European Common database on Nationally Designated Areas (National – CDDA)

2013 – Quality assessment

ETC/BD task team: Brian Mac Sharry (MNHN)

ETC/BD task manager: Sabine Roscher (roscher@mnhn.fr)

EEA project manager: Mette Patizsch Lund

Activity 1.2.1.B3

Date: 13/06/2013

Table of contents

1. Delivery of datasets

Not relevant for public view

2. Description of the QA/QC

2.1 Introduction

The latest version of the Common Database on Designated Areas (CDDA), version 11 from 2013, covers the entire geographical area of the countries that make up the EEA (including the 7 West Balkan countries that are 'cooperating countries' of the EEA) and includes the full geographical area under the responsibility of European countries as well as other States and Territories related to key initiatives in the European region1.
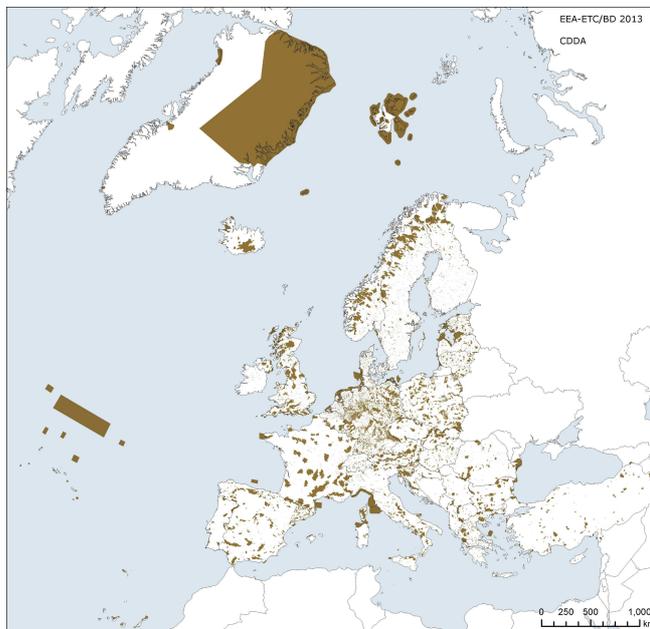

Figure 1: Extent of the ECDDA dataset,
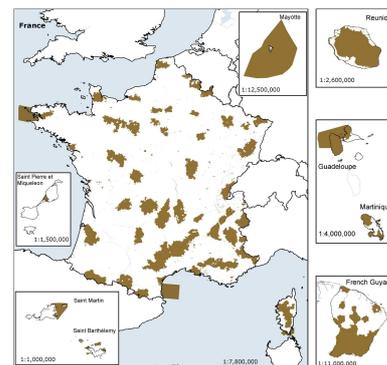
Figure 2: France (metropolitan, Dom-TOM, COM)

The resulting data covers the 39 EEA & EEA cooperating countries as well as Greenland (Denmark) and the French Overseas Departments and Territories[2] and Overseas Collectives[3] (figure 1 and figure 2).

In 2013 for inclusion into version 11 of the CDDA; 30 countries delivered tabular data (figure 3, & Appendix 1) and 31 countries delivered spatial data (figure 4 & Appendix 2). These datasets were subjected to a series of quality control and quality assurance (QA/AC) checks. Once the

---

[1] Memorandum of Cooperation between the EEA and the United Nations environment Programme–World Conservation Monitoring Centre (UNEP-WCMC) (2007).

[2] The *département d'outre-mer et territoires d'outre-mers* (DOM-TOM) of Martinique, Guadeloupe, French Guyana, Mayotte and Reunion (Article 2 point 287 (b) of the Lisbon Treaty)

[3] The *collectivités d'outre-mer* (COM) of Saint Martin (MAF), Saint Barthélemy (BLM) (these islands were formerly part of Guadeloupe but seceded to form a COM –Feb 22 2007) as well as Saint Pierre et Miquelon (SPM)

data passed these tests it was combined with data for those countries that did not submit data in 2013, which was extracted from the previous CDDA dataset, version 10.

