

Background document on specifications for reporting geographical data under UWWTD¹

Drafted by the EC DG Environment contractor -

umweltbundesamt^U

(FINAL version 08/05/2007)

Introduction

The aim of this document is to explain needs and specifications for spatial data in relation to reporting under UWWTD Directive (91/271/EEC). There is a need to obtain spatial data systematically, to improve the positional accuracy, the topological correctness and furthermore to harmonise the format of the data across the EU. Spatial data is of crucial importance to produce high quality maps which are in turn essential for presentational purposes and to make accurate GIS analyses.

The recommendations for spatial data handling in this document are based on the Guidance document N 9 of the Common Implementation Strategy for the Water Framework Directive (2000/60/EC) – Implementing the Geographical Information System Elements (GIS).

Overview Topics

The aim is to harmonise geographical data in relation to reporting under UWWTD. Geographical data should also be in line with future reporting in WISE. The following aspects are considered:

1. Coordinate reference system
2. Data accuracy
3. Format / structure / geometry of datasets -> Data Dictionary
4. Data quality
5. Data documentation (Metadata)

1. Coordinate reference system

The use of a common coordinate reference system is particularly important for the harmonisation of geographic information across Europe. Spatial data provided by Member States are often insufficiently documented (e.g., the used projection is unknown or is only partially described). The use of different coordinate systems is a potential source of errors when national data are imported into a European wide system. **Therefore it is important that Member States shall provide data for reporting under a single coordinate system.**

The ETRS89 (European Terrestrial Reference System) is recommended by the WFD GIS Working Group for pan Europe spatial data collection, storage and analyses.

¹ The document is available on WFD CIRCA at:
http://circa.europa.eu/Public/irc/env/wfd/library?l=/framework_directive/treatment_directive/00-uwtd_questionnaire/02-manual_supporting

Therefore it is recommended to use ETRS 89 for the national point feature and area feature datasets regarding the UWWTD Directive.

The description and definition of ETRS89 is based on the convention of ISO19111, the 'Spatial referencing by coordinates' standard. For further documentation on ETRS89, see <http://crs.bkg.bund.de/crs-eu/> and <http://www.eionet.eu.int/gis>.

2. Data accuracy

Scale and Positional Accuracy

The factor determining the required detail of spatial data is the size of the smallest feature to be shown on digital and paper maps.

This 'spatial detail' determines both the minimum mapping area and the number of coordinates used to describe an element. For example on a large scale map (1:250.000) a river is presented with more points than on a small scale map (1:1.000.000) where small meanders may not be represented.

For the WFD the level of spatial detail is set to 0.5 km² surface for lakes and to a 10 km² catchment area for rivers that leads to a scale of 1:250.000.

The GIS Working Group recommends for WFD needs a positional accuracy for the reporting of 1000 metres (corresponding to an input scale of approximately 1:1.000.000) in the short term. In the long term a positional accuracy of 125 m (1:250.000) is strongly recommended for higher positional data accuracy for web mapping purposes. The scale of 1:250,000 is preferable/advisable to use.

As for the scale and positional accuracy of GIS UWWTD data concerning point features (e.g. agglomerations, UWWTP's and discharge points), sensitive area features (e.g. catchments, lakes, rivers) the accuracy of 125 m (1:250.000) is strongly recommended as a minimum.

Data precision: coordinates – in decimal degrees in 6 digits behind the comma.

3. Format / Coding / Geometry of Datasets

Short term data exchange

The "**shape**" file (**ESRI Standard**) is recommended by the WFD GIS working group as data exchange format for vector data because it is compatible with systems operated by the Commission. This vector data format stores the location, shape and attributes of geographic features by sequences of points and supports the features types of points, lines and polygons.

Geometric data will be reported to the Commission in the so-called "Shape File" format. This format is placed in the public domain and therefore most GIS software packages are able to generate files in this format.

The format consists of the following files:

- shp – file (*.shp): File contains geometry of the geographic features. This is a direct access, variable record length file in which each record describes a shape with a list of its vertices. *[Note: xml schemas will be developed in the frame of WISE]*
- Attribute file (*.dbf): This contains features attributes with one record per feature. The one to one relationship between geometry and attribute is based on record number. It is a Dbase IV format file which can be read by most software packages.
- Projection file (*.prj): detailed description of the geographic reference system and should be included in each spatial dataset delivered to the Commission.

