CDDA version 13 (2015)

Prepared by / compiled by: Manuel Löhnertz Organisation: ETC/BD – space4environment EEA project manager: (Mette.Lund@eea.europa.eu) Task Manager: (Sabine Roscher/roscher@mnhn.fr) Work Package n°:1.7.2.A

Contents

1	Background & Introduction				
	1.1	CDDA	3		
	1.2	Definition of terms	3		
2	Deliv	very of datasets	5		
	2.1	The CDDA v13 delivery	5		
3	CDE	DA Production	9		
4	Veri				
	4.1	Verification overview			
	4.2	Verification steps tabular data			
	4.3	Verification steps spatial data	11		
5	14				
	5.1	First spatial verification information	14		
6	Concluding remarks1				

1 Background & Introduction

The present document describes the activities and procedures for verification of the CDDA version 13 (2015).

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 European datasets have been uploaded to: <u>https://svn.eionet.europa.eu/repositories/Workdata/CDDA/cdda_ver13/</u>

2.1 The CDDA v13 delivery

The latest version of the Common Database on Designated Areas (CDDA), version 13 from 2015, 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 39 EEA and EEA cooperating countries as well as Greenland (Denmark) and the French Overseas Departments and Territories and Overseas Collectives (Map 2-1).



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

32 countries delivered new tabular and spatial data in 2015, which had to be included into version 13 of CDDA. Four countries informed EEA that no updates were necessary and that the delivery from the previous year can be used in 2015; 3 countries did not deliver data.

All datasets were subjected to a series of quality control and quality assurance (QA/QC) checks. Once the data passed these tests it was combined with data from those countries that did not submit data in 2015. For those countries data was extracted from the previous CDDA dataset, version 12.

The combined and integrated dataset (i.e. 2015 CDDA, version 13) covers 39 countries, and consists of a total of **100 181** records in the tabular database and **97 752** spatial records.

CDDA version	tabular	Spatial
Version 2015 v_13	100 181	97 752
Version 2014 v_12	98 367	95 109
Version 2013 v_11	97 481	92 757
Version 2012 v_10	94 810	86 226

Table 2-1 Number of tabular and spatial records in the different CDDA versions

The following map presents the countries, which provided data for the CDDA version 13.

Map 2-1 Countries that delivered data in 2015



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).

N			Version 12	Version 13	
Name	ISO - 2 digit	ISO - 3 digit	2014	2015	
Albania	AL	ALB	Yes	no new data	
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	СҮ	СҮР	no	no new data	
Germany	DE	DEU	yes	yes	
Denmark	DK	DNK	yes	no	
Estonia	EE	EST	yes	yes	
Spain	ES	ESP	yes	yes	
Finland	FI	FIN	yes	yes	
France	FR	FRA	yes	yes	
Greece	GR	GRC	yes	yes	
Croatia	HR	HRV	yes	yes	
Hungary	HU	HUN	yes	no new data	
Ireland	IE	IRL	yes	yes	
Iceland	IS	ISL	yes	yes	
Italy	IT	ITA	yes	yes	
Liechtenstein	LI	LIE	no	no new data	
Lithuania	LT	LTU	no	no	
Luxembourg	LU	LUX	no	yes	
Latvia	LV	LVA	yes	yes	
FYROM	МК	MKD	yes	yes	
Malta	MT	MLT	yes	yes	
Montenegro	ME	MNE	yes	yes	
Netherlands	NL	NLD	yes	yes	
Norway	NO	NOR	yes	yes	
Poland	PL	POL	no	yes	
Portugal	PT	PRT	yes	yes	
Romania	RO	ROU	yes	yes	
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	XKX	yes	yes	

Table 2-2 Data deliveries respectively data updates for CDDA version 12 & 13 (tabular & spatial)

The national CDDA data were delivered as tabular and 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 v13 database and 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 individual national vector data files were first merged to one national spatial dataset. In a next step, all merged spatial datasets were used for the construction of the final European 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 v13 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 v13 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, e.g. where information was missing, the countries were asked to update the data.

After a successful quality check of the deliveries, the single tables were imported to the CDDA-v13 database template (provided by the EEA). Afterwards final QA checks were done using the integrated QA tool.