The combined and integrated dataset that is the 2013 CDDA (version 11) covers 39 countries, and consist of a total of **97 481** records in the tabular database and **92 291** spatial records (see Appendix 3 for more details). With approximately 95% of records having spatial boundaries this reflects the work that countries have been investing into their protected areas network. Of these more than 90 000 boundaries only 3.7% of them are currently restricted from being downloaded. The reasons for this restriction can vary from copyright issue to the sensitive nature of some of the sites.

Table 1.

CDDA version	Number of records tabular	Number of records spatial
Version 11	**97 481**	**88 882**
Version 10	94 810	86 226
Version 9*	120 456	112 274
Version 8*	109 634	101 818

Table 1. Number of tabular and spatial records in version 11

* previous versions included data from the EECCA countries so the perceived decline in numbers is an artefact of this. In addiction between versions 9 and 10 Germany removed a number of sites from the CDDA as they did not fit the definition of protected areas used in the CDDA.
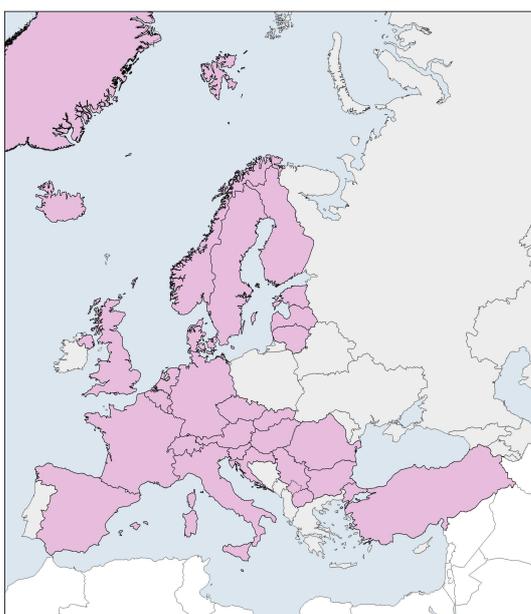

Figure 3: Countries that delivered tabular data.

Figure 4: Countries that delivered spatial data.

2.2 Main Issues

The majority of the data passed the QA/QC tests and was of a good quality. A number of the issues that were identified in the last reporting round were successfully resolved for the 2012 delivery. Those issues that remain, while minor, still result in a considerable extra work load in order to process the data.

The key issue remaining is the **use of identifiers for the spatial data**. In some cases the identifiers used are not globally unique, that is there are several records in the data set with the same identifier, and in other cases the identifier used is not directly linkable to the equivalent field in the tabular data. In these cases it is matter of examining every record and trying to successfully match spatial records with the matching record in the tabular data, for the vast majority of these cases this is a relatively simple but nonetheless a time consuming task.

The ideal identifier to use for the spatial data set is either the **'SITE_CODE'** (that is the CDDA/WDPA ID) or the **'SITE_CODE_NAT'** (national site code), where this is not possible the 'Site_Name' should be used, though it should be stressed that this option is far from ideal and can create a number of issues. If the 'Site_name' is to be used it should be a direct match for that in the tabular data, same character set, same case, same diacritical marks etc.

For the following countries there were a small number of issues regarding their delivery.

Croatia
There was one spatial site (Site name = 'Utjecajno podruèje (Tise–Debeljak na Papuku)') where a corresponding tabular record could not be found.

Luxembourg
There was one spatial site (Site code = ZH59) where a corresponding tabular record could not be found.

Malta
There was one spatial site (Site code = MALT_199) where a corresponding tabular record could not be found.

Poland
There was a significant problem in linking the Polish spatial data to the tabular data. The Polish spatial data used the name of the site as the unique identifier, though this name was often not globally unique, e.g. Jezioro Czarne and often there were differences in the name used in the

tabular and spatial data e.g. Chodelski versus Chodelski OChk in other cases there were issues with the diacritical characters used in the two datasets.