These three files must be zipped (compressed) into 1 file before transfer to the Commission. The latter is useful since a shape file can only be read if the three files are being provided.

The proposal for UWWTD reporting is to use the shape file format and apply state of the art geographic information technology focussing on accessing geographic data through internet browsers directly from the Member States.

Warning: The shapefile format proved to be limitative for the transfer of NON ASCII characters (for example Cyrillic, Greek but also an umlaut or accent grave).

The .dbf component (shapefile Dbase IV format) depends for its character set representation on the code page of the PC where it is made. Therefore the topographic names stored in the shapefile might be corrupted upon transfer to the Commission. In order to avoid the problem Commission asks ASCII representation of the names in addition to original language names. As for solving this problem the Commission recently proposed to do this with 3 columns regarding names:

- Name in original language
- Indication of the language
- Name in ASCII characters

There is still a discussion on this issue within the WISE technical group.

Coding

Concise coding will ease up electronic data transfer in the future and allow linkage to additional datasets that use the same coding mechanism. GIS feature coding is the assignment of unique identification codes to each table or spatial feature that will be referenced by GIS. This assignment needs to be managed to ensure uniqueness at national and international level. Standard code formats will ease electronic data transfer and enhance the possibility of central querying against distributed storage.

Unique European codes should be generated by placing the ISO 2 character national code for each Member State in front of up to 22 characters unique identifier codes generated within Member States.

Coding of point information

The Urban Waste Water Directive leads to the establishment of 3 features requiring unique coding at European Level. These features are:

- Agglomerations
- Waste Water Treatment Plants
- Discharge Points

By specifying a code one should ensure the persistence of the code in the future. Therefore it is recommended to use upper case and **characters from A to Z and figures from 0 to 9 only**. Codes for selected features (rivers, lakes etc) are presumed to be consistent with previously reported coding systems for the Water Framework directives Article 3 and 5 to be in line with further reporting within WISE.

Unique European codes are provided by following format

MS #1#2...#22 where:

MS = 2 character Member State identifier, in accordance with ISO 3166-1-Alpha-2 country codes, and

#1#2...#22 = an up to 22 character feature code that is unique within the Member State
symbol # = wildcard character (a wildcard character can be used to substitute for any other character or characters in a string).

Coding of sensitive areas and less sensitive areas

The feature code for **sensitive areas** should be unique also among the different types of sensitive areas. Therefore the following format for the sensitive areas code is recommended:

MS SA #1#2....#22 where:

MS = 2 character Member State identifier, in accordance with ISO 3166-1-Alpha-2 country codes, and

SA = a 2 character code for the sensitive area (see below)

#1#2...#22 = an up to 22 character feature code that is unique within the Member State
symbol # = wildcard character (a wildcard character can be used to substitute for any other character or characters in a string).

The following 2 character unique identifiers for sensitive areas are recommended:

RI	for Sensitive Area – river
LK	for Sensitive Area – lake
CL	for Sensitive Area – coastline
CA	for Sensitive Area – coast area
CM	for Sensitive Area – catchment
LS	for Less sensitive area (coastline)
TW	for Sensitive Areas and Less Sensitive Areas – transitional waters*

*Definitions:

Transitional waters - WFD Article 2(6): are bodies of surface water in the vicinity of river mouths which are partly saline in character as a result of their proximity to coastal waters but which are substantially influenced by freshwater flows. Transitional waters are for example estuaries and lagoons.

Estuary: is a semi-enclosed body of water where fresh water from the land, usually from a river, meets salt water from the sea. In the sense of UWWTD (Art. 2.12) an “estuary” means the transitional area at the mouth of a river between fresh-water and coastal waters.

Lagoon (tropical): is bordered by land on at least one side and is blocked from the sea by sandbars or coral reefs. Lagoons are typically shallow and do not have a major point source of fresh water input, such as a river, but do collect fresh water as it is discharged from the land through storm water runoff. The salinity of a lagoon depends on; seasonal rainfall, fresh water flowing from the land, evaporation, and the exposure to sea water.