Figure 4-1 Tabular data verification overview, version 13



Spatial:

The countries delivered various kinds and quantities of spatial data. Some countries delivered one polygon vector file, others more than 5 different polygon and point vector files.

Figure 4-2 Example: spatial delivery by one country



Therefore the first step of the verification was to check each spatial dataset prior to merging to one national dataset.

After the data passed the semantic check (projection, format, ...) without errors, all data was merged into one European CDDA dataset in a file-geodatabase. If a country did not delivered new data, the most recent data available to the EEA 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.





4.2 Verification steps tabular data

The verification of the tabular data was done on Eionet after data upload (<u>https://www.eionet.europa.eu/news/cdda2015/#ch5</u>).

The automatic QA checks following data quality indicators:

- availability of mandatory fields
- duplicities in primary key fields
- correctness of data types of provided values
- correctness of used codes where codelists are defined, including the correct use of site codes
- position of sites coordinates against the country boundaries

4.3 Verification steps spatial data

In the following chapter the verification of the spatial data will be explained.

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 with corresponding formats are available)

• projection (WGS84, ...)

- (if projection is not ETRS-LAEA, the dataset has been reprojected)
- 2. Mapping check
 - Unique identifiers (SITE_CODE) ([SITE_CODE] is mandatory!)
 - 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 center point location of each CDDA site was checked using an European boundary layer evaluating the sites position in relation to the country boundary: EuroBoundaryMap (full European coverage) - version 8.1, Apr. 2014

(http://sdi.eea.europa.eu/catalogue/srv/ger/catalog.search#/metadata/658f9a54-2820-46b0-9e2c-961f43e38154)

After the first verification check by spatial and tabular data type, cross-comparisons between the delivery types were made:

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



• SITE_CODE: It was checked if all codes reported in the spatial data 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. If the LAT/LON position was not to be calculated based on the spatial delivery the coordinate reported by the member state was used instead. The location of this centroid (LAT/LON) was compared against the country border. Three situations were distinguished, where the differences are <1km, >1km and >10km

(reference layer: EuroBoundaryMap + Administrative units 2010 at country level for CDDA regions outside the European continent)

- AREA: The area per polygon for each site was calculated and compared to the areas in the descriptive database for the same site, where the area was given.
- Major_ecosystem_type: The major reported ecosystem of the site is compared to the European coastline layer (<u>http://www.eea.europa.eu/data-and-maps/data/eea-coastline-for-analysis</u>) to determine whether the defined major ecosystem is correct.

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

Figure 5-1 Semantic check



Table 5.1 shows the results of the semantic checks (Fig 5.1) on the spatial deliveries.

|--|

Country	ISO3	Comment
Albania	ALB	• no new delivery – data from 2014 is used
Austria	AUT	 wrong projection (EPSG 31287) double sites-codes found sites which were already deleted in the past have been found: SITE_CODE: 5559 169133 169593 387354 387436 555513786
Bosnia - Herzegovina	BIH	no new delivery – data from a previous version is used
Belgium	BEL	• OK
Bulgaria	BGR	• OK
Switzerland	CHE	 wrong projection no SITE_CODE information double sites found (<i>site_code_nat</i>)
Czech Republic	CZE	• OK
Cyprus	СҮР	no new delivery – data from a previous version is used
Germany	DEU	 double site_codes: same SITE-code: "555558886" but different names: Gonsbachtal & Schwarzwald) → typing error: the site Gonsbachtal was changed to "555558888"
Denmark	DNK	no new delivery – data from a previous version is used
Estonia	EST	• wrong projection (EPSG 3301)

