EBM Data Quality Report

File name	EBM_2020_QualityReport		
Version	Author	Date	Comments
v1.0	Jörgen Spradau / BKG	06.01.2020	

Contents

1 Overview

BKG, as EBM production manager, has applied a complex and consistent set of quality checking procedures to guarantee a high quality level of the EuroBoundaryMap 2020 database. This comprises:

- Usage of database templates based on the EBM data specification to avoid errors in the data structure,
- Definition of core feature types and automatic routines to derive additional feature types:
  - Core feature types: feature classes EBM_A, ResidenceOfAuthority, related tables EBM_NAM, EBM_ISN, EBM_NUTS,
  - Derived feature types: feature classes AdministrativeUnit_x, AdministrativeBoundary, LAU, NUTS_x,
- Automatic quality checking procedures applied on the national contributions as well as on the final database,
- Clear workflow and documentation of data production, including cooperation with data providers for feedback and corrections.

This additional document gives an overview of different aspects of the data quality of EBM 2020. This report is based on the principles for describing the quality of geographic information established with ISO standard 19157 (former: ISO standard 19113). More details of national specifics are described in the metadata of each country.

2 Data quality elements

2.1 Completeness

2.1.1 Omission

Country datasets

EuroBoundaryMap contains national contributions from 39 National Mapping and Cadastral Agencies. This covers all 28 EU Member States, 3 candidate countries, 4 EFTA countries and 8 other European countries. Only some former Yugoslav federal states, Turkey, Russia and Belarus are missing for a complete coverage of Europe. EBM 2020 contains placeholders for these countries. For further details refer to *EBM_2020_Specification.pdf* (chapter 4.2).

Features

All features were captured according to the data specification. The completeness of features, in feature classes as well as tables, is assumed as 100%. It's difficult to validate the completeness of features. One possible check is to compare the administrative units in EBM with the list of LAU units provided by Eurostat. This matching resulted in a low number of inconsistencies, but it has to be clarified if the error is in EBM or in the LAU table (see *EBM_2020_LAU-NUTS-Matching.pdf* for details).

EBM contains only two optional features:

- Costal water: The territorial sea is provided for 9 of 25 countries.
- Residence of Authority: All national capitals are included in EBM 2020. Administrative seats of lower administrative levels are provided for 51 countries.

Attributes

All attributes were captured according to the data specification. The completeness of attributes is 100%.

Table EBM_NAM contains three attributes with additional information. These attributes were populated as follows:

- PPL (population): Attributed for most countries by the NMCAs from the official national register. Alternatively, population adopted from documents officially released by the national statistical institutes. For some countries, population figures are missing for lowest level units.
- ARA (area size): Populated with the official size of administrative units for a number of countries. Values for all other countries were calculated from the shape size of the polygons in the EBM database.

- effectiveDate: Attributed, at least, for administrative units changed before 01.01.2019. For some countries there were no changes in this period.

2.1.2 Commission

Minimum size of polygon features is in general 4 ha. Exceptions are allowed:

- For administrative units, where the main area is smaller 4 ha,
- For small islands or exclaves which are of major importance for the national territory.

For some countries it is questionable, if islands smaller 4 ha are really needed to describe the national territory. Nevertheless, such small islands were kept to get a consistent solution for the coastlines in EBM and ERM.

2.2 Temporal quality

The reference date of EBM 2020 is the 31 December 2018. For few countries it was not possible to get an update according to this date. For more details see *EBM_2020_Changes.doc* and *EBM_2020_ChangesDetails.xls.*

The reference date of population figures (attribute PPL in table EBM_NAM) is either 31 December 2018, the national date of Census 2011 or a date between these dates.

See *EBM_2020_QualityReport_CountryOverview.xls* for more details.

2.3 Positional accuracy

Due to the fact that EuroBoundaryMap is compiled of national contributions, the positional accuracy depends on the accuracy of the national source databases. EBM is intended to be used in map scale 1:100 000. For that scale a positional accuracy of about 50 m is suitable. All data providers were asked to deliver their data with that value of accuracy.

Some national datasets were derived from national large scale databases with a high positional accuracy of 10 m or better. Nevertheless, all those data have been generalised to get a harmonised EBM dataset. In that way it is difficult to describe the positional accuracy exactly.

The positional accuracy of all national contributions, as reported by the NMCAs, is listed in *EBM_2020_QualityReport_CountryOverview.xls*.

2.4 Logical consistency

2.4.1 Conceptual, domain and format consistency

The adherence with the conceptual schema of EBM is given, because all data is stored in database templates which were created based on the EBM specification. This consistency includes:

- General structure of the datasets,
- Spatial reference system is ETRS89,
- Spatial features have a valid geometry,
- Compliance of feature attributes with attribute domains,
- Linkage between feature classes and tables.

2.4.2 Topological consistency

The basic geometry of EBM is stored in feature class EBM_A. For all polygons in this feature class the compliance with the following topological rules can be guaranteed:

- Polygons must not overlap.
- Gaps between polygons are not allowed.
- Neighbouring polygons share the same set of coordinates on their border. This includes all polygons on international borders.
- No adjacent polygons have the same set of attribute values.
- National features are properly inside the national territory.
- Minimum distance separating all nodes and vertices of all outlines of polygons is 5 meters.

All other features classes (AdministrativeUnit_x, AdministrativeBoundary, LAU, NUTS_x) are derived from the polygons by automatic routines. The topological consistency is given by default, additional checks are not needed.