Examples how to use unique identifiers:

A sensitive area lake in Italy might have the identifier	ITLK45734
or an agglomeration in Spain might have the identifier	ES6712454212145

Use of the MS #1#2....#22 and MS SA #1#2....#22 is the only requirement for unique European feature identification codes. Codes of this format should be used for initial and subsequent references to features reporting to the Commission.

Underscores (“_”) must not be used directly after the Member State identifier (MS). This could result in difficulties when creating European codes by placing the ISO 2 character national code for each Member state in front of up to 22 characters unique identifier codes generated within Member States. However, underscores can be used as part of the 22 character feature code (e.g. an agglomeration in Spain might have the identifier ESAG_67124542145)

Geometry Types/Topological rules

The spatial datasets delivered to the Commission concerning the UWWTD Directive should be represented as following regarding the geometry types:

- ✓ **Point features: Agglomerations/UWWTP/discharge points in the ETRS 89 reference system delivered as table in the UWWTD-Questionnaire. Coordinates should be based on ISO**

6709:1983 (Standard representation of longitude, latitude and altitude for geographic point locations).

- ✓ **Line features: Sensitive areas – rivers, - coastline and less sensitive areas in the ETRS 89 reference system delivered as shape – file** (the corresponding shape-files will be uploaded via the electronic UWWTD questionnaire to the same physical location as the export files of the questionnaire. Tabular information on sensitive areas in the questionnaire has to refer to the corresponding name and location of the files)

Topological rules [the explanations to be developed in case it is needed]

Rivers:

- Must not have dangles (exceptions are sources and mouth)
- Must not overlap
- Must not intersect (nodes at intersections)
- Must not touch interior
- Must be covered by boundary of river basins
- Must not overlap with coastal waters, transitional waters
- Outlet of each feature must touch coastline

Coastline/less sensitive areas:

- Must not overlap
- Must not overlap with transitional waters, rivers, lakes
- Must touch transitional waters, national boundaries, river basins

- ✓ **Polygon features: Sensitive areas – lakes, – coast area and catchments in the ETRS 89 reference system delivered as shape file** (the corresponding shape-files will be uploaded via the electronic UWWTD questionnaire to the same physical location as the export files of the questionnaire. Tabular information on sensitive areas in the questionnaire has to refer to the corresponding name and location of the files)

Topological rules [the explanations to be developed in case this is needed]

Lakes:

- Must not overlap
- Must not overlap with coastal waters, transitional waters
- Must be covered by river basin districts

Coast area:

- Must not overlap
- Must not have gaps
- Must not overlap with transitional waters, rivers, lakes
- Must touch transitional waters, national boundaries, river

Catchments:

- Must not overlap
- Must not have gaps
- Must be covered by extent of river basin districts
- Must not overlap with coastal waters
- Must not overlap with transitional waters
- Boundary must be covered by river sub basins
- Must cover features of river sub basins
- Must contain at least one river
- Must touch the coastline

Data Dictionary

According to the WFD CIS Guidance Document the aim of a Data Dictionary is to provide a common understanding of the file / table structure that should be used for the report of both tabular data and geographic datasets to the Commission.

Note: Further developments of WISE-GIS guidance have to be considered here.

The recommended file names and geometry of the UWWTD tabular data and geographic data are listed in the Table 1 below.

Table 1

UWWTD datasets	Recommended File Name	Geometry Type
Agglomeration	MS*_Agglo	point/table
UWWTP	MS*_UWWTP	Point/table
Discharge Point	MS*_DisPoint	Point/table
Sensitive Area - river	MS*_SA_river	Line/shape
Sensitive Area - lake	MS*_SA_lake	Polygon/shape
Sensitive Area – coastline	MS*_SA_coastl	Line/shape
Sensitive Area – coast area	MS*_SA_coasta	Polygon/shape
Sensitive Area – transitional water	MS*_SA_TW	Polygon/shape
Sensitive Area - catchment	MS*_SA_catchm	Polygon/shape
Less sensitive Area – transitional water	MS*_LSA_TW	Polygon/shape
Less sensitive area - coastline	MS*_LSA_coastl	Line/shape

*MS = 2 character Member State identifier (for example UK_SA_river = Sensitive area river of United Kingdom)

The recommended attributes for each UWWTD geographic shape file are described in Tables 2 to 7. For each attribute a field name (shortened due to physical restriction on field name length e.g. dbase – 10 character), a text description, a generic field type and length, together with any restrictions (whether mandatory or optional) are provided.

