CDDA version 12

Prepared by / compiled by: Manuel Löhnertz Organisation: ETC/BD - GeoVille EEA project manager: (Mette.Lund@eea.europa.eu) Task Manager: (Sabine Roscher/roscher@mnhn.fr) Work Package n°:1.7.2.A

Version history

Version	Date	Author	Status and description	Distribution	
0.1	31/07/2014	Manuel Löhnertz	Pre-draft for ETC	Sabine Roscher	
1.0	31/07/2014	Manuel Löhnertz	First draft for EEA	Mette Lund	
2.0	08/10/2014	Manuel Löhnertz	Final draft for EEA	Mette Lund	

Contents

1	Back	kground & Introduction	3
	1.1	CDDA	3
	1.2	Definition of terms	3
2	Deliv	very of datasets	5
		The CDDA v12 delivery	
3	CDE	DA Production	9
4	Veri	fication	10
	4.1	Verification overview	10
	4.2	Verification steps	11
5	QA/	QC Results	13
	5.1	First spatial verification information	13
6	Con	cluding remarks	

1 Background & Introduction

The present document describes the activities and procedures for verification of the CDDA version 12 (2014).

1.1 CDDA

The Common Database on Designated Areas (CDDA) inventory holds information about protected areas and the national legislative instruments, which directly or indirectly create protected areas. The dataset contains data on individual nationally designated sites and designations in EEA member and collaborating countries.

1.2 Definition of terms

Before entering in details of the QA/QC process of the CDDA database, a definition of our understanding of a couple of important terms and related activities is required.

Validation / Quality control (QC)	Validation is the process by which the accuracy and consistency of products are evaluated and the associated uncertainties are quantified (Justice et al., 2000).
	Product <i>accuracy</i> is assessed by a comparison with independent data sources such as ground-based measurements, more detailed data or well-calibrated models.
	Inter-comparison with other equivalent products is also part of the validation process allowing building up a community reference product when no or not enough independent data are available.
	Quality control, or QC for short, is normally carried out after the end of the production and aims at providing the user with measurable / quantitative information how well the product meets the pre- defined specifications.
Verification / Quality assurance (QA)	The act of reviewing, inspecting, testing, checking, auditing, or otherwise establishing and documenting whether items, processes, services, or documents conform to specified requirements.
	Verification is a qualitative process in which intermediate or final results of the production process are commented and potential deviations from the specifications are highlighted. The verification will be

Table 1-1 Definition of terms

done during the course of production and is meant to increase data and production quality.
Quality Assurance (QA) is a way of preventing mistakes or defects in products and avoiding problems when delivering solutions or services to customers.
QA is applied to physical products in pre-production to verify what will be made meets specifications and requirements, and during manufacturing production by validating whether lot samples meet specified quality controls.
QA is also applied to software to verify that features and functionality meet business objectives, and that code is relatively bug free prior to shipping or releasing new software products and versions.

The QA/QC process carried out in the current project is - according to our understanding - a verification process, as the output does not provide quantitative results about the database quality and is used an element of a process to correct and improve the integrated European database version.

2 Delivery of datasets

The following tabular and spatial data have been uploaded on 10 July 2014 to <u>https://svn.eionet.europa.eu/repositories/Workdata/CDDA/cdda_ver12/version_2014-07-09/</u>

- SPATIAL.RAR
- TABULAR.RAR

2.1 The CDDA v12 delivery

The latest version of the Common Database on Designated Areas (CDDA), version 12 from 2014, covers the entire geographical area of the <u>countries that make up the EEA</u> (including the 6 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 region.

The resulting data covers the 40 EEA & EEA cooperating countries as well as Greenland (Denmark) and the French Overseas Departments and Territories and Overseas Collectives (Map).



Map 2-1 Extent of the CDDA dataset (DOM/TOMs are not shown here)

34 countries delivered tabular and spatial data in 2014, which had to be included into version 12 of CDDA.

These datasets were subjected to a series of quality control and quality assurance (QA/AC) checks. Once the data passed these tests they were combined with data from those countries that did not submit data in 2014, which was extracted from the previous CDDA dataset, version 11.

The combined and integrated dataset (i.e. 2014 CDDA, version 12) covers 40 countries, and consists of a total of **98 375** records in the tabular database and **95 176** spatial records.

CDDA version	Number of records	Number of records		
	tabular	spatial		
Version 12	98 375	95 176		
Version 11	97 481	92 757		
Version 10	94 810	86 226		
Version 9*	120 456	112 274		
Version 8*	109 634	101 818		

* previous versions included data from the EECCA countries so the perceived decline in numbers is an artefact of this. In addition, 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.