For the sites covered by designation PL02 out of 1 478 polygons supplied, 37 could not be matched, for sites covered by designation PL03 there were 2 sites that could not be matched for sites covered byPL04 out of 390 sites there were 241 sites that could not be matched. To match all these sites requires a significant amount of time on behalf of the ETC/BD

In future the Polish data should use the SITE_CODE as the unique identifier for the spatial data, as this is directly linkable to the tabular database.

Turkey

The principal issue with the Turkish data concerns the difficulty in linking the spatial data to the descriptive data. The unique identifier in the spatial data supplied by Turkey is the site name. In this field the site names are a combination of uppercase and lowercase, contain Turkish diacritical characters, underscores and abbreviations of the designation type at the end of the name e.g. _TKA, _MP. By contrast in the ECDDA database the "Site name" format consists of; the first letter of each word is upper case with the following letters being lower case, hyphenations, parentheses, some different characters (often lower case versions of the upper case diacritical characters) and the designation types are not at the end of the site name.

The Turkish NFP was very helpful in this regard.

The vast majority of issues relating to the spatial data are due to the SITE_CODE not being used as the unique identifier. This results in a significant amount of additional time on behalf of the agency to resolve these issues. Once again **it should be stressed that "SITE_CODE" should be used as the unique identifier**

2.3 Spatial Validation

The 31 countries that submitted spatial data did so in the form of shape files or personal geodatabases. All the data from these countries was converted to the shapefile format[4] and subjected to a series of spatial QA/QC checks. There were over 3 500 records supplied as points, these were buffered by 50m and merged with the polygon data to have just 1 spatial dataset. This was done for technical reasons to allow the data to be easily integrated into map viewers;

The spatial validation consisted of the following stages:

---

[4] This rational for using shapefiles over personal geodatabases is that tehse geodatabases are often version specific and in having the data as shapefiles this bypasses this issue. In future following discussion between the ETC/BD, EEA and WCMC the data may ultimately be delivered as personal geodatabases.

- 2.3.1 Projection validation
- 2.3.2 Geometry validation
    - o 2.3.2.1 Geometry must be valid if not Repair geometry.
    - o 2.3.2.2 Multipart polygons must not be present, if so "Dissolve"
- 2.3.3 Geographical and Attribute validation
    - o 2.3.3.1 Data must lie within the country extent (terrestrial + marine).
    - o 2.3.3.2 Check if coordinates in the database are within the country.
    - o 2.3.3.3. Attribute validation, check that each feature has a site code, if not link by another field, if not possible check site name and try to link using site name, grid coordinates, area
    - o 2.3.3.4 Calculate coordinates for each polygon and compare them to the coordinates as supplied by country.
    - o 2.3.3.5 Comparison of the Area, area calculated using GIS and compared to that supplied by the Country.

2.3.1 Projection validation

All data were checked to ensure they add a projection file. All files passed this first step. The Table in Appendix 4 details the native projection or projections of the data. All the data were transformed to ETRS LAEA 5210 to incorporate them into a European wide projection system.

2.3.2 Geometry validation

2.3.2.1 Geometry must be valid

The rule for this check was that the geometry must be valid. The geometry of all the files was checked using the 'Check Geometry' Tool in Arc GIS 10.1. Where this QA/QC identified errors the 'Repair Geometry' Tool was run in order to repair them. Common geometrical issues were self intersections or incorrect ring ordering.

2.3.2.2 Multipart polygons must not be present:

The rule for this check was that multipart polygons must not be present. After the geometry had been validated all the files were dissolved using the 'Dissolve' command in Arc GIS 10.1. All features were aggregated based on the unique identifier.

2.3.3 Geographical and attribute validation:

2.3.3.1 Data must lie within the member State extent:

The rule for this test was that all data should lie within the terrestrial and marine extent of the country. Due to the differences in the borders of countries between what they have available

nationally and what is available at the European level a buffer of 5km was created around the country boundaries and the Marine extents[5]. All the data passed this test.