Remark: The tables for point information (agglomeration, UWWTPs and discharge points) are covered by the electronic UWWTD-Questionnaire 2007 and therefore, not further explained in this document.

Table 2: MS*_SA_river

MS*_SA_river						
Attribute	FieldName	Definition	Values	Type	Length	Restrictions
shape	SHAPE			Geometry (line)		
Name	NAME	Locally used name		String	100	Mandatory**
MemberStateSARiverCode	MSCD_SA_RV	Unique code for Receiving area at Member State level		String	100	Mandatory**, as per coding guidelines
EuropeanSARiverCode	EUCD_SA_RV	Unique code for Receiving area at EU level	ISO3166 & MSCD SA_RV	String	26	Mandatory**, as per coding guidelines
Length	LENGTHKM	Total length of the River		Double	9,2	Optional**

**mandatory' and 'optional' in technical terms in sense of GIS data

Table 3: MS* SA lake

MS* SA lake						
Attribute	FieldName	Definition	Values	Type	Length	Restrictions
shape	SHAPE			Geometry (polygon)		
Name	NAME	Locally used name		String	100	Mandatory**
EuropeanSALakeCode	EUCD_SA_LK	Unique code for Receiving area at EU level	ISO3166 & MSCD_SA_LK	String	26	Mandatory**, as per coding guidelines
MemberStateSALakeCode	MSCD_SA_LK	Unique code for Receiving area at Member State level		String	100	Mandatory**, as per coding guidelines
AreaKM2	AREAKM2	Area in square kilometers		Double	9,2	Optional**

**mandatory' and 'optional' in technical terms in sense of GIS data

Table 4: MS* SA coastl

MS* SA coastl						
Attribute	FieldName	Definition	Values	Type	Length	Restrictions
shape	SHAPE			Geometry (line)		
Name	NAME	Locally used name		String	100	Mandatory**
EuropeanSACoastlineCode	EUCD_SA_CL	Unique code for Receiving area at EU level	ISO3166 & MSCD_SA_CO	String	26	Mandatory**, as per coding guidelines
MemberStateSACoastlineCode	MSCD_SA_CL	Unique code for Receiving area at Member State level		String	100	Mandatory**, as per coding guidelines
Length	LENGTHKM	Total length of the coastline		Double	9,2	Optional**

**mandatory' and 'optional' in technical terms in sense of GIS data

Table 5: MS* SA coasta

MS* SA coasta						
Attribute	FieldName	Definition	Values	Type	Length	Restrictions
shape	SHAPE			Geometry (polygon)		
Name	NAME	Locally used name		String	100	Mandatory**
EuropeanSACoastareaCode	EUCD_SA_CA	Unique code for Receiving area at EU level	ISO3166 & MSCD_SA_CA	String	26	Mandatory**, as per coding guidelines
MemberStateSACoastareaCode	MSCD_SA_CA	Unique code for Receiving area at Member State level		String	100	Mandatory**, as per coding guidelines
AreaKM2	AREAKM2	Area in square kilometers		Double	9,2	Mandatory**

**mandatory' and 'optional' in technical terms in sense of GIS data

Table 6: MS* SA_TW

MS* SA_TW						
Attribute	FieldName	Definition	Values	Type	Length	Restrictions
shape	SHAPE			Geometry (polygon)		
Name	NAME	Locally used name		String	100	Mandatory**
EuropeanSATransitionalwaterCode	EUCD_SA_TW	Unique code for Receiving area at EU level	ISO3166 & MSCD_SA_TW	String	26	Mandatory**, as per coding guidelines
MemberStateEuropeanSATransitionalwaterCode	MSCD_SA_TW	Unique code for Receiving area at Member State level		String	100	Mandatory**, as per coding guidelines
AreaKM2	AREAKM2	Area in square kilometers		Double	9,2	Mandatory**