The following map presents the countries, which provided data for the CDDA version 12.



Map 1-2 Countries that delivered data

A more detailed overview is given in the following table. The cells marked with orange show countries which have not delivered data for various reasons (e.g. no new national CDDA).

	ISO - 2		Version 12	Version 11	
Name	digit	ISO - 3 digit	2014	2013	
Albania	AL	ALB	yes	no	
Austria	AT	AUT	yes	yes	
Bosnia - Herzegovina	BA	BIH	no	no	
Belgium	BE	BEL	yes	yes	
Bulgaria	BG	BGR	yes	yes	
Switzerland	СН	CHE	yes	yes	
Czech Republic	CZ	CZE	yes	yes	
Cyprus	CY	СҮР	no	no	
Germany	DE	DEU	yes	yes	
Denmark	DK	DNK	yes	yes	
Estonia	EE	EST	yes	yes	
Spain	ES	ESP	yes	yes	
Finland	FI	FIN	yes	yes	
France	FR	FRA	yes	yes	
Greece	GR	GRC	yes	no	
Croatia	HR	HRV	yes	yes	
Hungary	HU	HUN	yes	yes	
Ireland	IE	IRL	yes	no	
Iceland	IS	ISL	yes	yes	
Italy	IT	ITA	yes	yes	
Liechtenstein	LI	LIE	no	no	
Lithuania	LT	LTU	no	yes	
Luxembourg	LU	LUX	yes	yes	
Latvia	LV	LVA	yes	yes	
Monaco	MC	МСО	no	no	
FYROM	MK	MKD	yes	yes	
Malta	MT	MLT	yes	yes	
Montenegro	ME	MNE	yes	no	
Netherlands	NL	NLD	yes	yes	
Norway	NO	NOR	yes	yes	
Poland	PL	POL	no	no	
Portugal	РТ	PRT	yes	no	
Romania	RO	ROU	yes	no	
Serbia	RS	SRB	yes	yes	
Sweden	SE	SWE	yes	yes	
Slovakia	SK	SVK	yes	yes	
Slovenia	SI	SVN	yes	yes	
Turkey	TR	TUR	yes	yes	
United Kingdom	UK	GBR	yes	yes	
Kosovo (UNSCR 1244/99)	ХК	ХКК	yes	no	

Table 2-2 Data deliveries respectively data updates for CDDA version 12 & 11 (tabular & spatial)

The national CDDA data were delivered as table and as spatial dataset.

Tabular data:

EEA provides the different national institutes with an MS-Access CDDA template database and technical specifications via the <u>Central data repository</u> (CDR).

Figure 2-1 CDDA MS-Access data model & technical specification document



Using the template and the specification the countries update the table with the national CDDA information. After the national update, the database is uploaded again to the CDR.

Spatial boundary data:

Next to the tabular data the countries are asked to update their spatial CDDA data and to upload the data to the CDR.

3 CDDA Production

The final CDDA v12 database & map is the union of all single national databases & maps.

TABULAR:

Once the data deliveries could pass the quality check without errors the single tables were imported to a new MS-Access database.

SPATIAL:

Once the data deliveries could pass the quality check without errors the single vector data were first merged to a national spatial dataset. Next all merged spatial datasets were used for the construction of the final spatial CDDA dataset in a file-geodatabase.

Figure 3-1 Workflow spatial map production



4 Verification

The spatial and tabular data were checked at different points during the CDDA v12 production. First the input deliveries were checked for completeness. Then semantic checks of the single deliveries were carried out. After the final production of the full CDDA versions a second verification was done.

4.1 Verification overview

Tabular and spatial data were checked using different methods.

Tabular:

The countries uploaded the national CDDA v12 MS-Access versions on the CDR. An automatic QA, which is integrated in the database, performed semantic checks of the different tables. Errors were analysed and corrected. If necessary information were missing, the countries were asked for a updating the data.

After a successful quality check of the deliveries, the single tables were imported to the CDDA-v12 database template (provided by the EEA). Afterwards final QA checks were done using the integrated QA tool.

Figure 4-1 Tabular data verification overview



Spatial:

The countries delivered various kinds and numbers of spatial data. Some countries delivered one polygon vector file, others more than 20 different polygon and point vectors files.

Figure 4-2 Example: spatial delivery by one country



Therefore the first step of the verification was to check all the single spatial datasets. After the data passed the semantic check without errors, they were transferred to a new national CDDA-v12 dataset in ETRS-LAEA89 projection (EPSG: 3035). Next the single national spatial datasets were merged to one European CDDA dataset in a file-geodatabase. If countries did not deliver new data, the corresponding data from the CDDA version 11 was imported. If spatial data was delivered as point dataset the points were buffered using a 50m buffer and merged with the polygon layer. With this operation the final spatial dataset only consist of polygon vector layers.