2.3.3.2 Check if coordinates are within the Member State:
The descriptive database contains two fields (LAT, LON) that are used by countries to add coordinate information to the sites. The coordinates as supplied by the countries were converted to a point and a projection (WGS 84) added to the points. The location of these points was checked against the extents of the countries. A small number of sites occurred outside the country extent (Appendix 7) the principal cause of these 'errors' was the latitude and longitude being switched. Where this occurred the coordinates were corrected in a separate field and the test run again.

A number of countries did not supply coordinates for their site, instead they asked that the coordinates be calculated automatically where spatial datasets were provided[6]. The centroid of the polygon was generated so that it would lie inside the polygon feature using the 'Feature to Point' tool in Arc GIS 10.1

2.3.3.3 Attribute validation:
All spatial features were checked against the descriptive data to ensure that a unique identifier within the spatial data could be linked to the descriptive data and ultimately to the site code. All countries had unique identifiers in the spatial data that could be linked to the descriptive data base (version 9) and the site code filled, except for the cases mentioned in section 2.

2.3.3.4 Calculate coordinates for each polygon and compare them to the coordinates as supplied by country.
The centroid of the polygon was calculated to lie within the polygon. The location of this centroid was compared with the latitude and longitude as described in the descriptive database, where it exists. Four situations were distinguished, where the differences is >5km, >10km, >50km and >100km (Appendix 6 shows a table giving the number of sites per country that fall into these classes).

2.3.3.5 Comparison of the Area: GIS calculated area compared to that supplied by the Member State

The area of the polygons for each site was calculated and compared to the areas in the descriptive database for the same site, where the area was given. Three situations were identified, where the difference was >10%, >50% or >100% (Appendix 7 shows a table giving the number of sites per country that fall into these classes).

2.4. Results

Once the data from the 31 EEA and EEA collaborating countries that delivered in 2013 underwent the QA/QC procedures it was merged into a single polygon feature. The data for

---

[5] The boundaries used were generated during the Article 17 process and consist of the National GeoSpatial Agency (NGA) coastline data (global shoreline data, satellite derived high water line data) supplemented by EU Member State data where supplied, the internal boundaries are based on EEA supplied Euroboundary map data, the marine extents are based on the EEZ obtained from the VLIZ (http://www.vliz.be/En/INTRO) which are based on the UN law of the Sea.

[6] These countries denoted this by filling in the value "02" in the filed CDDA_Coordinate_Code in the sites table or via correspondence stating that they wanted the coordinates to be calculated

those countries that did not submit spatial data in 2013 were extracted from the previous CDDA dataset version 10.

The 2013 CDDA data set covers 39 countries (Appendix 3) with over 97 000 records in the database and over 90 000 spatial records. As mentioned previously there are still a small number of restrictions on the dissemination of the data. Table 2 highlights the 3 options for data dissemination; this table is extracted from the CDDA Data Dictionary. This field is included in the attributes of the shape files as the field 'CDDA_Dissemination_code'. The dissemination code was taken from the field 'CDDA_Dissemination_code' from the table 'sites_boundaries'

Values	Definition
01	Yes, public dissemination by EEA allowed
02	No dissemination, except to UNEP-WCMC for internal use
03	No dissemination, only for EEA internal use

Table 2. Dissemination instruction code (CDDA Data dictionary, EEA).

3. Concluding remarks

- The majority of the data delivered under the CDDA reporting cycle is of a high quality.

- The download of datasets from Reportnet and the merging of datasets into a European one is a time consuming task if carried out manually. It is hoped that the development of the automated harvesting, merging of datasets and creation of xml-based QA/QC reports as carried out by the EEA.

- The provision of the helpdesk, as well as the initial automatic QA/QC of the data in addition to communication with data providers, where necessary in case of data issues, is considered as being very beneficial in order to assure - or where necessary to improve - the high quality of the CDDA data flow.

- The new site code generation tool has been used by about 50% of the countries who requested codes. The main reason for the remaining countries not using the tool was the lack of a user manual. Therefore the ETC/BD created such a user manual and sent it to the countries having difficulties in using the tool. (The user manual has been included in the CDDA data flow documentation; it might be useful to make it available online for the next collection round.)