**mandatory' and 'optional' in technical terms in sense of GIS data

Table 7: MS*_SA_catchm

MS*_SA_catchm						
Attribute	FieldName	Definition	Values	Type	Length	Restrictions
shape	SHAPE			Geometry (polygon)		
Name	NAME	Locally used name		String	100	Mandatory**
EuropeanSACatchmentCode	EUCD_SA_CM	Unique code for Receiving area at EU level	ISO3166 & MSCD_SA_CM	String	26	Mandatory**, as per coding guidelines
MemberStateSACatchmentCode	MSCD_SA_CM	Unique code for Receiving area at Member State level		String	100	Mandatory**, as per coding guidelines
AreaKM2	AREAKM2	Area in square kilometers		Double	9,2	Mandatory**

**mandatory' and 'optional' in technical terms in sense of GIS data

Table 8: MS*_LSA_TW

MS*_LSA_TW						
Attribute	FieldName	Definition	Values	Type	Length	Restrictions
shape	SHAPE			Geometry (polygon)		
Name	NAME	Locally used name		String	100	Mandatory**
EuropeanLSATransitionalwaterCode	EUCD_LSATW	Unique code for Receiving area at EU level	ISO3166 & MSCD_LSATW	String	26	Mandatory**, as per coding guidelines
MemberStateLSATransitionalwaterCode	MSCD_LSATW	Unique code for Receiving area at Member State level		String	100	Mandatory* as per coding guidelines
AreaKM2	AREAKM2	Area in square kilometers		Double	9,2	Mandatory**

**mandatory' and 'optional' in technical terms in sense of GIS data

Table 9: MS*_LSA_coastl

MS*_LSA_coastl						
Attribute	FieldName	Definition	Values	Type	Length	Restrictions
shape	SHAPE			Geometry (line)		
Name	NAME	Locally used name		String	100	Mandatory**
EuropeanLSACoastlineCode	EUCD_LSACL	Unique code for Receiving area at EU level	ISO3166 & MSCD_LSACL	String	26	Mandatory**, as per coding guidelines
MemberStateLSACoastlineCode	MSCD_LSACL	Unique code for Receiving area at Member State level		String	100	Mandatory**, as per coding guidelines
AreaKM2	AREAKM2	Area in square kilometers		Double	9,2	Mandatory**

**mandatory' and 'optional' in technical terms in sense of GIS data

4. Data quality

It is essential that data delivered to the Commission are homogeneous and of general good quality. Quality control includes the following questions:

- Do values and codes in tabular data are within the range defined in guideline documents?
- Are 'IDs' in accordance with the given standards?
- Does the geometry, either being points, lines or polygons, have an acceptable structure and topology, e.g. are polygons closed, do they overlap, do they have gaps and lines forming networks linked properly to nodes?
- Do the geometric accuracy, coordinate reference system, projection and file format follow specifications?

Has metadata been filled in, including aspects of accuracy, coordinate system, methodology and source? Quality control must be carried out and documented before uploaded to the questionnaire's system.

5. Data documentation (Metadata)

Metadata is the information and documentation, which makes data understandable and shareable for users over time.

The WFD GIS working group recommends the application of the rules laid down in **ISO 19115** for creating a metadata profile.

Some of the main aspects covered in a metadataset are: name of the data, description of data, coordinate reference system, source and methodology – including version of specification on which the compilation is based, responsible party, ownership, user rights, process steps and dataset accuracy

The WISE metadata profile applies to spatial datasets, dataset series and services, and may be applicable to other resource types. It defines the minimum requirements for the functional concepts of discovery and data usage. The proposal is in line with the draft implementing rules for metadata of the INSPIRE initiative. The profile is mainly based on the guidelines for metadata included within the document "Guidance Document on Implementing the GIS Elements of the Water Framework Directive".

For UWWTD reporting it is strongly recommended to use the basic elements from this form and to integrate this information as part of the UWWTD Questionnaire.