Figure 4-3 Spatial data verification overview



4.2 Verification steps

In the following chapter only the verification of the spatial data will be explained. Information of the QA/QC of the tabular data can be found in the MS-Access database and on the CDR.

For improving the entire verification process, a python script was created. For the python script the typical standard python modules were used (qgis, arcpy, gdal, ...).

The different verification processes can be divided into three major steps:

- 1. Formal check
- 2. Mapping check
- 3. Topology check

The different checks were carried out with all single delivered spatial data and the results were documented.

- 1. Formal check
 - file format (shp, gdb, mdb, ...) (file readable?)
 - data type (point, line or polygon) (only point and polygons allowed)
 - attributes name (check if all necessary attributes available)
 - projection (WGS84, ...)
 (if projection not ETRS-LAEA, the dataset was projected)
- 2. Mapping check
 - Unique identifiers (SITE_CODE) ([SITE_CODE] is necessary!)
 - Valid site codes (double values existing?, NULL values existing?, ...)
- 3. Topology & location check
 - Multipart polygons

Multipart polygon – dissolved by SITE_CODE – are allowed. No double SITE-CODES in the attribute table are allowed!

• Geometry

The geometry was automatically check and repaired

- Data location (CDDA sites should be inside the corresponding country) The location of the single CDDA sites of the every country was checked using a European boundary layer: EuroBoundaryMap (full European coverage) version 7.0, Oct. 2013 <u>https://sdi.eea.europa.eu/data/continental/europe/external_db/eurogeographics/</u> <u>EuroBoundaryMap/EuroBoundaryMap_v70/EuroBoundaryMap_v70_shape/F</u> ullEurope/
- LAT/LON attribute in spatial data

Some countries delivered LAT/LON attributes (in WGS84 projection), a check was done, if the coordinates a located inside the country

After the first spatial and tabular verification checks, some comparisons between both deliveries were made:

Figure 4-4 Verification of cross relationships between spatial and tabular dataset



- SITE_CODE: it was checked if all spatial codes are found in the tabular data (and vice versa).
- LAT_LON: the location of the tabular LAT_LON values were checked using a GIS

The centroid of the polygon was calculated to be located 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

• AREA: 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.

5 QA/QC Results

In this chapter a selection of information and results about the new CDDA dataset will be presented.

5.1 First spatial verification information

The single spatial delivery were analysed and checked.

Figure 5-1 Semantic check



Table 5-1 Selection o	f semantic check i	results and information
-----------------------	--------------------	-------------------------

Country	ISO3	Comment			
Albania	ALB	 Two polygon shp files National projection no SITE_CODE attribute no LAT/LON attribute 			
Austria	AUT	 One polygone shp file MGI_Austria projection 1281 features 			
Bosnia - Herzegovina	BIH	no new data delivery			
Belgium	BEL	 18 different polygone shp files Belge_Lambert_1972 projection 			
Bulgaria	BGR	 6 different polygones shp files WGS_1984_UTM_Zone_35N projection no LAT/LON attribute 			
Switzerland	CHE	One polygone shp file CH1903_LV03 projection no LAT/LON attribute			
Czech Republic	CZE	 One polygone shp file S-JTSK_Krovak_East_North projection no LAT/LON attribute 			
Cyprus	СҮР	no new data delivery			
Germany	DEU	 3 different polygones shp files DHDN_3_Degree_Gauss_Zone_3 projection no LAT/LON attribute 			
Denmark	DNK	 Two different polygones shp files (the CDDA-DNK-v11 version plus the NEW single CDDA sites) 			