  The tool currently has a limit of 500 codes per country. The ETC/BD helpdesk received some requests (FR, DE, FI) for more than 500 codes. Therefore it is recommended to increase the limit to 1000 codes per country.

- Over the last 3 years the CDDA data has been used for the calculation of Indicators (SEBI 07 on nationally protected areas, CSI 008 on designated areas), it was discussed in the 2nd Message of the '10 messages for 2010' as well as being used in the State of the Environment 2010 (SOER 210) report specifically Part B on Biodiversity in Protected Areas. In addition the CDDA was a very important dataset for the EEA report on 'Protected Areas'. The CDDA was the key dataset used in the production of statistics and further analysis of the nature of Protected Areas in Europe.

- With the advert of the Water Framework Directive a number of 'protected sites' are being created such as 'Nitrate vulnerable zones', it needs to be discussed whether these should be covered by the CDDA dataflow or not.

Appendix 1 30 countries that supplied tabular data for ver. 11

Table listing which countries delivered descriptive data in 2013.

Country	ISO3
Austria	AUT
Belgium	BEL
Bulgaria	BGR
Croatia	HRV
Czech Republic	CZE
Denmark	DNK
Estonia	EST
Finland	FIN
France	FRA
FYROMacedonia	MKD
Germany	DEU
Hungary	HUN
Iceland	ISL
Italy	ITA
Kosovo (UNSCR 1244/99)	XKX
Latvia	LVA
Lithuania	LTU
Luxembourg	LUX
Malta	MLT
Netherlands	NLD
Norway	NOR
Romania	ROM
Serbia	SRB
Slovakia	SVK
Slovenia	SVN
Spain	ESP
Sweden	SWE
Switzerland	CHE
Turkey	TUR
United Kingdom	GBR

Appendix 2: 31 countries that supplied spatial data for ver. 11

Table listing which countries delivered spatial data in 2013.

Country	ISO3
Austria	AUT
Belgium	BEL
Bulgaria	BGR
Croatia	HRV
Czech Republic	CZE
Denmark	DNK
Estonia	EST
Finland	FIN
France	FRA
FYROMacedonia	MKD
Germany	DEU
Hungary	HUN
Iceland	ISL
Italy	ITA
Kosovo (UNSCR 1244/99)	XKX
Latvia	LVA
Lithuania	LTU
Luxembourg	LUX
Malta	MLT
Netherlands	NLD
Norway	NOR
Poland	POL
Romania	ROM
Serbia	SRB
Slovakia	SVK
Slovenia	SVN
Spain	ESP
Sweden	SWE
Switzerland	CHE
Turkey	TUR
United Kingdom	GBR

Appendix 3: 39 countries

Table listing which countries make up the 2013 CDDA and the number of records in the descriptive database and spatial data.

Country	ISO3	No. of records in database	No. of records GIS -poly	No. of records GIS -point	Removed features
Albania	ALB	796	48		
Austria	AUT	1202	1182		
Belgium	BEL	1904	1357		
Bulgaria	BGR	1016	884	131	
Bosnia - Herzegovina	BIH	156	3	30	
Switzerland	CHE	5910	5877		
Cyprus	CYP	45	17	4	
Czech Republic	CZE	2367	2367		
Germany	DEU	16255	16243		
Denmark	DNK	2267	1917		
Spain	ESP	1561	1561		
Estonia	EST	11446	10328	1042	2736
Finland	FIN	10627	10625		
France	FRA	2706	2609		
United Kingdom	GBR	9120	9079		
Greece	GRC	839	773		
Croatia	HRV	431	330	96	
Hungary	HUN	279	218		218
Ireland	IRL	309	155		
Iceland	ISL	109	109		
Italy	ITA	871	870		
Liechtenstein	LIE	41	41		
Lithuania	LTU	359	352		
Luxembourg	LUX	105	104		
Latvia	LVA	706	704		
FYROM	MKD	74	52	16	
Malta	MLT	202	200		
Montenegro	MNE	37	0	5	
Netherlands	NLD	303	301		
Norway	NOR	2886	2799		
Poland	POL	2186	1639		
Portugal	PRT	224	223		
Romania	ROU	978	884		104
Serbia	SRB	247	156		
Slovakia	SVK	1141	1127		
Slovenia	SVN	1947	1071	868	
Sweden	SWE	13486	12114	1372	
Turkey	TUR	2245	350		350
Kosovo (UNSCR 1244/99)	XKX	98	26		
Total		**97481**	**88727**	**3564**	**3667**