Table 10: Dataset metadata form mapped to ISO19115

Label	Name	Example	Value
Metadata (information on metadata)	MD_Metadata		
CharacterSet	characterSet	007 (ISO/IEC 8859-1)	
Language	language	ENG	
Standard Name	metadataStandardName	ISO 19115:2003	
Standard Version	metadataStandardVersion	WISE 1.0	
Creation Date	dateStamp	2006-01-29	
Contact	contact		
Name	individualName		
Organisation Name	organisationName	European Commission . EG Environment	
position name	positionName		
role	role	007 (pointOfContact)	
Data Identification	MDDataIdentification		
citation	citation		
Title		Main Rivers	
date	date		
Date	refDate	2007-03-22	
Date type	dateType	002 (publication)	

Label	Name	Example	Value
Abstract	abstract	Vector dataset with the main rivers of Europe, whose catchment is exceeding 0000 km2.	
<u>Contact Info</u>	<i>contactInfo</i>		
Name	individualName	either name, organisation name or position name have to be documented	
Organisation Name	organisationName		
position name	positionName		
- Address	- address		
	<i>Cit</i>	<i>city</i>	
	<i>Postal code</i>	<i>postalCode</i>	
	<i>Country</i>	<i>country</i>	
	<i>Administrative Area</i>	<i>administrativeArea</i>	
	<i>email</i>	<i>electronicMailAddress</i>	
Phone	<i>phone</i>		
	<i>Voice phone</i>	<i>voice</i>	
	<i>Fax</i>	<i>facsimile</i>	
Online Resource	<i>onlineResource</i>		
Hours of service	<i>hoursOfService</i>		
Contact instructions	<i>contactInstructions</i>		
- linkage	linkage	http://www.wise.eu	
Topic Category	topicCategory	13 (inland waters)	
Spatial representation type	spatialRepresentationType	001 (vector)	
<u>Spatial Resolution</u>	spatialResolution		
equivalent scale	equivalentScale		
Denominator	denominator	10000000	
distance	distance		
Language	language	ENG	
Characterset		007 (ISO/IEC 8859-1)	
<u>Extent</u>	extent		
Geographic Element	geographicElement		
Ex GeographicBoundingBo x	EX_GeographicBoundingBo x		
Most western coordinate	westBoundLongitude	-20	
Most eastern coordinate	eastBoundLongitude	50	
Most southern coordinate	southBoundLatitude	25	
Most northern coordinate	northBoundLatitude	80	
Ex GeographicDescription	Ex_GeographicDescription	or	
Geographic Identifier	geographicIdentifier		
code	code	Europe, 15 (RBD code)	
Ex BoundingPolygon	EX_BoundingPolygon		
- polygon	polygon		
Distribution Information	distributionInfo		
distributionFormat	distributionFormat		
Name	name	GML	
Version	version	3.0	
digitalTransferOptions	transferOptions		
online	online		
linkage	linkage	http://www.wise.eu/download	
Data quality information	dataQualityInfo		

Label	Name	Example	Value
<u>Scope</u>	scope		
Level	level	dataset	
<u>Lineage</u>	lineage		
statement	steatement	The dataset has been create from national contributions due to reporting obligations according to the water framework directive.	
Process step	processStep		
description	decription		
source	source		
description	description		
Reference system Information		referenceSystemInfo	if not WGS84 or ETRS89
Name of coordinate reference system (CRS)	referenceSystemIdentifier	AT_MGI / AT_TM	
-	Prime Meridian	Prime Meridian	Greenwich
-	Prime meridian greenwich longitude	Prime meridian greenwich longitude	0°
code	code	it is recommended to use EPSG codes	
<u>Projection</u>	projection	Transverse Mercator Projection	
-	False Easting	falseEasting	500 000 m
-	False Northing	falseNorthing	0 m
-	Central Meridian	centralMeridian	24° E
-	Latitide of Origin	latitudeOfOrigin	0°
-	Scale factor at central meridian	scaleFactor	0.9996
code	code	it is recommended to use EPSG codes	
<u>Ellipsoid</u>	ellipsoid	Bessel 1841	
Ellipsoid semi major axis	Ellipsoid semi major axis	6 378 137 m	
llipsoid semi minor axis	Ellipsoid semi minor axis	635 911 m	
Ellipsoid inverse flattening	Ellipsoid inverse flattening	298.2572222	
code	code	it is recommended to use EPSG codes	
<u>datum</u>	datum	MGI	
code	code	it is recommended to use EPSG codes	