Country	ISO3	Comment
		 ETRS_1989_LAEA projection no LAT/LON attribute
Estonia	EST	 2 polygons & one point shp file
		 MI_0 projection
		no LAT/LON attribute
Spain	ESP	One polygone shp file
		 ETRS_1989_LAEA projection no LAT/LON attribute
Finland	FIN	One polygone shp file
		 ETRS_1989_LAEA projection
		no LAT/LON attribute
France	FRA	• 43 polygons & 1 point shp file
		Different projections: ETDS 1080 LAFA
		 ETRS_1989_LAEA WGS_1984_UTM_Zone_20N
		 RGFG95_UTM_Zone_22N
		 WGS_1984_UTM_Zone_38S
		• IGN62_UTM_Zone_42S
		 RGR92_UTM_Zone_40S WGS_1984_UTM_Zone_21N
		• GCS_WGS_1984
		no LAT/LON attribute
Greece	GRC	One polygone shp file
		GREEK_Grid projection
Croatia	HRV	 no LAT/LON attribute 2 polygons & one point shp file
Cibatia	TIK V	 GCS_WGS_1984
		 no LAT/LON attribute
Hungary	HUN	One polygon shp file
		• GCS_WGS_1984
* * *		no LAT/LON attribute
Ireland	IRL	One polygone shp file TTPS 1000 LAFA projection
		 ETRS_1989_LAEA projection no LAT/LON attribute
Iceland	ISL	One polygone shp file
		• ETRS_1989_LAEA projection
		no LAT/LON attribute
Italy	ITA	One polygone shp file
		ETRS_1989_LAEA projection
		no LAT/LON attribute
		Comments:
		The data was not put on the CDR. Therefore the data was downloaded from an Italian
		server and copied to the correct place
Liechtenstein	LIE	no new data delivery
Lithuania	LTU	no new data delivery
Luxembourg	LUX	Six polygone shp files
Lancinoourg	Lon	 LUREF projection
		• For two files no SIDE_CODE exists (SITE_CODE was manually updated)
		no LAT/LON attribute
T a tasia	LVA	
Latvia	LVA	 One polygone shp file LKS_1992_Latvia_TM projection
		 no LAT/LON attribute
Monaco	MCO	no new data delivery
FYROM	MKD	2 polygons & one point shp file
		• GCS_WGS_1984
N 1.		no LAT/LON attribute
Malta	MLT	One polygone shp file
		 MLT national projection no LAT/LON attribute
Montenegro	MNE	One polygone shp file
C		 ETRS_1989_LAEA projection
		no LAT/LON attribute

Country	ISO3	Comment					
Netherlands	NLD	 Two polygone shp files NLD national projection no LAT/LON attribute 					
Norway	NOR	 Two polygone shp files WGS_1984_UTM_Zone_33N projection no LAT/LON attribute 					
Poland	POL	no new data delivery					
Portugal	PRT	 Six polygone shp files Different projections: GCS_WGS_1984 Porto_Santo_1995_UTM_Zone_28N Some missing SITE_CODES (SITE_CODE was manually updated) no LAT/LON attribute 					
Romania	ROU	 One polygone shp file ETRS_1989_LAEA projection no LAT/LON attribute 					
Serbia	SRB	One polygone shp file ETRS_1989_LAEA projection no LAT/LON attribute					
Sweden	SWE	 2 polygons & one point shp file GCS_WGS_1984 no LAT/LON attribute 					
Slovakia	SVK	One polygone shp file ETRS_1989_LAEA projection no LAT/LON attribute					
Slovenia	SVN	 2 polygons & one point shp file GCS_WGS_1984 no LAT/LON attribute 					
Turkey	TUR	 10 polygons point shp files Missing SITE_CODE different projections: GCS_European_1950 GCS_WGS_1984 no LAT/LON attribute <u>Comments:</u> Inside the spatial data a high number of sites are without SITE_CODE information. It was not possible to update all missing SITE_CODE information's 					
United Kingdom	GBR	 One polygone shp file ETRS_1989_LAEA projection no LAT/LON attribute 					
Kosovo (UNSCR 1244/99)	ХКК	 One polygone shp file ETRS_1989_LAEA projection no LAT/LON attribute 					

After the first verification all data were transformed to the ETRS_1989_LAEA projection and merged into a national CDDA_v12 layer and finally to the European CDDA_v12 spatial dataset. Using the final aggregated spatial dataset different additional tests were done. First for all sites the area and coordinates were calculated. Then the total CDDA area was compared between version 12 and 11 country by country:

hame	iso2	iso3	count 2013	count 2014	Difference 2014-2013	km² 2013	km ² 2014	Difference km²2014-2013
Albania	AL	ALB	48		6		4810,06	
Austria	AT	AUT	1182	1198	16	·	26273,66	
Bosnia - Herzegovina		BIH	33	33	0			data from 2013
Belgium		BEL	1357	1557	200	,		
Bulgaria		BGR	884	1042	158	· · · · ·		
Switzerland	СН	CHE	5877	5879	2	3032,00	3032,98	0,98
Czech Republic	CZ	CZE	2367	2434	67	13074,18		
Cyprus	CY	СҮР	17	17	0	203,48	203,48	data from 2013
Germany	DE	DEU	16243	16814	571	133169,19	132773,40	-395,79
Denmark	DK	DNK	1917	1923	6	985837,12	985842,80	5,68
Estonia	EE	EST	10328	11482	1154	23129,94	22759,55	-370,38
Spain	ES	ESP	1561	1561	0	68171,94	74576,82	6404,87
Finland	FI	FIN	10625	10875	250	34410,53	34550,88	140,35
France	FR	FRA	2609	2779	170	214708,13	270635,47	55927,34
Greece	GR	GRC	773	799	26	34357,17	38034,61	3677,44
Croatia	HR	HRV	330	415	85	7216,44	7170,70	-45,74
Hungary	HU	HUN	218	218	0	8495,62	8495,62	0,00
Ireland	IE	IRL	155	155	0	603,62	603,62	0,00
Iceland	IS	ISL	109	108	-1	20065,14	19287,33	-777,81
Italy	IT	ITA	870	871	1	59446,00	59387,53	-58,48
Liechtenstein	LI	LIE	41	41	0	81,78	81,78	data from 2013
Lithuania	LT	LTU	352	352	0	10435,31	10435,31	data from 2013
Luxembourg	LU	LUX	104	104	0	1257,69	1257,69	0,00
Latvia	LV	LVA	704	707	3	16851,53	16839,08	-12,45
Monaco	MC	MCO						data from 2013
FYROM	MK	MKD	52	75	23	2042,02	2046,20	4,18
Malta	MT	MLT	200	203	3	337,62	337,75	0,13
Montenegro	ME	MNE		5	5		0,01	0,01
Netherlands	NL	NLD	301	256	-45	17250,72	12783,41	-4467,31
Norway	NO	NOR	2829	2863	34	181188,66	181980,78	792,12
Poland	PL	POL	1641	1641	0	60039,81	60039,81	data from 2013
Portugal	РТ	PRT	223	186	-37	121249,44	121021,51	-227,92
Romania	RO	ROU	884			20186,28	20032,82	
Serbia	RS	SRB	156					
Sweden	SE		12114	13958			57330,54	
Slovakia		SVK	1127	1123	-4		11321,87	

Table 5-2 CDDA area comparison: CDDA_v12 vs CDDA_v11

name	iso2	iso3	count 2013	count 2014	Difference 2014-2013	km² 2013	km² 2014	Difference km²2014-2013
Slovenia	SI	SVN	1071	2045	974	27062,84	27974,74	911,90
Turkey	TR	TUR	350	1302	952	21909,76	38441,69	16531,93
United Kingdom	UK	GBR	9079	9080	1	94636,54	102761,22	8124,68
Kosovo (UNSCR 1244/99)	XK	ХКК		26	26		1245,18	1245,18

During the area comparison large differences were detected. The extreme values such as France (increase CDDA area: 55 927 km² between 2013 & 2014) were checked in detail. Both CDDA versions were imported to a GIS and analysed: Map 5-1 CDDA France 2013 vs. 2014 example



The map shows the changes between 2013 & 2014 for France. In the new CDDA version new large sites (marine and terrestrial) were added. As no clear errors could be found, the data were imported.

In the Netherlands, about 4500km² CDDA area was "lost" between 2013 & 2014. The delivered addition information from the Netherlands was controlled and the following comment was found:

Category 'natuurreservaten met beheersubsidie' / 'Nature reserve' (NL21)

A major change has taken place here. As these areas are not truly protected under any kind of national law (they have no legal status), and do not represent actual sites with consistent additional data as "Name", "Year of establishment", "Management", etc., we decided to remove these areas from the CDDA database and deliver them as "Spatial coverage of Protected Areas not defined as Sites" only.

These sites have some kind of protection, being owned or managed by a national, regional or local nature protection organisation, such as the Dutch State Forestry, but no legal status.

Therefore, we can also in this case accept the CDDA delivery from the Netherland.

6 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.
- The provision of the helpdesk, as well as the initial automatic QA/QC of the tabular 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 countries delivered the spatial data in very different forms:

Some countries delivers point and polygon data	Some countries only polygon layers				
Some countries delivers one polygon	Some countries delivers more than 20 different polygons				
Some countries upload all spatial data in national projection	Some countries delivers the data already in the final LAEA89 projection				
Some countries produced a high number of attributes	Some countries upload spatial data with no SITE_CODE information				
different attribute names [SITE_CODE]; [SITECODE]; [CODE];					
no clear file names (RN.shp, TB_31122013.shp, de07_lsg2013.shp,)					
not all files are delivered with metadata information					

Table 6-1 Overview of the different types of spatial data delivery

As shown in the table, the entire spectrum of types of spatial data deliveries was uploaded to the EEA server. Therefore, first all data had to be brought up to the same standard (LAEA89-projection, same attributes, ...). This circumstance made an automatic validation significantly more difficult. Therefore for the next CDDA data collection it is recommended to provide more strict specifications for the spatial data.