Appendix 4

Check if coordinates supplied by countries (in the 2013 tabular data) are within the country.

This is only run on those countries that supplied data in 2013.

Belgium

There is one Belgian site where the supplied coordinates are outside the country.

SITECODE	NAME	LAT	LON	Actual Location
5593	Kalmthoutse Heide	47.7395	13.036	Austria

Denmark

There is one Danish site where the supplied coordinates are outside the country.

SITECODE	NAME	LAT	LON	Actual Location
5758	Damsted Klit, Østre Damsted	47.8810	13.154	Austria

Serbia:

There are 4 Serbian sites where the supplied coordinates are outside the country.

SITECODE	NAME	LAT	LON	
16394	Grmija	40.083333	21.21667	Grevena, Greece
328881	Prirodni prostor oko manastira Koporin	44.3152	19.001	In Bosnia Herzegovina
328883	Gamzigrad	44.909	22.174	In Romania
555552431	Lesni profil kod Starog Slankamena	45.142	22.254	In Romania

Appendix 5

Table highlighting differences in the coordinates from the descriptive data and those calculated by GIS. The table shows the number of sites with a difference of over 10km, over 50km & over 100km.

Country	Diff over 10km	Diff over 50km	Diff over 100km
AUT	2		
BEL	1	1	1
BGR	5	1	1
CHE	6		
DNK	5	4	3
ESP	5		
EST	2		
FRA	85	7	5
GBR	155	20	9
HRV	1		
ISL	4	1	
LTU	13		
LUX	2		
LVA	7	2	
MLT	5		
NLD	11		
NOR	5		
ROU	14	5	5
SRB	28	22	18
SVK	1146	987	694
SVN	7	7	7
SWE	35	1	0
TUR	32	18	13
XKX	3		

Appendix 6:

Table highlighting differences in the area of the sites comparing the areas as supplied in the tabular data to the GIS derived areas. The table shows the number of sites with a difference in area of over 10%, 50% & over 100%.

Country	>10%	>50%	> 100%
AUT	6	5	4
BEL	13	6	0
BGR	400	320	239
CHE	100	43	2
CZE	48	17	9
DEU	254	66	19
DNK	131	48	23
ESP	191	109	14
EST	2	1	0
FIN	1	1	1
FRA	376	195	117
GBR	120	80	54
HRV	3	1	0
HUN	5	1	1
ITA	158	68	34
LUX	4	1	0
LVA	1	1	0
MLT	2	0	0
NLD	2	2	1
NOR	10	4	3
SRB	26	17	14
SVK	724	662	307
SVN	7	5	4
TUR	15	10	5
XKX	5	4	2

Appendix 7: Field names of spatial data

Field Name	Description
SITE_CODE	Site code of the record
Parent_ISO	ISO 3 digit code as taken from the tabular dataset
ISO3	Parent ISO 3 digit code as taken from the tabular dataset
ODESIGNATE	Name of the designation of the record in the native language, as taken from the 'designations' field in the tabular dataset
YEAR	Year of designation of the record, as taken from the tabular dataset
SITE_AREA	Area of the record taken from the tabular data (Site_Area from the 'sites' table)
SITE_NAME	Name of the record as taken from the tabular data set
DESIGNATE	Name of the designation of the record in English as taken from the 'designations' field in the tabular dataset
IUCNCAT	IUCN category as taken from the tabular dataset
CDDA_Disse	Dissemination code for the record as taken from the field 'CDDA_Dissemination_Code' from the 'site_boundaries' table.