2.5 Thematic accuracy

Until EBM v11 Eurostat requested two LAU levels. With the beginning of EBM v12 Eurostat changed their internal processes and requested only one LAU level. This LAU Level corresponds exactly the lowest administrative level in each country. With regard to earlier releases of EBM some countries provides also units on lower level, which are electoral districts or parishes. This comprises the following countries:

- Denmark: Parishes (Sogne) were integrated in EBM as lowest level, although these units are not of administrative meaning.
- Slovenia: Settlements (Naselje) were integrated in EBM as lowest level, although these units are not of administrative meaning.

The administrative hierarchies of all countries were carefully checked to achieve consistent and harmonized structures of hierarchical levels. For some countries a slight generalisation was necessary to improve the understanding of national specifics. The administrative levels are stored in table EBM_ISN. For further details refer to the national metadata.

SHN code is used as European-wide harmonized and unique identifier for all administrative units. This code is a strictly hierarchically built identifier for all administrative units on each administrative level. In general, SHN corresponds to the national administrative code. The only deviations are the SHN codes of Great Britain and Northern Ireland. A new coding system has been introduced by the Office for National Statistics (ONS). The new codes are not compliant with the EBM requirements for unique identifiers of European administrative units. The ONS codes identify administrative units only within each administrative level – they don't reference the upper administrative levels.

All names of administrative units are included as provided by the national data providers. The names contain all national diacritics, but an additional conversion in simple ASCII code is attached. Transliterations of non-Latin geographical names use the official national conversion schema. The completeness and correctness of the names were checked. In some cases, the official names of the administrative units may differ from commonly used names. For some countries it is not possible to exclude the designations of hierarchical levels from the names. According to the EBM specification, all designations have to be stores in table EBM_ISN.

3 EBM quality checking procedures

3.1 Methodology

A number of automatic tools have been developed to validate the quality of national EBM contributions as well as the quality of the full Europe database. The EBM QC workflow is described in the following list:

- Import of national data contributions into an EBM database template. This import provides information about:
    - Features not compliant with EBM data schema
    - Missing Features
    - Additional features not needed for EBM
- Comparison of new delivery with data from previous release. This analysis gives an impression of the amount of changes. Sometimes systematic errors can be detected.
- Geometry checks for EBM_A:
    - Search for multipart polygons (has to be single part)
    - Check minimum size (4ha)
    - Topology:
        - Outline of national territory (dissolved area of EBM_A) must be covered by international boundaries
        - Areas must not overlap
        - Areas must not have gaps
    - Check outlines of polygons:
        - No duplicate vertices within 5m
        - No cutbacks (angle between connected line segments) smaller 5°
        - Calculate average distance between vertices. This provides information about generalisation degree.
    - Visual assessment of generalisation degree
- Attribute checks for EBM_A:
    - Neighbouring polygons must have different SHN
    - Check of attribute TAA: every unit (SHN) must have exactly one mainland
    - Specific attribute check for water areas
- Attribute checks for EBM_NAM:
    - Interactive analysis of SHN structure and referring table EBM_ISN
    - Check attribute USE
    - Check name conversion (NAMN to NAMA)
    - Interactive validation of names and language codes
    - Check attributes PPL, ARA, effectiveDate
- Validation of NUTS matching (table EBM_NUTS):
    - Comparison of EBM data and LAU tables provided by Eurostat
- Check linkage between EBM_A and additional tables (EBM_NAM, EBM_ISN, EBM_NUTS)
- Automatic deriving of additional feature classes:
    - AdministrativeBoundary
    - AdministrativeUnit_x
    - Statistical regions LAU, NUTS_x

3.2 Results of quality checking

The automatic tools described above list a number of remaining inconsistencies. A summary is given below. Please refer to the national metadata for detailed information.

- Geometry checks for EBM_A:
  - Polygons smaller 4 ha (EBM minimum size) are included because of the following reasons:
    - Denmark (2 areas), Spain (1 area), Slovenia (1 area), France (1 area), Bulgaria (1 area): main areas
    - Germany (1 area), Ireland (1 areas), France (5 areas): coastal waters cannot be logically merged with other units
    - Austria (1 area), Germany (15 areas), Denmark (1 area) ,: branch areas cannot be logically merged with other units
    - Croatia (1 area),: exclaves
    - France (7 areas), Portugal (2 areas): marginal smaller 4ha
    - Great Britain (53 areas): intersection of the coastline with the administrative units is very complex, small polygons cannot be avoided in every case
    - Cyprus (5 areas), Greece (35 areas), Italy (1 areas), Spain (5 areas), Faroe Islands (2 areas), Netherlands (1 area): small islands which are of major importance to outline the territorial sea
    - French Guiana (423 areas): small islands which depends on a complete new geometries for French Guiana from France. These are more or less important in the source database.
- Linkage between EBM_A and additional tables:
  - EBM_A to EBM_NUTS:
    - No entries in table EBM_NUTS for the following countries or territories: Albania, Andorra, Bosnia and Herzegovina, Faroe Islands, Gibraltar, Greenland, Kosovo, Moldova, Monaco, San Marino, Ukraine and Vatican.
    - In some countries specific water bodies are own units with SHN codes, e.g. territorial waters. Those units have no entry in EBM_NUTS. This concerns water bodies in the following countries: Bulgaria, Germany, Greece, Croatia, Lithuania, and Slovenia.
    - All in dispute areas have no linked entry in EBM_NUTS.
  - EBM_A to EBM_NAM:
    - The cardinality is set up in the EBM data model as 1..* to 1. This cardinality is applicable only for the direction EBM_A to EBM_NAM.
    - The direction EBM_NAM to EBM_A is not applicable because table EBM_NAM contains also entries for the upper levels in the national administrative hierarchy.
    - That's why the linkage has to be checked from EBM_NAM to all administrative levels (AdministrativeUnit_x). Remaining inconsistencies concern the countries with specific water bodies (see above).
    - All in dispute areas have no linked entry in EBM_NAM.