used in further analysis or only after a careful consideration. The detected data quality inconsistencies will be tried to be solved in the near future.

## 5. Outlier detection tests

Detection of outliers was performed on the Mean values in the “Nutrients” and “SuppDet” data table.

Different methods of outlier detection were used, from simple comparison of measurement value with the defined limit value for particular determinand, to more complex statistical tests.

Sometime the 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.

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

A special QA field (QA\_outlier) has been added to the tables and records, where the any of the situations mentioned above has been detected, have been flagged in this field as follows:

401 – 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 is marked as an outlier by a content expert

402 – Measurements are probably taken from a highly polluted locations but information was not confirmed

403 – The whole country delivery is considered as problematic because it contains too many quality issues

491 – Outlier has been confirmed by country as correct value

492 – Outlier has been confirmed by (ETC) content expert as correct value

493 – Measurement has been confirmed by country to be taken from a highly polluted area

It is recommended that the records where QA\_outlier field contains codes 401-403 and eventually also code 493, 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.

## 6. Stations tests

A number issues in stations records and in related records in other tables are checked in these tests. A special QA\_field (QA\_station\_issues) was added to all data tables where these issues, if detected, are indicated by appropriate flag as follows:

500 station coordinates fall slightly outside the respective country boundary, but were confirmed as correct by country

501 station coordinates fall outside the respective country boundary

502 one or both station coordinates are missing

503 more stations with the same coordinates (if it might indicate an error)

511 Water Category value is incompatible with this particular dataset

512 station coordinates fall outside the respective River Basin District

513 Catchment Area is suspicious or illogical

599 station is not defined in the station table

These issues should be taken into the account in further use and analysis of the data. The detected data quality inconsistencies will be tried to be solved in the near future.

## 7. 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 than 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; ...]*

This field serves as an indication why some of the values are missing, as a reference for solving similar problems in the future or in certain cases as background information for future update of Data dictionary.