Country	ISO3	Comment
Spain	ESP	• OK
Finland	FIN	• OK
France	FRA	• OK
Greece	GRC	• OK
Croatia	HRV	 "CDDA_kod" instead of SITE_CODE attribute double site_codes: (377986: 15642: 377841)
		• spatial site "3493049" cannot be found in the tabular database
Hungary	HUN	• OK
Ireland	IRL	• OK
Iceland	ISL	• OK
Italy	ITA	• OK
Liechtenstein	LIE	no new delivery – data from a previous version is used
Lithuania	LTU	no new delivery – data from October 2014 is used
Luxembourg	LUX	 wrong projection five different spatial datasets
		 in some of the datasets field SITE_CODE is missing double SITE-CODES missing CDDA sites (16358: 28416: 337459)
Latvia	LVA	wrong projection
FYROM	MKD	• OK
Malta	MLT	no new delivery – data from a previous version is used
Montenegro	MNE	• site with site-code <u>15749</u> (Skadarsko lake partial) cannot be found in the tabular database
Netherlands	NLD	number of single spatial files is not correct wrong projection
		• SITE with SITE-CODE <u>555561871</u> cannot be found in the database
Norway	NOR	double SITE_CODES
Poland	POL	• OK
Portugal	PRT	wrong projection
Romania	ROU	• OK
Serbia	SRB	 wrong projection double SITE_CODES
Sweden	SWE	• OK
Slovakia	SVK	double SITES_CODES
Slovenia	SVN	 double SITE_CODES in the point dataset some sites can be found in the point AND in the polygon dataset
Turkey	TUR	 11 different spatial datasets wrong projection
		missing SITE_CODE information
United Kingdom	GBR	 some sites are marked as "disappeared" in the EIONET table minor geometrical shift including same geometry for two sites (968 vs. 139760) google-map_link
Kosovo (UNSCR 1244/99)	XKX	wrong projection

Another output of the semantic checks is the comparison of the site area between the new spatial CDDA version and the previous version 12 from 2014.

PARENT_ISO	area km² 2014	area km² 2015	Difference in area 2015/2014 [km ²]	Difference in area 2015/2014 [%]	INFO
ALB	4810	4810	0	0	no new data
AUT	26274	26780	506	1.93	increase
BEL	5056	5171	114	2.26	increase
BGR	17926	6271	-11 655	-65.02	decrease
BIH	99	99	0	0	no new data
CHE	3033	3027	-6	-0.21	no changes (±1%)
СҮР	203	203	0	0	no new data
CZE	13102	13205	103	0.79	no changes (±1%)
DEU	132773	133953	1 180	0.89	no changes (±1%)
DNK	985843	985843	0	0	no new data
ESP	74577	79599	5 022	6.73	increase
EST	22760	23716	957	4.2	increase
FIN	34551	35369	818	2.37	increase
FRA	270635	297905	27 270	10.08	increase
GBR	102761	162799	60 038	58.42	increase
GRC	38035	38035	0	0	no changes (±1%)
HRV	7171	7202	31	0.43	no changes (±1%)
HUN	8496	8496	0	0	no new data
IRL	604	603	0	0	no changes (±1%)
ISL	20082	20087	5	0.02	no changes (±1%)
ITA	59388	59388	0	0	no changes (±1%)
LIE		82	0	0	no new data
LTU	10935	10935	0	0	no new data
LUX	1258	641	-617	-49.05	decrease
LVA	16839	16840	0	0	no changes (±1%)
MKD	2046	2046	0	0	no changes (±1%)
MLT	338	340	2	0.65	no changes (±1%)
MNE	0,006	0,005	0,001	-20.62	decrease
NLD	12783	12799	16	0.12	no changes (±1%)
NOR	181981	182105	124	0.07	no changes (±1%)
POL	60040	103957	43 917	73.15	increase
PRT	121022	121433	411	0.34	no changes (±1%)
ROU	20033	20007	-26	-0.13	no changes (±1%)
SRB	5693	5106	-586	-10.3	decrease
SVK	11322	12451	1 129	9.98	increase
SVN	27975	28113	138	0.49	no changes (±1%)
SWE	57327	57975	648	1.13	increase
TUR	38186	36161	-2 025	-5.3	decrease
XKX	1245	1240	-5	-0.44	no changes (±1%)
	1400407	1529898	129 491	9.25	increase

Table 5-2 CDDA area comparison: CDDA_v12 vs CDDA_v13 (spatial datasets)

The tabular database contains more CDDA sites than the spatial dataset. Therefore the area statistic looks different for the tabular dataset:

Country	PARENT_ISO	area km² 2014	area km² 2015	Difference 2015- 2014 km ²	Difference 2015- 2014 [%]	INFO
Albania	ALB	4648	4648	0	0	no new data
Austria	AUT	26312	26312	0	0	no changes (±1%)
Belgium	BEL	7081	8340	1 259	17.78	increase
Bulgaria	BGR	14451	15158	707	4.89	increase
Bosnia -	DIII	201	201			
Herzegovina	BIH	391	391	0	0	no new data
Switzerland	CHE	3479	3479	0	0	no changes (±1%)
Cyprus	CYP	3789	3789	0	0	no new data
Czech Republic	CZE	13077	13178	101	0.77	no changes (±1%)
Germany	DEU	132358	133801	1 443	1.09	increase
Denmark - Greenland	DNK	995884	995884	0	0	no new data
Spain	ESP	75170	79682	4 512	6	increase
Estonia	EST	23550	23716	166	0.7	no changes (±1%)
Finland	FIN	30388	35322	4 934	16.24	increase
France DOM/TOM	FRA	291263	299801	8 538	2.93	increase
United Kingdom	GBR	100949	162741	61 792	61.21	increase
Greece	GRC	38130	38130	0	0	no changes (±1%)
Croatia	HRV	8081	8078	-3	-0.04	no changes (±1%)
Hungary	HUN	11319	11319	0	0	no new data
Ireland	IRL	2009	2006	-3	-0.15	no changes (±1%)
Iceland	ISL	20084	20089	5	0.02	no changes (±1%)
Italy	ITA	60227	60227	0	0	no changes (±1%)
Liechtenstein	LIE	82	82	0	0	no new data
Lithuania	LTU	10582	10933	352	3.32	no new data
Luxembourg	LUX	1260	1279	18	1.45	increase
Latvia	LVA	16831	16856	25	0.15	no changes (±1%)
Macedonia	MKD	2297	2297	0	0	no changes (±1%)
Malta	MLT	340	340	0	-0.01	no changes (±1%)
Montenegro	MNE	1313	1314	0	0.01	no changes (±1%)
Netherlands	NLD	12783	12800	17	0.13	no changes (±1%)
Norway	NOR	181446	182332	886	0.49	no changes (±1%)
Poland	POL	91865	103886	12 021	13.09	increase
Portugal	PRT	121636	121884	248	0.2	no changes (±1%)
Romania	ROU	20354	20354	0	0	no changes (±1%)
Serbia	SRB	5354	5727	373	6.97	increase
Slovakia	SVK	12180	12224	44	0.36	no changes (±1%)
Slovenia	SVN	27972	28109	138	0.49	no changes (±1%)
Sweden	SWE	57321	57964	642	1.12	increase
Turkey	TUR	39790	111538	71 749	180.32	increase
Kosovo	XKX	1237	1241	4	0.32	no changes (±1%)
L	TOTAL	2467286	2637251	+169 966 km²	+6.89%	increase

Map 5-1 CDDA area comparison: CDDA_v12 vs CDDA_v13 (tabular datasets)

The total tabular CDDA area increased from 2 467 286 km² to 2 637 251 km² between versions 12 and 13. That's an increase of 169 966 km². Larger increases have been identified in Poland and in the United Kingdom. In Poland a large number of new CDDA sites was added. In the following map the difference between the versions can be clearly seen:

Figure 5-2 CDDA Poland version 2014 vs. 2015



Poland CDDA 2014

Poland CDDA 2015

Also in the UK a large number of sites – especially marine sites – was added to the new CDDA dataset as shown in the map below.





UK CDDA 2014

UK CDDA 2015

For Turkey, the tabular CDDA area increased by 71 749 km². However, the figure does not reflect the reality. During the data integration a number of problems occurred. Sites could not be allocated correctly because of missing SITE_CODEs in both spatial and tabular datasets. As a result, the Turkish data could not be fully imported to the new European CDDA dataset and a number of reported sites are missing.

6 Concluding remarks

While the majority of the data delivered under the CDDA reporting cycle is of a high quality, there are still some problems during the data processing due to the following:

Tabular dataset

- Not 100% specification conform, despite the specification and the automatic QA some records are not correctly filed.
- [designation_boundaries] information delivered by some countries but without providing the corresponding spatial data

Spatial dataset

The new spatial CDDA specifications are used by the most countries for data delivery, but some countries still deliver spatial data which is not 100% specification conform:

- Some countries delivered more than one shape-file
- Some countries delivered in national projection
- Deliveries received from few countries have shown inconsistent or erroneous differences in the spatial data compared to previous datasets